DRAFT. Katie Steele

‘International Paretianism’ and the question of ‘feasible’ climate solutions

forthcoming in *Philosophy and Climate Change*, Budolfson, M., McPherson, T. and Plunkett, D (eds.) Oxford University Press.

Proponents of *International Paretianism (IP)*—the principle that international agreements should not make any state worse-off and should make some at least better off—argue that it is the only feasible approach to reducing the harms of climate change (see, especially, Posner and Weisbach 2010). They draw on some key assumptions regarding the meaning of ‘feasibility’ and the nature of the Pareto improvements associated with coordinated action on climate change. This chapter challenges these assumptions, in effect weakening the case for IP and allowing for broader thinking about what counts as a ‘feasible’ climate solution.

1 Introduction

In view of the considerable suffering that climate change threatens in the medium- and long-term future, the global response to this problem has been woefully inadequate. This very concerning state of affairs has lead many to reflect on the nature of the problem and the prospects for solving it. Accordingly, the notion of *political feasibility* has become prominent in debate about who can be counted on to bear the costs of climate change—the costs of mitigation and adaptation. While it is all well and good arguing over what are better and worse distributions of these costs, actual progress can only be achieved through action, which is constrained by what is feasible. Or so the rhetoric might go. Indeed, it is hard to deny that, ultimately, it is important that we actually advance on the status quo—that climate-related suffering is reduced by *some amount*, the more the better, even if this falls far short of a morally good state of affairs.

A specific principle or maxim has been pitched as the answer to the call for ‘feasible’ climate solutions—*International Paretianism* (IP)—which basically recommends framing international negotiations on problems like climate change in terms of mutual ‘national interest’. The idea is that, as far as feasibility goes, any proposal for determining responsibilities for tackling climate change must advance the interests of

all states, or at least not frustrate their interests. That is, a genuine climate solution must provide a *Pareto Improvement* with respect to the interests of states: No state is made 'worse-off', and at least one state is 'better off' than they fare under the status quo. One might question whether states really are the key international players in climate negotiations, but let us assume that this feature of international IP is plausible. It is also ambiguous as to what counts as a state's 'self or national interest', but let us assume it is an appropriate aggregate of the welfare of present and future citizens (more about this later). As such, the costs of a mitigation strategy for present citizens can be counterbalanced by the benefits for future citizens, such that there is a net benefit in terms of national interest.

Advocates argue that IP is uniquely placed with respect to political feasibility in the international setting (see, especially, Eric Posner and David Weisbach's 2010 book *Climate Justice*). They claim that the only feasible proposals for significantly reducing the harms of climate change are IP proposals, and thus treaties intended to actually reduce suffering associated with climate change ought to satisfy IP criteria. While IP is not a radically new approach to international agreements, it is nonetheless provocative in climate justice circles.[1] Discussions of climate justice have traditionally focussed rather on historical responsibility and duties of reparation, or simply duties of assistance, that are owed by rich, high-emitting states.[2] IP, on the other hand, redirects attention to what is good for rich, high-emitting states (as well as what is good for other states). Thus IP, as an approach to climate change, may seem rather shocking. In spite of this, there is certainly something compelling, hopeful even, about IP, once we focus on political feasibility. IP is not about states recognising their duties and stepping up to make whatever demanding reparations for climate change are morally required of them. It is rather about promoting a climate solution that the relevant actors would have no reason to resist and may have material

[1] Note, however, that some philosophers have also claimed that it may be productive, at least in the short term, to focus on climate solutions that amount to Pareto improvements on the status quo; notably Broome (2010); see also Schokkaert and Eyckmans (1998). The focus here is on the work of Posner and Weisbach because they make stronger, more definitive claims about feasibility in the international setting. Moreover, Broome takes individual persons, rather than nation states, to be the basic agents that are subject to climate-related costs and benefits, and so does not defend International Paretianism *per se*.

[2] Typically, the focus is on the *costs* of mitigation, assuming that an appropriate global mitigation target has been decided. The question is how to divide up the costs of meeting this target; a variety of moral perspectives lead to the conclusion that the rich high-emitting states should bear the bulk of the costs.

interest in pursuing. We would be fools to let an overly aspirational treaty prevent easy agreement on a morally inferior, but nonetheless mutually beneficial, one. Or so the thinking might go.

Political feasibility is certainly an important consideration in the debate about climate change. But once we look to the closer details of what feasibility stands for, its relationship with furthering self-interest, let alone the national interest of a state, is not as persuasive as first appears. In order to arrive at a substantial global response to climate change, IP must appeal to a notion of national interest that is rather revisionary given the existing apparent motivations of state actors. Furthermore, closer inspection of the form of the cooperative dilemma associated with climate change reveals that the pursuit of national interest alone may turn out to be self-defeating. This is more worrisome for the feasibility argument for IP than its advocates suggest. Or so this chapter will argue. More generally, the aim here is to shed light on International Paretianism, the climate change predicament, and the meaning and role of 'feasibility' in ethical-political debate. Section 2 begins with the latter conceptual issue. Section 3 turns to the climate predicament itself—the kind of Pareto improvement that may be in the offing, and explores whether the achievement of such a Pareto improvement does in fact depend solely on 'self interest'. Section 4 reflects on the implications of these considerations for the overall feasibility claims made by IP proponents.

