Review of International Studies (2007), 33, 435–449 Copyright © British International Studies Association
doi:10.1017/S0260210507007590

Enforcing the Kyoto Protocol: can punitive consequences restore compliance?

JON HOVI, CAMILLA BRETTEVILLE FROYN AND GURI BANG*

Abstract. To enhance effective cooperation, the Marrakesh Accords provide a compliance system for the international climate regime. An innovative part of this system is an Enforcement Branch authorised to apply punitive consequences against countries that fail to comply with their Kyoto obligations. While previous research has primarily focused on the ability of this compliance system to deter non-compliance, this article discusses whether the actual use of punitive consequences will induce a non-compliant country to *return* to compliance. The Marrakesh Accords explicitly emphasise that the punitive consequences shall be aimed at the restoration of compliance. However, we show that the application of punitive consequences will accomplish this in only *exceptional* circumstances.

1. Introduction

The 1992 UN Framework Convention on Climate Change (UNFCCC), to which virtually all UN member countries are parties, acts as an important framework for a continued international climate regime. It states that industrialised countries have a historic responsibility to take the lead in reducing greenhouse gas (GHG) emissions. The Kyoto Protocol, which came into force in February 2005 and is now legally binding, commits industrialised countries to reduce their emissions of six GHGs by 5 per cent below 1990 levels on average in the period 2008–2012.[1] Key features of the Protocol are the three flexibility mechanisms: emissions trading, joint implementation, and the clean development mechanism. These three flexibility mechanisms are intended to limit the cost of reducing emissions, by allowing for flexibility in how parties fulfil their obligations.

The Marrakesh Accords provide detailed prescriptions for a compliance system for the Kyoto Protocol. While most other international environmental agreements have (at best) weak institutions for compliance, an innovative part of Kyoto's compliance system is an Enforcement Branch authorised to apply punitive 'consequences' to countries that fail to comply with their Kyoto obligations. Most importantly, a country that fails to meet its assigned emissions target must cover its

* We are indebted to Kristin Rypdal and three anonymous referees for helpful comments and to Frank Azevedo for excellent editorial assistance. Financial support from the Research Council of Norway is gratefully acknowledged.

[1] The commitment to reduce emissions by 5 per cent is described in Annex I of the Kyoto Protocol, and therefore countries that are bound by the commitment are often referred to as 'Annex I countries'.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

deficit plus an additional 30 per cent in the next commitment period, and loses its eligibility to sell emission permits.[2]

There is a growing body of literature on the question of compliance with the Kyoto Protocol. One strand of this literature characterises and comments upon the Protocol's overall approach to compliance.[3] A second strand considers how the climate regime can achieve clarity regarding commitment, compliance, and regime response.[4] Finally, a third strand of literature proposes measures that the climate regime might use against a non-compliant country,[5] and identifies the conditions under which a compliance system for the climate regime might effectively deter non-compliance.[6]

This third strand of literature points out a number of potential weaknesses in Kyoto's compliance system. Barrett notes five major weaknesses.[7] First, the Marrakesh Accords do not include any enforcement provisions addressing failure by a non-compliant country to accept its punishment. Hence, the compliance system relies heavily on 'self-punishment', in the sense that implementation of the punitive consequences requires cooperation by the non-compliant country. Second, nothing in the Accords prevents a country from postponing implementation of the punishment to a later commitment period. In fact, implementation could conceivably be delayed indefinitely. Third, the compliance system is not legally binding and can be made so only through an amendment which requires a three-fourths majority vote by the Meeting of the Parties. Even if such an amendment is passed, the compliance system becomes binding only on those countries that *ratify* the amendment. Fourth, the anticipation of being punished is likely to induce countries to hold out for a low target in the upcoming negotiations on emission targets for the second commitment period. If they were successful in this it would reduce the *de facto* punishment. Fifth, any party is entitled to withdraw from the Kyoto Protocol with 12 months' notice. Withdrawal is not cost-free, since it will exclude the country in question from participating in the Kyoto flexibility mechanisms, reduce its influence in future climate negotiations, and damage its reputation. Nevertheless, a non-compliant country can escape punishment if it withdraws from the climate regime.[8]

[2] For further details on Kyoto's compliance system, see Section 2 below.

[3] See Daniel Bodansky 'International Law and the Design of a Climate Change Regime', in U. Luterbacher and D. F. Sprinz (eds.), *International Relations and Global Climate Change* (Cambridge, MA: MIT Press, 2001), Michael Grubb, Christiaan Vrolijk and Duncan Brack, *The Kyoto Protocol. A Guide and Assessment* (London: The Royal Institute of International Affairs, 1999), and Sebastian Oberthür and Hermann E. Ott, *The Kyoto Protocol. International Climate Policy for the 21st Century* (Berlin: Springer Verlag, 1999).

[4] Ronald B. Mitchell, 'Institutional Aspects of Implementation, Compliance, and Effectiveness', in Urs Luterbacher and Detlef F. Sprinz (eds.), *International Relations and Global Climate Change* (Cambridge, MA: MIT Press, 2001) and Ronald B. Mitchell, 'Flexibility, Compliance and Norm Development in the Climate Regime', in Olav S. Stokke, Jon Hovi and Geir Ulfstein (eds.), *Implementing the Climate Regime: International Compliance* (London: Earthscan, 2005).