2 The concept of 'Political Feasibility'

As suggested above, Posner and Weisbach (2010) largely rest their case for an IP-style climate treaty on the fact that only this style of climate treaty is 'politically feasible'.[3] But what is the precise meaning and significance of that term? As

[3] A qualification must be made here: Posner and Weisbach (2010) also argue that a climate treaty should be concerned only with advancing mutual self-interest because anything more ambitious in terms of rich states paying would amount to trying to solve other problems of global distributive justice in a climate treaty, which would not be an *efficient* way to solve these other important problems. So they argue for IP on efficiency grounds (in the broader scheme of distributive justice) as well as feasibility grounds. It is not at all obvious that the optimal IP climate solution would in fact yield the amount of climate change mitigation that a benevolent global planner would choose, which is an aspect of the efficiency argument. Later in Section 3.1 some further remarks are made on this point. But it is beyond the scope of this chapter to discuss the efficiency argument in detail. See, however, Frisch

suggested above, the reason to focus on feasibility at all stems from an interest in actual social change, in 'getting things done'. So 'feasibility' is a property that is invoked in the interests of actual change. Beyond that, however, the proper usage of the term is not obvious. And yet the plausibility of the IP proponents' claim – the tight relationship between feasibility and Pareto improvements – depends on these details.

An initial concern is that the concept is inherently vague and so the IP proponents' claim will resist any definitive analysis. The very target of the concept – what supposedly has the property of 'feasibility' – seems to vary in general talk and debate, let alone the rules for determining whether/to what extent the property is present. We talk about whether a goal, say, of swimming 50m in less than half a minute, is feasible; about whether change, say, towards a less racist society, is feasible; about whether a program, say, a new teaching curriculum, will feasibly improve students' learning outcomes; about whether it is feasible that I will not procrastinate while doing some task. Some of these examples seem to be about whether actions will realistically be performed (e.g., work without procrastination), and others about whether goals will realistically be achieved (e.g., the fast swim or a less racist society) perhaps in addition to or else conditional on some action being performed (e.g., the improvement of students' learning outcomes given reforms to the teaching curriculum).

The further concern is that there are apparently no set criteria for determining whether the intended target – a goal or action or both – counts as feasible, i.e., how plausible its realisation must be and what is the basis for such an assessment. Perhaps we should accept that 'feasibility' is simply a smokescreen for differing substantial views about what are worthwhile pursuits. In the political context, the question might boil down to old disputes about what are viable prospects for social change. If IP proponents were simply staking a 'realist' position regarding international relations, say, then here again their claims would seem to resist any searching analysis. Quite simply: those sympathetic to the 'realist' view (roughly that 'might is right') in the international arena will be sympathetic to the IP view, unlike those of more 'idealist' persuasion.

(2012), Baer (2013), Jamieson (2013) and Shue (2013), for critiques of Posner and Weisbach's claims regarding the ethical optimality, so to speak, of IP.

There is something to the above concerns, but they are a bit quick. We will see that analysing the IP proponents' claims about feasibility and climate change is more interesting and fruitful than first meets the eye. To begin with, the growing philosophical literature on political feasibility suggests that the concept has subtleties revealing a need to clarify what legitimate role(s) it may play in debate. This will become apparent in the attempt to pin down a working definition of the term in 2.1 below. Note that while there is some divergence amongst the prominent definitions in the literature (and moreover some suggest that 'feasibility' may play multiple normative roles and so accordingly have multiple meanings) there are nonetheless a number of points of agreement amongst feasibility scholars. These points of agreement will be taken as 'fixed points' in our analysis. Beyond that, the strategy here is to fashion a notion of feasibility that is, as it were, charitable to the IP proponents' position. It should be a useful concept for evaluating climate change proposals and flexible enough to *potentially* vindicate the IP proponents' claim regarding a strong relationship between feasibility and Pareto improvements with respect to climate change. This more substantial issue is introduced in 2.2 and considered in more detail in the remainder of the chapter.

2.1 A working definition of 'political feasibility'

Despite there being a range of views about the finer details of the concept of 'political feasibility', there is reasonable agreement about the basics. In particular, there is broad agreement (also in line with the discussion thus far) on the general function of assessments of feasibility: They bear on directives for action and, roughly speaking, concern the possibility of success. In the spirit of 'ought implies can', what an agent *ought to do* is *somehow sensitive* to what is feasible (see Brennan and Southwood 2007, Gilabert and Lawford-Smith 2012, Lawford-Smith 2013, Southwood 2016, Wiens 2015).[4] Note that this accords with the examples given above of how the term is variously employed in debate. All these examples relate to directives for action,

[4] There are many further subtleties concerning this phrase that I will not explicitly address in this discussion - concerning the way and extent to which 'what one ought to do is sensitive to what is feasible'. Southwood (2016), for instance, discusses different notions of 'ought' that have differing logical relationships with his stated notion of 'feasibility'. The point here is simply that the notion of feasibility should be defined in such a way that it plays *some* guiding role in determining right action.

although, as noted, the precise nature of the target seems to vary – the feasibility assessment applies sometimes to an act, sometimes to a goal or state of affairs to be acted upon, and sometimes to a combination of the two. (We will return to this diversity shortly.)

Given the function just described, it follows that feasibility assessments pertain to a given agential positioning or decision-making context (the more precisely specified the more precise will be the feasibility assessment). All agree that feasibility has to do with whether the target may be plausibly realised, but, as mentioned, there is disagreement about the target. Perhaps the most natural target is an action that an agent might consider performing. This is the line taken by Brennan and Southwood (2007) and Southwood (2016). We can interpret these authors as holding that the 'feasible acts' are akin to what decision theorists might refer to as the 'available acts' for some decision-maker, or in other words, the 'choice set'. It follows that feasibility, understood in this way, is a binary matter: Candidate acts are either feasible (i.e., in the choice set), or not, depending on worldly constraints and whether the agent can 'bring him/herself to perform the act in question'[5]. The motivational constraint here is that the agent must be able to decide, if it follows from their psychological attitudes, upon the act in question, to initiate the act if it is chosen, and also to follow through in performing the act.[6] So the agent must be sufficiently likely to actually go ahead and perform the act, conditional on being psychologically disposed, in some sense, to do so.

Others also endorse a binary notion of feasibility that is a property of acts. There are differences, however, in emphasis, and in the kind of 'acts' that are taken to be at issue. Wiens (2015), for instance, attends much more closely to how motivational and resource constraints together affect what acts/ options in the political domain are sufficiently likely to be realised and are thus 'feasible'. In so doing, the acts he considers are rather more complicated ones involving multiple agents. There is not such an obvious connection between feasibility and a decision-maker's 'choice set'

[5] This is, roughly, the phrase used by Southwood (2016, 11), who attributes it to Estlund (2011). Note that even if probabilities are employed in reasoning about what is feasible, on this account ultimately the judgment must be binary because arguably it does not make sense for an act to be only partially in an agent's choice set.
[6] As per Southwood (2016, section II)

(to use the decision theory language introduced above). Gilabert and Lawford-Smith (2012) also extend the target of feasibility assessments beyond simple acts; they include a goal as part of their account, thus proposing a four-place predicate for feasibility: On their view, feasibility pertains to a particular agent performing a specified act in order to achieve some goal in a given context. Again, some of the acts (paired with goals) they consider are rather complicated ones. (They also introduce more and less minimalist feasibility assessments, presumably to cater for the term having multiple roles in ethical discourse; one is a binary notion and depends on 'hard' logical-metaphysical constraints, while the latter is a graded notion and depends on further 'soft' empirical details.[7])

One might diagnose the situation as follows. There is disagreement about what is the most useful notion of 'feasibility' when it comes to directives for action. Specifically, should the target be the decision theoretic 'acts' that would, if feasible, be the basic items of a decision-maker's choice set? (The possible consequences of these acts would then be another, further matter.) Or should the target rather be more complicated activities—complex acts that may take years to carry out, say, and are perhaps also associated with a goal? The question is important because it may not always be fruitful to think of complex acts as options that a decision maker can, depending on a binary feasibility assessment, simply choose to perform. At least some kinds of complex acts seem to call for a notion of feasibility that comes in degrees,[8] and accordingly, are not best conceived as the choice options available to a decision maker.

To avoid ambiguity, it helps to introduce a new term for the kinds of 'complex acts' referred to above: Let us call them *projects* or *multi-stage plans*. This seems to best characterise the target for feasibility assessments in the climate debate. (The account developed here need not be the only useful notion of feasibility.) Climate treaties or proposals are more like projects or multi-stage plans than simple acts that a decision maker may or may not choose to perform. Indeed, climate treaties not only concern a

[7] While Southwood (2016) focuses on different logical roles for 'feasibility', depending on the moral concept at issue, Gilabert and Lawford-Smith seem to focus rather on different definitions of 'feasibility', depending on the moral concept at issue.
[8] As per Gilabert and Lawford-Smith's *graded* notion of feasibility that takes into account 'soft' constraints

course of action that spans some considerable length of time; they also involve more than one agent (with the further complication that these agents are themselves groups). The feasibility (or predicted success) of such a project surely comes in degrees. But given projects are not the sort of thing a single agent may simply decide to perform, does feasibility still concern directives for action and pertain to an agent's deliberations? Yes. For starters, it is hard to make sense of the feasibility of a project or multi-stage plan, complex as this series of acts may be, absent the perspective of a deliberating agent and a specific way of initiating the project (even if these matters of perspective are not always made explicit in our everyday talk). Second, there are good reasons for a decision-maker to assess the feasibility or likelihood of success of a project, were it initiated in a specific way. The most obvious reason is that the decision maker is considering whether it is worth doing the initiation herself. This clearly depends on the likelihood that, were she to take this action, the project would eventually be successfully implemented. That is, the full implementation of the project is one possible consequence of the act of initiation, and thus the likelihood of this consequence – the feasibility of the project were it initiated in the given way – is relevant to the agent's deliberations about whether the project ought to be initiated in that way.[9]