[5] Scott Barrett and Robert N. Stavins, 'Increasing Participation and Compliance in International Climate Change Agreements', *International Environmental Agreements: Politics, Law, and Economics*, 3 (2003); Tim Hargrave, Ned Helme, Suzi Kerr and Tim Denne, *Defining Kyoto Protocol Non-Compliance Procedures and Mechanisms* (Leiden: Center for Clean Air Policy, 1999).

[6] Scott Barrett, 'Consensus Treaties', *Journal of Institutional and Theoretical Economics*, 158 (2002), Scott Barrett, *Environment and Statecraft. The Strategy of Environmental Treaty-Making* (New York: Oxford University Press, 2003), Cathrine Hagem and Hege Westskog, 'Effective Enforcement and Double-Edged Deterrents', in Stokke et al. (eds.), *Implementing the Climate Regime*.

[7] Barrett, *Environment and Statecraft*, pp. 385–6.

[8] It is a standard feature of most international agreements that parties to them have the right to withdraw, provided that they give due notice. However, it is not a standard feature that a

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

At least one more weakness can be added to Barrett's list. The punitive consequences will produce negative economic effects not only for the non-compliant country itself, but also for some other countries.[9] For example, because the punished country will have to reduce its emissions more (because of the additional deficit penalty) than it might otherwise have had to, it will probably, as a result, be forced to find ways to use less fossil fuel. Consequently, the demand for, and therefore the price of, fossil fuels will fall.[10] While this will benefit countries that are net importers of fossil fuels, it will harm countries that are net exporters. Also, the punitive consequences will reduce the supply of emission permits, thereby causing the permit price to go up.[11] For countries that will eventually be both exporters of fossil fuels and buyers of permits, the welfare effects might actually be worse if *another* country (notably Canada, Japan or Russia) is punished than if they are punished themselves.[12] Obviously, this will likely be seen as unfair. Moreover, it creates incentives for strategic behaviour for the members of the Enforcement Branch. Even though the punitive consequences are both predetermined and automatic, the members of the Enforcement Branch will likely have some discretion in deciding whether a country is in compliance (see Section 4 below). If at least two Enforcement Branch members from Annex I countries take the above-mentioned effects into consideration, a country that is technically in non-compliance might nevertheless be able to escape punishment.[13]

As this brief review demonstrates, there is a good deal of literature addressing the capacity of Kyoto's compliance system to deter non-compliance. A major conclusion is that this capacity is likely to be insufficient to achieve a compliance rate near 100 per cent. It is therefore interesting to consider whether the use of punitive consequences will induce a country that has violated its Kyoto commitments to *return* to compliance. The Marrakesh Accords explicitly emphasise that the punitive consequences to be applied by the Enforcement Branch 'shall be aimed at the restoration of non-compliance (sic!) to ensure environmental integrity, and shall provide for an incentive to comply'.[14] Nevertheless, the question of whether applied punitive consequences can actually restore compliance has so far been largely ignored in the literature. The aim of this article is to fill this gap.

Intuition suggests that punitive consequences should cause a return to compliance if those consequences are relevant, potent, and credible. However, we argue that these are basically conditions necessary for the punitive consequences to effectively *deter* non-compliance. If punitive consequences are *applied*, however, deterrence would

non-compliant country can *escape punishment* by pulling out of the agreement. For example, withdrawal by a non-compliant member of the World Trade Organization (WTO) could easily *magnify* the effect of its punishment by the Dispute Settlement Body because withdrawal would jeopardise the non-compliant country's most-favoured-nation status *vis-à-vis* remaining members of the WTO.

[9] Hagem and Westskog, 'Effective Enforcement and Double-Edged Deterrents', pp. 110–13.

[10] The size of this effect will obviously be larger for major than for minor countries.

[11] Note that it is the right to *sell* emissions permits that is suspended (for example, see: ⟨http://unfccc.int/kyoto_mechanisms/compliance/items/3024.php⟩). Hence, such suspension will cause the supply of permits to be reduced, but will not impact demand. As a result, the permit price will be higher than otherwise.

[12] Steffen Kallbekken and Jon Hovi, 'The Price of Non-compliance with the Kyoto Protocol: The Remarkable Case of Norway', *International Environmental Agreements* (forthcoming).

[13] Hagem and Westskog, 'Effective Enforcement and Double-Edged Deterrents', pp. 114–15.

[14] UNFCCC, 'The Marrakesh Accords and the Marrakesh Declaration', p. 132. ⟨http://unfccc.int/cop7/accords_draft.pdf⟩.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

have already *failed*. Hence, to identify the conditions necessary for punitive consequences to *restore* compliance, one needs to take this fact into account.

We argue that only under exceptional circumstances should one expect the use of punitive consequences to cause a return to compliance. So, the Kyoto compliance system is not only *weak in deterring non-compliance* but also *limited in its ability to restore compliance* by applying punitive consequences. However, its limited ability to restore compliance should not be considered a weakness. In fact, we argue that effective deterrence is the best that can be hoped for from a punitive compliance system.[15]

The literature suggests that compliance is determined not only by incentives, but also by several other factors: state identities and norms of appropriate behaviour,[16] the perceived legitimacy of international rules,[17] the extent to which international rules are adopted as domestic rules,[18] member states' capacities to comply,[19] and social and economic changes over time.[20] While some of these other factors *might* cause a non-compliant country to return to compliance, this possibility is not considered in this article. Rather, we argue that the punitive consequences are unlikely to make much of a *difference* in a country's decision about whether to return to compliance.