It is worth clarifying the project account of feasibility before moving on to the more substantial questions concerning self-interest and Pareto improvements. We see that projects are of concern to a deliberating agent. The difference with basic acts is that, at the point of choice, the agent can merely initiate the project. Subsequent steps in the completion of the project are not directly under the agent's present control; they are the province of other agents or perhaps the 'future selves' of the initiating agent. So projects are rather complex and there is room for a variety of eventualities once they are initiated. One of these eventualities or possible consequences is the full implementation or successful realisation of the project. The likelihood of this outcome is the feasibility of the project.

Note that we refer to projects or multi-stage plans in diverse ways and so many feasibility statements could conceivably refer to projects. (Consider the examples

[9] By the same token, the nature and likelihood of the project *not* being successfully realized is also relevant to the agent's deliberations.

given earlier.) Projects are sometimes described as if they were single actions, albeit complex ones that would take a number of steps to implement over some period of time. Consider the 'act' of me working without procrastinating, or, at the grander scale, the 'act' of Australia reducing its greenhouse gas (GHG) emissions by 26% on 2005 levels by 2030.[10] Note that the latter example of Australia reducing its emissions may equally be regarded a goal. Indeed, projects are often described in the language of goals, for instance, 'the goal of limiting climate change to 2 degrees C rise in average global temperature', or 'the achievement of universal access to high-speed internet'. Moreover, projects may be more or less detailed. The goal of limiting climate change to 2 degrees C may be stated in such a way that the precise means to this end are specified, or else rather left open.[11] But however they are referred to in casual talk, the feasibility of a project, on the account given here, depends on the way it would be initiated by the relevant agent and what the whole project amounts to, or what are the conditions for its success; the more precisely these details are specified, the more precise the feasibility assessment.

A further issue is how we should think about the feasibility of projects involving multiple agents. One may refer loosely to 'group projects' but the idea here is that any project must somehow be initiated, and this concerns the choice of a single agent or decision maker.[12] The choice concerns whether to perform the initiating act. Whether the project will succeed then depends, amongst other things, on the *predicted actions of the other agents* who are relevant to the success of the project. In other words, the behaviour of these other agents, from the point of view of the decision maker, are simply aspects of the world, like whether or not it will rain, that bear on the consequences of the initiating act. Note that the decision maker may relate to the wider group in a variety of ways: He/she may be a member of the group who, for the purposes at hand, is regarded the deliberating agent. Or else he/she may be an onlooker to the group who is nonetheless interested in the initiation of the group

[10] As agreed to in Paris in 2015

[11] Compare the following two projects regarding the mitigation of climate change: i) Reduce greenhouse gas (GHG) emissions by 26% ii) Implement extensive program of carbon-capture-and-storage so as to reduce GHG emissions by 26%. The latter project is obviously more detailed in terms of the steps to be taken, and as such, its feasibility can be more precisely specified.

[12] Admittedly, a group may count as a single agent if it is sufficiently unified in the relevant ways (as per List and Pettit 2011). But the conditions for agency are relatively difficult to satisfy, and one would be hard pressed to argue that the groups of nation states that concern us here count as single agents (even if each nation state separately counts roughly as a single agent).

project. To see the distinction, consider the assessment of a proposed international climate treaty by i) a representative of a participating state and ii) a UN official charged with brokering the deal. These agents may justifiably have differing assessments of the feasibility of the treaty, due to differences in how they would initiate the project.

So feasibility is the probability that a project or multi-stage plan would be successfully realised if initiated in a particular way by some agent. One further important question concerns the nature of this probability judgment. How exactly should we interpret the probability in question? Is it some agent's credence that a project would be successfully realised if initiated in a particular way, say, the credence of the decision-maker in question? Or is it supposed to be an objective chance? There is room for some latitude here, depending on the precise function that the feasibility assessment is intended to play – perhaps it informs the advice an onlooker gives to an agent about how to act, or perhaps it is part of the deliberations of the agent him/herself. Whatever the precise details, it seems intuitive that feasibility is in some sense an objective matter. Perhaps it is best considered an objective epistemic probability – the credence that the decision-maker *should ideally* have, if appropriately informed, regarding whether the project in question would be successful if initiated in the specified way.[13]

2.2 Feasibility and Self-Interest

The characterisation of 'feasibility' above clearly does not *directly* concern self-interest, let alone Pareto improvements (which we return to in later sections). Feasibility is not a property of a value function *per se*; it is not a measure of the extent to which a theory or account of value is in line with self-interest. This point is worth elaborating. It is generally agreed that feasibility is not pertinent to the evaluation or ranking of states of affairs.[14] Thus if IP were understood as a theory of value (with

[13] Of course, this raises further questions (which, again, come back to the precise function of the feasibility assessment) as to what it means for the decision-maker to be 'ideally/appropriately informed'. The broad account given here, however, is sufficient for the purposes of this discussion.
[14] Wiens (2015, 461) explicitly makes the distinction between directives for action and purely evaluative ethical claims, and notes that feasibility is pertinent to the former as opposed to the latter.

Pareto improvements on the status quo ranked higher than states of affairs that are not Pareto improvements), then feasibility would not be a relevant consideration in assessing IP. More generally, when the question is one of evaluation—say, how we should evaluate different distributions of the costs of responding to climate change—feasibility is not relevant. Note that much discussion about the ethics of climate change concerns the aforementioned question of evaluation. It is only once we start talking about directives for action that feasibility becomes relevant, and as discussed, it concerns the likelihood that certain outcomes/events will come about, as opposed to the value of these outcomes.

Presumably when proponents say that 'IP is the most (or only) feasible approach to climate change', they are not making a purely evaluative statement but rather something in line with our working definition of feasibility above: That climate treaties designed to achieve a Pareto improvement on the status quo are likely to succeed in a way that climate treaties designed to involve sacrifices, are not. (IP proponents in fact tend to speak of feasibility as a binary matter, but as per the discussion above, a graded assessment is arguably more appropriate.) This is supposedly because the agents involved will only play their part in a treaty project if it furthers their self-interest. If true, this would amount to a more *indirect* relationship between feasibility and self-interest. It is an empirical claim about what reliably motivates the agents that climate projects depend upon.

Let us not forget that, throughout our discussion, the function of feasibility assessments is to aid a decision-maker. One thing to note is that, if we focussed just on whether it would be possible for a decision-maker to *initiate* a project in a specified way (whether this is a feasible act a la Southwood and co.), the association with self-interest would be tenuous. Recall why a simple act may not be a viable choice for an agent: If he/she would suffer some kind of irrationality or weakness of will that would ultimately prevent the performance of the act, *even if he/she preferred it and tried to choose it.* This does not let the agent off the hook easily when it comes to onerous acts involving self-sacrifice. In most cases, agents who act selfishly do not do so because it was not viable (or *feasible*, by the lights of Southwood and co.) for them to do otherwise. More likely a range of acts were viable, and yet the agent preferred the selfish one. To give an example, it is surely perfectly viable for me to set

up a regular online donation to a charity. That is, if this were what I preferred to do, I would succeed in doing it. If I do not set up the regular donation, it is rather because this is not in fact what I prefer to do, given my beliefs and values.

It may thus be perfectly viable for a decision-maker to initiate both more and less demanding climate treaties. But some of these treaty projects may be more likely to be fully realised than others, i.e., some of these treaty projects may be more feasible than others. This is the sense of feasibility that was argued above to be most apt for our discussion here. It is easy to see why this notion of feasibility is important for a decision-maker – it concerns how likely are the planned consequences of an initiating act – and moreover, why feasibility, thus understood, may be associated with self-interest. It all depends on the context and the relevant empirical facts. The likelihood that a complex project will be realised, if initiated in a specified way, depends on whether other agents the project depends on (perhaps even 'future selves' of the decision maker) will play their part.

The substantial question then is: Should a decision maker be more confident that a project or multi-stage plan she initiates would be realised to the extent that others are expected merely to act on their self-interest? Common wisdom suggests the answer is 'yes': Arguably, people more reliably act on self-interest than on other motivations. In that case, a project in which others are expected to act on self-interest as opposed to other motivations would be more likely to succeed, and would consequently be more feasible. But this picture of motivation may of course be overly simplistic. Moreover, it surely depends on the type of agent involved, whether an individual or a group entity such as a state. For now, let us simply note that the association between feasibility and self-interested motivations is an empirical matter, and one that is sensitive to context.

3 International Paretianism and Climate Change

Putting aside, for now, the relationship between self-interest and feasibility, let us examine what the climate predicament looks like if cashed out in terms of the self-interest of the nation state actors. The first thing to note is that the self-interest of

nation states – being large sprawling groups extending over time and space – is not a straightforward matter. As noted in Section 1, here we assume that the 'national interest' is an appropriate aggregate of the welfare (however this is cashed out) of present and future citizens (with considerable, if not equal, weight to future citizens). That is, we assume that there is some appropriate measure of the national interest of a state, as per the self-interest of an individual.[15]

The initial aim here, in Section 3.1, is to canvas plausible models of the climate-change predicament when posed in terms of the national interest of state actors. Whether or not state leaders are actually motivated to pursue the national interest, as roughly defined here, is a further issue that will be discussed in Section 3.2, along with other motivational concerns about climate treaties that aim for Pareto improvements. That is, the game models here do not necessarily track states' existing motivations (or rather, their apparent motivations/values given their choices). The models rather present one way of conceiving the climate-change predicament – where (the outcomes of) strategies are evaluated according to the national interest of the respective actors.