The remainder of the article is organised as follows: Section 2 provides a brief account of the Kyoto-Marrakesh compliance system. Section 3 distinguishes two types of non-compliance. Section 4 discusses in some detail the conditions under which applied punitive consequences will cause a country to return to compliance, and argues that such cases will likely be rare. While the discussion in Section 4 assumes perfect monitoring of emissions inventories, we explain in Section 5 that our conclusion holds even with imperfect monitoring. Section 6 explains why the inability to restore compliance is not a weakness of the compliance system. Finally, Section 7 presents some conclusions.

2. Kyoto's compliance system

Kyoto's compliance system is laid out by the Marrakesh Accords. The Accords establish a Compliance Committee with both a Facilitative Branch and an Enforcement Branch to control compliance with the commitments in the Kyoto Protocol.

[15] The discussion in this article focuses on the compliance system for the current climate regime. However, the underlying logic of our argument is more general. Therefore, it could be applied to other mechanisms for achieving compliance with international norms, such as economic sanctions. See Jon Hovi, Robert Huseby and Detlef F. Sprinz, 'When Do (Imposed) Economic Sanctions Work?', *World Politics*, 57 (2005), pp. 479–99.

[16] See Jeffrey Checkel, 'Why Comply? Social Learning and European Identity Change', *International Organization*, 55 (2001), pp. 557–8.

[17] Thomas M. Franck, 'Legitimacy in the International System', *American Journal of International Law*, 82 (1988), p. 706.

[18] Harold H. Koh, 'Transnational Legal Process', *Nebraska Law Review*, 75 (1996), p. 205.

[19] Abram Chayes, Antonia H. Chayes and Ronald B. Mitchell, 'Managing Compliance: A Comparative Perspective', in E. B. Weiss and H. K. Jacobson (eds.), *Engaging Countries: Strengthening Compliance with International Environmental Accords* (Cambridge, MA: MIT Press, 1998), pp 52–4.

[20] Abram Chayes and Antonia H. Chayes, 'On Compliance', *International Organization*, 47 (1993), pp. 177–88.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

The Facilitative Branch is 'responsible for providing advice and facilitation to Parties in implementing the Protocol, and for promoting compliance by Parties with their commitments under the Protocol'.[21] It shall provide for early warning of potential non-compliance, formulate recommendations, and facilitate financial and technical assistance, including technology transfer and capacity building from sources other than those established under the Convention and the Protocol for the developing countries. Finally, it is specifically responsible for reviewing the Parties' reports showing that their use of the flexibility mechanisms supplements domestic action.

The Enforcement Branch is responsible for determining whether Annex I countries comply with their Kyoto commitments, and for pushing non-compliant countries to return to compliance. The enforcement 'consequences' to be applied when countries fail to meet their obligations, and the procedures that will determine the consequences and apply them, are the punitive (or 'hard') elements of Kyoto's compliance regime.[22]

If a country fails to meet its assigned emissions target, the Marrakesh Accords outline three means to enforce compliance. First, the non-compliant country must present a plan demonstrating how it intends to restore compliance with the Protocol. Second, in the next commitment period it must cover its deficit, *plus* another 30 per cent of that deficit, in addition to whatever its commitment would be for that period. Finally, it loses its eligibility to make transfers (sell emission permits) under the emissions trading provisions of the Protocol.[23]

If the eligibility requirements for participating in the Kyoto flexibility mechanisms are not fulfilled, the Marakesh Accords state that the eligibility of the non-compliant country shall be suspended in accordance with relevant provisions under articles 6, 12, and 17.[24]

According to Article 18 of the Kyoto Protocol, any procedures and mechanisms entailing binding consequences must be adopted by means of an amendment to the Protocol. At the first meeting of the Conference of the Parties acting as the Meeting of the Parties (the COP/MOP) in December 2005, the decision whether to make the consequences legally binding was deferred to COP/MOP 3 in 2007.

3. Types of non-compliance

There are two main types of non-compliance. *Unintentional* non-compliance is caused by incapacity;[25] therefore, scholars concerned with unintentional non-compliance tend to emphasise that non-compliance is sometimes unavoidable.[26] At least two

[21] UNFCCC, 'The Marrakesh Accords and the Marrakesh Declaration', p. 131.
[22] Geir Ulfstein and Jacob Werksman, 'The Kyoto Compliance System: Towards Hard Enforcement', in Olav S. Stokke, Jon Hovi and Geir Ulfstein (eds.), *International Compliance: Implementing the Climate Regime* (London: Earthscan, 2005), p. 40.
[23] UNFCCC, 'The Marrakesh Accords and the Marrakesh Declaration', p. 139.
[24] Ibid., p. 139. For a more detailed presentation of the compliance system, see, for example, Ulfstein and Werksman, 'The Kyoto Compliance System: Towards Hard Enforcement', pp. 41–9.
[25] Note, however, that not all non-compliance related to incapacity is unintentional. See Section 3.2 for an elaboration of this point.
[26] Chayes and Chayes, 'On Compliance', pp. 187–8.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

types of incapacity may be distinguished.[27] *Financial* incapacity means that a country has inadequate economic resources to ensure compliance. *Administrative* incapacity refers to inadequate bureaucratic resources to issue rules and regulations, and to monitor their enforcement. *Administrative* incapacity tends to be greatest in countries that lack competent and powerful governmental bureaucracies.

In contrast, *intentional* non-compliance is caused by 'free-riding'; deliberate attempts to escape one's share of the costs of reducing emissions. Global warming will affect countries asymmetrically, and most, if not all, will benefit to some degree from reductions of GHG concentrations in the atmosphere. The more countries that contribute to reducing GHG emissions, the greater the benefits will be. However, a single country's efforts will only have a limited impact on global emissions. The exact impact will depend on the size and composition of the country's economy. Furthermore, while the benefits from reductions will be shared by all countries in a non-exclusive and non-rival manner, each country will have to bear its own, possibly high, costs of abatement. Free-riding thus enables a country to avoid costly abatement, with only minor detrimental effects on global environmental quality. According to the theory of public goods, the result may be under-provision of abatement. An extreme scenario is that no country will act to reduce emissions, but a more likely scenario is that only a few wealthy countries act, resulting in sub-optimal provision.[28]

There are two types of incentives for free-riding: the incentive for a country not to join the agreement since it would benefit nevertheless from the signatories' abatement efforts (non-participation), and the incentive for a signatory to violate its commitments under the agreement (non-compliance).[29] Creating incentives to ensure participation and compliance is a major challenge for the parties in the international climate negotiations.[30]

Overcoming the public goods problem of under-provision[31] is difficult because no supranational body exists to enforce agreements.[32] Nor is it realistic to expect nations to transfer enforcement authority to such a body in the foreseeable future. Furthermore, even if limited (or greater) policy-making authority were to be delegated to such international bodies, domestic legislation would remain the method for implementing international environmental policies.

The difference between free-riding and incapacity is not crystal clear. For example, because (intentional) non-compliance caused by free-riding is less likely to be

[27] Abram Chayes, Antonia H. Chayes and Ronald B. Mitchell, 'Managing Compliance: A Comparative Perspective', pp. 52–4; David Vogel and Timothy Kessler, 'How Compliance Happens and Doesn't Happen Domestically', in Weiss and Jacobson (eds.), *Engaging Countries*, pp. 20–3.

[28] Todd Sandler, 'Global and Regional Public Goods: A Prognosis for Collective Action', *Fiscal Studies*, 19 (1998), p. 222.

[29] Michael Finus, *Game Theory and International Environmental Cooperation* (Cheltenham: E. Elgar, 2001), ch. 2.

[30] Barrett and Stavins, 'Increasing Participation and Compliance in International Climate Change Agreements', p. 349.

[31] For a discussion of public goods provision, see for example, Dennis C. Mueller, *Public Choice III* (Cambridge: Cambridge University Press, 2003).

[32] We follow the commonly held view that mitigation of climate change is a pure public good. An alternative view is that global atmospheric problems are better considered as common-pool resource problems (for example, see J. Samuel Barkin and George E. Shambaugh (eds.), *Anarchy and Environment* (Albany, NY: State University of New York Press, 1999).

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

tolerated than (unintentional) non-compliance caused by incapacity, it might be tempting to under-invest in increased administrative capacity for compliance in order to camouflage free-riding as incapacity. A country could then claim to merit a disproportionately low emissions reduction target because of financial or administrative constraints,[33] or could use the same ruse for not signing (or ratifying) an agreement,[34] or for not complying after signing and ratifying an agreement.[35] Due to asymmetric information, determining the legitimacy of such claims could prove extremely difficult. In turn, uncertainty as to whether a country that exceeded its assigned emissions target did so intentionally could have a bearing on whether the Enforcement Branch determines the country to be in compliance or non-compliance. Importantly, however, the Marrakesh Accords make no distinction between intentional and unintentional non-compliance.

4. Deterring non-compliance versus restoring compliance

Suppose that a country fails to fulfil its Kyoto obligations for the first commitment period, and that the Enforcement Branch applies punitive consequences as specified by the Marrakesh Accords. Under what conditions would these punitive consequences cause the country to return to compliance in the second commitment period? We proceed in four steps to answer this question. In step one, we briefly remind the reader of the conditions under which punitive consequences effectively *deter* non-compliance. Because non-compliance takes place only when deterrence fails, the next three steps of our argument consider what happens when the conditions for effective deterrence are not met. In step two, we demonstrate that when complete information exists the application of punitive consequences *cannot* induce a non-compliant country to return to compliance. In step three, we argue that for the application of the punitive consequences to cause a return to compliance, a country must initially either underestimate the determination of the Enforcement Branch to apply punitive consequences, or underestimate the impact of the applied consequences. Finally, in step four, we discuss whether the requirements for punitive consequences to cause a return to compliance are likely to be fulfilled under Kyoto's compliance system. Throughout our discussion of these four steps in this section we assume that the expert review teams will be able to monitor emissions inventories perfectly. In the next section we consider the implications of relaxing this assumption.

[33] For example, during the climate negotiations Norway emphasised its dependency on the petroleum sector, as well as high abatement costs due to its traditional reliance on hydro-power. The result was a Kyoto target that allows a 1 per cent increase of GHG emissions in 2012, compared to the baseline in 1990. This could be considered a relatively easy target compared to the targets of most other Annex I countries, particularly bearing in mind Norway's relative wealth.