3.1 Prospects for IP Climate Treaties

This section considers what an IP climate deal might look like by appeal to key findings in public economics regarding the usage of 'common pool resources'. The key question is: What kind of game model plausibly fits the empirical facts and our rough definition of national interest? This is crucial for determining what sort of Pareto improvement is in the offing. Posner and Weisbach (2010, p. 6), for instance, suggest (in some places at least) that IP would support a global mitigation effort that is optimal (at least by utilitarian standards) and moreover substantial. We will see that this is not necessarily the case; indeed a number of factors affect the collective

[15] Note that Posner and Weisbach (2010) do not consistently treat the 'national interest' this way in their defense of the IP approach to climate change. What they mean by a nation's interest seems to vary, depending on whether they are discussing the feasibility merits or rather the moral merits of IP. When it comes to feasibility, they seem to allow future welfare, for instance, to be weighed in the 'national interest' just as much as current citizens see fit (see, e.g., the discussion of the importance of respecting a state's sovereignty with respect to its own future in the discussion of the asteroid analogy on p.77). In many other places, however, they emphasize the ethical significance of the mitigation that would come from an IP treaty, suggesting a more ethically robust notion of 'national interest' as per our discussion in this chapter (see especially their comments on p.177 to this effect).

optimality and extent of mitigation under IP, as well as the nature of individual contributions.

The clearest reason for there being an unrealised opportunity for Pareto improvement on the status quo is a failure of collective action of some sort. Environmental resources are prone to collective action problems. Consider, for example, the over-exploitation of fisheries, deforestation, and of course climate change. The diagnosis: Many environmental resources such as those mentioned (abundant fisheries and forests, a stable climate, etc.) are *common pool resources* shared amongst agents who are primarily concerned with their own consumption of it. Technically speaking, common pool resources are effectively *non-excludable*, meaning that there are no barriers to anyone using the resource.[16] Common pool resources have a tendency (absent any institutional arrangement) to be over-exploited, because no one in particular has the responsibility/ability/motivation to provide or maintain the resource given that others cannot be excluded from using/spoiling it.

Climate stability is surely a classic example of a global common pool resource.[17] It is commonly assumed that the collective action problem takes the form of a Prisoners' Dilemma of international magnitude, where the 'cooperative solution' would be a clear Pareto improvement over the 'non-cooperative solution'. But this need not be the structure of the *game*, even assuming that the state actors pursue their 'national interest'. The Prisoners' Dilemma, considered particularly problematic when it comes to cooperation for mutual benefit, can be contrasted with 'mere' coordination problems. Moreover, even when it comes to Prisoners' Dilemmas, there is variation in the size of the Pareto improvement in question. In what follows, these different game scenarios are discussed in more detail: Section 3.1.1 briefly considers the

[16] Some environmental resources may also be *non-rivalrous* (at least up to some threshold), meaning that the resource in not depleted (up to the threshold) by additional users. Goods that are both *non-excludable* and *non-rivalrous* are referred to as *public goods*. (Strictly speaking, it is a matter of degree as to whether a good has these properties.) While the literature on environmental dilemmas often refers to *public goods*, it is generally only the property of non-excludability that is at issue, in which case, better simply to refer to *common pool resources*.

[17] Note that there are various ways to depict climate change as involving a common pool resource. Here we refer to the resource of 'climate stability', as per Ostrom (2009). Others refer to the 'carbon sink', or the Earth's ability to absorb greenhouse gas emissions, as the common pool resource, rather than a stable climate *per se*. The difference is not a substantive one.

coordination-game account of climate change, after which Section 3.1.2 turns to the Prisoners' Dilemma account and the variety of forms it may take.[18]

3.1.1 Climate change as a coordination game

Some suggest that, due to the presence of a dangerous threshold of emissions that would bring catastrophe to all, swamping other impacts, climate change may be best conceived as a *coordination game* (see Barrett 2011 for discussion).[19] Here the cooperative solution simply requires coordination: If all states were confident that the others were pursuing a common project for emissions abatement that just meets the threshold for avoiding catastrophe, then it would be in the interests of each to pursue this project as well. In other words, the presence of a dangerous threshold would, ironically, be good news for collective action, because there would exist arrangements for emissions abatement that are stable in the sense that no actor would have an incentive to defect. These sorts of joint strategies that resist unilateral defection are referred to in game theory as *Nash equilibria*; they are commonly regarded the 'solutions' to the game.