[34] For example, Russia was for a long time reluctant to ratify the Kyoto Protocol, officially on the grounds that they expected their future economic growth to entail very high compliance costs (especially after the United States declined to ratify Kyoto).

[35] For example, a country might claim that economically vital domestic stake-holders exert pressure (for instance, by threatening to relocate businesses abroad) that makes it impossible for the country to abide by its commitments.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

4.1. Conditions for effective deterrence of non-compliance

To deter non-compliance punitive consequences must satisfy three conditions.[36] First, the consequences must be *relevant*, in the sense that a country needs sufficient technical, financial and political capacity to comply. If it lacks the capacity to comply, then there is obviously no way in which deterrence can succeed (at least not in the short term). Second, the punitive consequences must be *sufficiently potent* such that a country would prefer to comply, rather than to suffer the consequences of non-compliance. Obviously, a country would almost certainly prefer to avoid punitive consequences. But however much a country might regret the application of such consequences, such regret will not motivate it to comply if doing so entails higher costs than those caused by non-compliance. Finally, the threat of punitive consequences must be *credible*, that is, a country must believe that failure to comply will cause the Enforcement Branch to apply punitive consequences and that those consequences will be maintained until it returns to compliance. If the threat of the punitive consequences is credible, and the consequences themselves are relevant and potent, this threat will effectively deter non-compliance. Thus, if these three conditions are fulfilled for all Annex I countries, *there will not be any non-compliance* (either intentional or unintentional). Rather, when the three conditions are fulfilled the compliance system will effectively induce all Annex I countries to meet their emission targets and other obligations under the Kyoto Protocol.

Next, assuming that an Annex I country has *already* defied the threat of punitive consequences, under what conditions could punitive consequences cause this non-compliant country to *return* to compliance? Obviously in such a case, it must also be assumed that at least one of the three conditions for effective deterrence has failed; therefore, we need to address the following question: Under what circumstances might a country (1) violate its obligations in the first commitment period, thereby defying the *threat* of punitive consequences, and yet (2) return to compliance *because* those consequences are applied in the second commitment period?

4.2. Failure of deterrence and restoration of compliance: complete information

If the threat of punitive consequences fails to deter non-compliance, then what happens next will depend on *why* deterrence failed. It follows from the discussion in the previous section that deterrence might fail because the threat of punitive consequences is not credible, or because the consequences themselves are irrelevant or insufficiently potent. Should we expect *applied* punitive consequences to work in any of these cases?

Note first that if a country is *incapable* of compliance, then neither threatened nor applied consequences have any chance of deterring non-compliance (assuming that the country's capacity for compliance cannot be improved). In other words, in cases in which non-compliance with the Kyoto Protocol turns out to be caused by capacity problems which render the threat of punitive consequences irrelevant, the punitive

[36] For a substantiation of this claim, see Jon Hovi, *Games, Threats and Treaties. Understanding Commitments in International Relations* (London: Pinter, 1998), ch. 2.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

consequences *cannot* have any bearing on the parties' behaviour. The tools available to the Facilitation Branch – advice and technical assistance – will then be the only means of restoring compliance. However, while capacity problems are mostly relevant for developing countries, the punitive consequences under Kyoto's compliance system apply only to industrialised countries (Annex I countries). In the remainder of this article we therefore concentrate on cases in which deterrence of non-compliance fails – either because the threat of punitive consequences is not credible or because the punitive consequences themselves are not sufficiently potent. For the moment, we assume that both the non-compliant country and the members of the Enforcement Branch are able to assess all relevant aspects of the situation correctly – that both have complete information. (This assumption will be relaxed shortly.)

First, consider a case in which the threat of punitive consequences fails because it is an empty threat. Empty threats are just that – empty. Thus, in such cases, non-compliance would not cause punitive consequences to be applied. Obviously, consequences that are *known never to be applied* cannot induce compliance (or deter non-compliance). It follows that in such cases deterrence would fail, and countries would enter into non-compliance. Yet, no punitive consequences would be applied, and the compliance system would fail to restore compliance.

Next, consider a case in which a *credible* threat fails because a country considers the cost of compliance to outweigh the cost of suffering punitive consequences. In such a case, non-compliance *would* cause the punitive consequences to be applied. Nevertheless, we should not expect a party that fails to stand by its commitments in the first commitment period to return to compliance in the second commitment period. If the threat to apply punitive consequences were credible, a country would *correctly conclude* that non-compliance would cause the Enforcement Branch to apply them. Thus, one would have to assume that the calculations underlying its decision not to comply in the first commitment period would take into consideration the cost of punitive consequences in the second commitment period. If this cost were seen as intolerable, it would be in the best interest of a country to meet its obligations in the first commitment period. Hence it would enter into non-compliance only if the cost of compliance were to outweigh the cost of the punitive consequences. If all the costs and benefits of non-compliance were common knowledge, the actual application of the punitive consequences would not change a country's assessment of such costs and benefits. Thus, assuming that all parties have complete information, if the threat of punitive consequences were to fail to *deter* a country from non-compliance, then their actual application would not cause it to *return* to compliance.

4.3. Failure of deterrence and restoration of compliance: incomplete information

Two preliminary conclusions emerge from the discussion thus far. First, deliberate non-compliance would occur only if a country were to believe either that the threat of punitive consequences were empty or that the consequences themselves were insufficiently potent. Second, assuming that all parties have complete information, then one of two scenarios would materialise:

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

- The threat of punitive consequences would effectively deter non-compliance; hence, there would be no need to actually apply them; or
- Deterrence would fail, but either the punitive consequences would not be applied because the threat would have been empty, or they would be applied but would fail to restore compliance because of their insufficient potency.

This means that, assuming all parties have complete information, *applied* punitive consequences *cannot* work.

The other side of the coin is that for applied punitive consequences to restore compliance there must be *in*complete information. Specifically, there are two conditions under which applied punitive consequences could cause a return to compliance. First, the application of punitive consequences must be unanticipated by the country contemplating non-compliance. If this is the case, a country might enter into non-compliance expecting that no punitive consequences will be applied, and return to compliance if, contrary to its expectations, punitive consequences *are* applied.

The second condition under which applied consequences can bring about compliance is that a country initially underestimates the cost of the (anticipated) punitive consequences. It might then choose non-compliance, expecting that the punitive consequences will entail only moderate costs, but return to compliance if the applied consequences prove harsher than expected.[37] More precisely, a country might return to compliance if it comes to believe that the cost of continuing to suffer the punitive consequences outweighs the cost of compliance.

The question we address in the next section is whether either of these two conditions is likely to be fulfilled in the international climate regime.

4.4. Restoring compliance in the climate regime

4.4.1. Threats that are erroneously believed to be empty. Consider a country that fails to fulfil its obligations for the first commitment period because it (erroneously) believes the threat of punitive consequences to be empty. In other words it believes that, even if it does not live up to its Kyoto obligations, the Enforcement Branch will either (1) not determine that it is in non-compliance, or (2) determine that it is in non-compliance, but fail to apply the punitive consequences. In such a case, a country might be induced to reconsider if the Enforcement Branch proves its resolve by determining that it is in non-compliance *and* imposing the punitive consequences. Clearly, for this to happen, the punitive consequences must be sufficiently potent. Therefore, in this case the *unexpected* application of (potent) punitive consequences act as the key factor that induces a country to return to compliance.

Is it reasonable to expect that at least some countries might consider the punitive consequences available to the Enforcement Branch to be mere paper tigers, or consider the Enforcement Branch itself to be a paper tiger? The Marrakesh Accords

[37] We disregard the logical possibility that these factors are correctly estimated at the threat stage, and that – after sanctions have been applied – the target is deceived into erroneously *over*estimating (1) the harshness of the consequences, or (2) the resolve of the Enforcement Branch.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

emphasise two provisions: that the members of the Enforcement Branch shall serve in their personal capacities and that they shall have legal expertise.[38] The motivation for these provisions is a desire to minimise the impact of political considerations in the compliance system. Moreover, the consequences of non-compliance are fully *predetermined*.[39] The Enforcement Branch shall only ascertain whether the Party is in non-compliance (either for failing to meet its emissions target or for failing to fulfil the preconditions for eligibility to utilise the Kyoto flexibility mechanisms). It has no discretion to choose which punitive consequences are to be applied. Finally, the punitive consequences are also *automatic*. This means that once it determines a country to be in non-compliance, the Enforcement Branch is 'responsible for applying the consequences'.[40] The decision of the Enforcement Branch can be appealed to the COP/MOP, but this possibility is very limited.[41] All of this implies that the threat to apply punitive consequences is credible.

One might object, as Ulfstein and Werksman remind us, that determining whether a country is non-compliant requires considerable judgment and discretion.[42] Also, regardless of how clearly or well procedures are written, it is difficult to completely avoid political jockeying in determining whether a country is non-compliant. Thus, one cannot entirely rule out the possibility that the Enforcement Branch's final decision might be affected if the votes of a sufficient number of members are influenced by their respective national self-interests or other political considerations. Moreover, it is in the best interests of most Annex I countries if the Enforcement Branch does *not* determine a large seller of permits (such as Russia) to be in non-compliance. If Russia is punished, the price of emissions permits will increase significantly for all buyers of permits. Since the vast majority of Annex I countries will likely be buyers of emissions permits, one might reasonably question whether the Enforcement Branch is really prepared to determine Russia to be in non-compliance.

How significant is this objection? Recall that, for the punitive consequences to cause a *return* to compliance, it is not enough that countries believe the threat of punitive consequences to be empty. That belief must prove to be *erroneous*. Over time, the Parties to the Kyoto Protocol will probably learn the circumstances (if any) under which the threatened consequences are likely to be applied. Thus, while it is not inconceivable that a country might underestimate the willingness of the Enforcement Branch to apply sanctions, such a case is likely to be increasingly rare once the compliance system has been in operation for some time. As countries gain experience with the system, they will be better able to correctly assess the credibility of the threat of punitive consequences. Greater experience will, of course, cause this credibility to be strengthened or weakened, depending on the actions of the Enforcement Branch. Consequently – and again depending on the Enforcement Branch's actions – the overall number of cases of non-compliance may go up or down. In any event, because of experience, non-compliance caused by an *erroneous* belief that the threat of punitive consequences is empty should quickly become rare.