The presence of a dangerous climate-change threshold is unlikely to be as rosy as first appears, however, even from the perspective of collective action. To begin with, there is the problem of too many Nash equilibria. There would be countless projects or joint strategies involving just enough emissions abatement to avert catastrophe; some of these would involve heroic abatement efforts on the part of any given actor with little abatement from others, whereas some of them would involve very little abatement on the part of the actor in question and much greater efforts from others. It would be no trivial task for the actors to settle on one particular stable joint strategy – a tough

[18] The models of this section follow the classic work of Barrett (1990) and Carraro and Siniscalco (1993) on international protection of the environment. Note that Gardiner (2001, Section VII) has a broadly similar discussion of the alternative ways to conceive of the intra-generational problem of environmental protection (primarily climate change); a key difference however, is that, in the first instance at least, Gardiner focuses on individual persons as the actors rather than nation states. It is worth noting also a further issue, which will not be explored here (in line with the literature on international environmental protection): How the structure of an existing game may be changed via new incentives. Ostrom (1990), for instance, has done very important work on the diversity of governance schemes that may resolve the over-exploitation of a common pool resource, by effectively changing the nature of the game. For the most part, however, the situation is such that, even if actors consent to a scheme for governing the resource, an external authority of some sort is required to ensure compliance with this scheme. It is typically assumed that such an authority is not available in the international setting.

[19] Typically the game is simplified such that there is negligible benefit to emissions abatement that is short/in excess of the designated threshold.

bargain. As such, it might be risky for any given actor to pursue a particular joint strategy for averting catastrophe, given there is no guarantee that others would also opt for the same equilibrium strategy. The riskless Nash equilibrium would be for all to do little and suffer catastrophe without additionally engaging in costly and potentially futile abatement.[20]

An even more crucial issue, however, is that while climate change plausibly involves a danger threshold, there is uncertainty within the scientific community about its location. (The critical amount of 2 degrees C warming has gained traction in international debate, but apart from any other complications, there remains uncertainty about what *likelihood* for this temperature increase should be treated as the threshold, and there is furthermore uncertainty about the stock of atmospheric GHGs that corresponds to the chosen likelihood.) According to Barrett and Dannenberg (2012), this sort of uncertainty would effectively turn the prima facie coordination game into the more difficult Prisoners' Dilemma (PD) game. For this reason, our main focus here will be the PD account of climate change. Section 3.1.2 considers the variety of forms, when it comes to the magnitude of the effect (so to speak), that the climate PD could take.

3.1.2 Climate change as a Prisoners' Dilemma

It helps to begin with the simplest characterisation of how climate stability may give rise to a Prisoners' Dilemma. In our simple model, there are N relevant actors that affect the climate by (potentially) emitting greenhouse gases (GHGs).[21] Assume the N nations have equivalent circumstances and so the game is entirely symmetric (we return to this assumption later): Each actor can reduce or abate emissions, relative to the status quo or 'business as usual' (BAU), at a constant cost of c and a constant benefit of b per unit of abatement. Given that climate stability is a *common pool resource*, the costs of emissions abatement are private (borne just be the actor doing the abatement) but the benefits of this abatement accrue to all actors. So if each of N

[20] This is to suggest that the coordination problem has the form of the so-called 'Stag Hunt' (see Skyrms 2004 for extensive discussion of this game). Others suggest the climate problem has the form of the roughly similar 'Battle of the Sexes' game (as discussed in Gardiner 2001).
[21] Even though there are other ways to hasten/mitigate climate change (e.g. (de)forestation), GHG emissions play a key role and thus are our focus here.

nations abates one unit of emissions, they each receive a benefit of Nb. The classic Prisoners' Dilemma results from the following pattern of costs and benefits:

$$Nb > c > b$$

Since, for each actor, the individual cost, c, of their own personal abatement effort is greater than the benefit, b, they would receive from this effort, there is no incentive to act. Whatever others do, it is better for the individual to do nothing. And yet, all individuals would benefit if they pulled together, and therein lies the dilemma: If all individuals were to similarly engage in emissions abatement, the benefit to each and every actor, Nb, surpasses their respective individual costs, c.

To use more formal language: In the PD game just described, the *non-cooperative* solution, where each actor chooses their level of emissions abatement based on their own interests, holding fixed what others do, amounts to zero total abatement and so zero total net benefit. This is because, for each actor, regardless of what others do, their own marginal contribution to the climate through abatement brings less private benefit b than the private cost c. This is the *Nash equilibrium* for the PD game—the solution that is stable in that no actor has an incentive to defect. The *fully cooperative* solution, on the other hand, is what brings maximum benefit to each actor; it is also optimal for the group in the aggregate welfare sense. Here it involves all actors pursuing maximum abatement x, which brings $(Nb - c)x$ net benefit to each actor. Note, however, that this Pareto improvement is not a (stable) Nash equilibrium as all actors have an incentive to defect and not abate at all.