[38] Ulfstein and Werksman, 'The Kyoto Compliance System: Towards Hard Enforcement', p. 47.
[39] Jutta Brunnée, 'The Kyoto Protocol: Testing Ground for Compliance Theories?' *Heidelberg Journal of International Law*, 63 (2003), p. 264.
[40] See Marrakesh Accords, Decision 24/CP.7, Annex, Section V(6).
[41] Ibid., Section XI.
[42] Ulfstein and Werksman, 'The Kyoto Compliance System: Towards Hard Enforcement', pp. 51–2.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

4.4.2. Underestimating the potency of the consequences. Next, consider the case of a country that ignores the threat of punitive consequences because it underestimates their potency. In this case, the country anticipates that non-compliance will cause the Enforcement Branch to apply punitive consequences. Yet it deliberately enters into non-compliance because it expects that the consequences will be less costly than compliance. However, after the punitive consequences have been applied, the country might update its estimate of their true costs and come to consider them intolerable. Assuming that the country expects the consequences to remain in place until it returns to compliance, it is possible that it will prefer to comply rather than to continue paying the costs of the punitive consequences.

However, scenarios of this kind are unlikely to occur very often in the climate regime for at least three reasons. First, recall that the compliance system of the Marrakesh Accords suffers from several weaknesses which are likely to limit the costs of punitive consequences. For example, reducing emissions further (with a penalty rate for the deficit – see above) will be mandatory only for countries that ratify an amendment which makes the compliance system legally binding. And even if the compliance system becomes binding there is nothing preventing a country from postponing implementation of the punishment for one or even several periods. Also, the costs of the consequences in a given commitment period could be moderate if a country successfully holds out for a generous emissions target for that period. Of course, the fact that the cost of the consequences is likely to be limited does not completely rule out the possibility that it might be underestimated. However, the less potent (costly) the consequences are, the smaller the extent to which they can be underestimated. Thus, there will likely be few (if any) cases in which a country *significantly* underestimates the potency of the punitive consequences.

Second, countries with limited financial and administrative capacities are more likely to miscalculate the cost of punitive consequences. Countries with ample financial and administrative capacity should be able to gauge the cost of punitive consequences reasonably well. Because punitive consequences apply only to Annex I countries, countries *with* such capacity, it is thus unlikely that (serious) miscalculation will be a frequent phenomenon in the climate regime.

Finally, punitive consequences entail real costs for a country only if it chooses to comply with the punishment. Ultimately, the country being punished (partly) controls implementation of the consequences (see Section 1). Since (by assumption) we are dealing with *intentional* non-compliance, we are considering a country that has *deliberately* violated its legally binding commitment for the first commitment period. Hence, whenever punitive consequences are applied in such cases, the country will have already defied the threat of such consequences by intentionally violating its Kyoto commitments. Given this circumstance, it is not unreasonable to question the country's willingness to implement the punitive consequences as well.[43] Indeed, a country that considers unbearable *both* the cost of compliance *and* the cost of punitive consequences might simply choose to withdraw from the treaty.

4.4.3. External pressure. So far, we have demonstrated that the application of punitive consequences is unlikely to cause a country to return to compliance.

[43] Brunnée, 'The Kyoto Protocol: Testing Ground for Compliance Theories?', p. 263.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

However, a country that is determined by the Enforcement Branch to be in non-compliance might well experience other types of pressure also – both from other countries and from non-governmental organisations (NGOs).[44] Some observers have even suggested that such external pressure could be more important than the punitive consequences of the climate regime. For example, at the 2001 climate meeting in Marrakesh, Canada's top negotiator said, 'The matter [of a legally binding compliance system for the Kyoto Protocol] is largely symbolic. Even if countries don't face sanctions, countries that signed on to the accord but did not abide by its rules would face pressure from other signatories.'[45]

If such external pressure to return to compliance were taken into account, would we need to modify the above conclusions? The answer is no, because the logic for punitive consequences outlined in this article holds for external pressure as well. External pressure would no doubt make non-compliance more costly than it would otherwise be. However, *anticipation* of external pressure would simply add to the deterrent effect of punitive consequences, thereby serving to further discourage a country from non-compliance. In contrast, if external pressure comes *unexpectedly*, it might cause a country to return to compliance, provided that the combined effect of punitive consequences and external pressure is sufficiently potent. Similarly, if a country foresees that violating its commitments would entail external pressure plus punitive consequences, and yet decides to enter into non-compliance, it is likely to *return* to compliance only if the combined effect of punitive consequences and external pressure is greater than it expected. In short, external pressure would cause – exactly as the punitive consequences would cause – a return to compliance only if this pressure were (1) unanticipated by the noncompliant country, or (2) more potent than it expected.

However, there is at least one important difference between punitive consequences applied by an international regime and external pressure. As we have already mentioned, the functioning of the compliance system is likely to become fairly predictable after it has been in place for some time. However, predicting correctly the likelihood and potency of external pressure could be more difficult. The potency of external pressure depends on the degree of attention which NGOs and the media give to the non-compliance. If this attention were to vary both over time and across countries, then underestimation of the external pressure's potency might occur more often than underestimation of the punitive consequences' potency.[46] We thus predict that in the climate regime, external pressure will restore compliance more often than punitive consequences will.

5. Imperfect monitoring of inventories and restoration of compliance

The discussion in Section 4 assumed perfect monitoring of emissions inventories, an assumption which is no doubt unrealistic in the context of the climate regime.

[44] For discussions of the potential effect of external pressure from other states and green NGOs in promoting compliance with the climate regime, see Stokke et al. (eds.), *Implementing the Climate Regime*, chs. 6–8.

[45] WSJ.com 27.11.2001. ⟨http://interactive.wsj.com/fr/emailthis/retrieve.cgi?id=ID-CO-20011103-000334.djml⟩.

[46] One should also remember that many Annex I countries are relatively small democratic countries. Such countries might be more susceptible to pressure than (say) the United States, the world's leading power, or a large undemocratic country like China.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

However, assume that monitoring is imperfect, which might make it possible for a country to get away with creative accounting or even with plain lies about inventories. How does this assumption affect the conclusions reached above?

In several judicial and compliance systems, a non-compliant subject will, when exposed, face punitive consequences more or less automatically. If a subject, knowing that monitoring is imperfect, takes its chances, violates a rule, and gets caught, then automatically applied punitive consequences might cause this subject to return to compliance, provided that the consequences are sufficiently potent.[47]

However, the climate regime is different, because a country found to be using creative accounting or to be lying about inventories will *not* automatically face punitive consequences. For a period of 100 days after the completion of the expert review of the commitment period's final year (often called the 'true up' period), a country may continue to acquire assigned amount units (AAUs), emission reduction units (ERUs), and certified emission reductions (CERs) from the just-expired commitment period.[48] Hence, even if a country gets caught violating its commitment, it can nevertheless return to compliance – and avoid punitive consequences – if it buys enough permits during the true-up period. Once caught, a non-compliant country will do this, if it considers the threat of punitive consequences to be credible and the consequences themselves to be sufficiently potent.[49] Conversely, a non-compliant country will decline to buy permits in the true-up period if it (1) believes that it can somehow avoid punitive consequences or (2) believes that the cost of the consequences is outweighed by the cost of doing what is necessary to be in compliance.

We are concerned with determining the circumstances under which *applied* punitive consequences can restore compliance. As we have demonstrated in this section, with imperfect monitoring there is an additional sense in which punitive consequences can restore compliance. However, if a country returns to compliance in the true-up period, punitive consequences will *not* be applied. Indeed, punitive consequences will be applied only if a country continues to be in non-compliance at the end of the true-up period. Hence, the conclusion that *applied* punitive consequences will rarely cause a return to compliance still stands.

6. Is the inability to restore compliance a weakness of the compliance system?

We have argued that in the climate regime, *applied* punitive consequences will restore compliance only in exceptional cases. Should one infer that this is a weakness of the compliance system – something that ought to be corrected? To enable applied punitive consequences to restore compliance, one would need to *create* uncertainty about either their application or their cost to the non-compliant country, or both. For example, one could intentionally deceive member countries to erroneously

[47] A similar logic would apply if the violator bases its actions on expected-utility calculations using subjective probabilities that turn out to be incorrect.
[48] See the non-compliance procedure, XIII.
[49] Similarly, a country which *inadvertently* (or unintentionally) is in violation of its assigned emissions target at the beginning of the true-up period would buy enough permits to get into compliance by the end of the true-up period.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press

believe that non-compliance would not cause the application of punitive consequences, or one could misinform member countries so that they would underestimate the cost of the consequences. If successful, such deceitful manoeuvres would improve the likelihood that applied consequences would restore compliance. However, such manoeuvres would also undermine the ability of the consequences to *deter* non-compliance and thus might cause more countries to violate their commitments. For environmental, economic and administrative reasons, effective deterrence is preferable to restoring compliance *after* deterrence has failed. Therefore, policymakers should aim for a compliance system that achieves effective deterrence. Hence, although the analysis in this article helps one to understand how the compliance system of the climate regime might work, it does not provide a rationale for reform. On the contrary, it suggests that the Marrakesh Accords' emphasis on predetermined and automatic consequences makes good sense.

Conclusions

While previous research on the Kyoto compliance system has primarily focused on the compliance system's ability to deter non-compliance, this article has discussed whether *application* of punitive consequences could induce a non-compliant country to *return* to compliance. We have argued that the threat of punitive consequences might push a country to return to compliance during the true-up period. However, if this happens, the consequences will *not* be applied. In contrast, *applied* punitive consequences can restore compliance only under exceptional circumstances: *Either* their application must be unanticipated by the non-compliant country, *or* the consequences themselves must prove harsher than the non-compliant country expected. The first condition is unlikely to be fulfilled under Kyoto's compliance system because of its strong emphasis on the automatic application of predetermined consequences once a case of non-compliance has been established. The second condition is also unlikely to be fulfilled because the cost of suffering the consequences will be limited, and because the compliance system applies only to wealthy countries that should be able, due to their financial and administrative capacities, to estimate this cost reasonably well. Also, a non-compliant country can *solve* the problem of unexpectedly harsh consequences by withdrawing from the Kyoto Protocol. It is thus likely that only in exceptional circumstances will *applied* punitive consequences cause a non-compliant country to return to compliance. In fact, external pressure might be able to accomplish a return to compliance more often than punitive consequences do.

https://doi.org/10.1017/S0260210507007590 Published online by Cambridge University Press