A more realistic model has cost and benefit functions that are not constant but rather depend on levels of emissions abatement.[22] Assume $i = 1, \ldots, N$ symmetrical actors as before. Let the total emissions abatement, Q, be the sum of all individual abatement efforts; so $Q = \sum_N q_i$, with q_i being the abatement of each actor. Assume that the marginal benefit of a unit of abatement for each actor, B_i, depends on total abatement as follows:

$$B_i(Q) = b(aQ - \frac{Q^2}{2})N \qquad \text{where } a \text{ and } b \text{ are positive parameters}$$

[22] This is considered more realistic for a pollutant that does not accumulate in the atmosphere, unlike GHGs. Barrett (1994) notes that Nordhaus (1990) employs a logarithmic cost function that depends on percentage emissions abatement, to model costs associated with GHG emissions abatement in the US. The precise forms of the cost and benefit functions, however, do not affect our (more general) discussion here.

This is a concave function with respect to global emissions abatement, which is to say that there are diminishing marginal returns for each additional unit of global abatement. The cost of abatement for each actor, C_i, is a convex function of its own abatement levels:

$$C_i(q_i) = cq_i^2/2$$

This amounts to increasing marginal costs of abatement: the more a nation individually abates, the more costly the next unit of abatement.

As per the previous model, the *non-cooperative* (Nash equilibrium) solution for non-constant cost/benefit functions is determined according to abatement levels that individual actors would choose, holding fixed the abatement of others. The solution, corresponding to total abatement Q_0, is such that the marginal abatement cost for each individual equals the marginal abatement benefit for that individual alone. The *fully cooperative* solution, on the other hand, corresponding to total abatement Q_C, is such that the marginal abatement cost for each individual is equal to the global marginal abatement benefit. That is, Q_C maximises the total benefit to the group, measured as the sum of benefits to individuals minus the sum of costs. The abatement and the net benefits for each actor under full cooperation are greater than the abatement and the net benefits for each actor under the non-cooperative solution; the difference, Q_C - Q_0, being the gain from full cooperation.

The size of this gain depends on the ratio *c/b* (the cost factor to the benefit factor) and the size of *c* (see Barrett 1994, pp. 880-1). When *c* is small and *b* is large, the cooperative gain is small because actors already benefit significantly from unilateral or non-cooperative abatement. When *c* is large and *b* is small, the cooperative gain is small because it makes little sense to abate, whether unilaterally or cooperatively. The biggest gains from full cooperation, both with respect to total abatement and net benefits, are when *c* is comparable to *b* and both are large.

So we see that if the international climate predicament were best modelled as a Prisoners' Dilemma (PD), the size of the Pareto improvement (let alone whether or not it is achievable) is yet a further question. The model above well illustrates this point. For instance, in one of the PD scenarios mentioned just above—where the individual costs of emissions abatement are small and the benefits large—little is

gained from a cooperative climate deal *per se*. Of course, such a model, if plausible, may nonetheless inspire change in revealing that the international community falls short even of the *non-cooperative* solution with respect to GHG emissions abatement; perhaps states need to start acting on their 'national interest', so to speak, rather than overcome a further collective action problem.[23] Some have indeed argued that it is not well enough appreciated that the benefits a single state receives from its own emissions abatement does to a large extent surpass the costs. Just because climate stability is a common-pool resource does not automatically mean that it is against a state's national interest to unilaterally do anything about it. There may well be indirect national benefits (or, in other words, lesser national costs than appreciated) associated with emissions abatement, such that the difference in abatement under the cooperative and the non-cooperative solutions is small. Ostrom (2009), for instance, emphasises local gains (or co-benefits) associated with emissions abatement, including savings in energy costs and health benefits from lower pollution (cf. Maslin and Austin 2012 on energy security and air quality).[24]

The other PD scenario where the gains of cooperation are slight, is when the cost factor, *c*, for emissions abatement is extremely large compared with the benefit factor, *b*. In this case, there is little reason, whether non-cooperatively or cooperatively, to engage in emissions abatement. This kind of model would only be plausible if, for each state, i) the BAU climate change path was relatively innocuous and/or ii) the welfare of future citizens was largely irrelevant to the 'national interest'. The first scenario is not very likely, given our best science. There is much uncertainty, possibly irresolvable, about how particular regions of the world will be affected by any particular rise in average global temperature, but nonetheless it is most plausible that the *expected* (in the mathematical sense) consequences are negative for all, compared

[23] If this were the case, the supposed motivational power of the 'national interest' that is implicit in the defence of IP may be undermined, because one reason for states not currently pursuing their national interest and achieving the non-cooperative solution is that this is politically difficult. Section 3.2 will revisit this general issue of how easy it is for states to pursue their national interest.

[24] Others focus on the dynamics of the problem, and argue that the costs of emissions abatement decrease rather than increase, the greater the abatement effort to date (for various accounts of this phenomenon, see, for instance, Victor 2011, Heal and Kunreuther 2011, Tavoni 2013). The basic idea is that new partnerships and energy economies associated with emissions abatement would rapidly reach some critical mass whereby it is less costly to be in than out.

to a future with a more stable climate.[25] The second reason is a non-starter in the context of our discussion here. It is true that, the more the future is discounted, when it comes to a state's 'national interest', the less benefit and the more cost associated with emissions abatement. But a heavily discounted future contravenes one of the premises of our discussion, namely that the 'national interest' of a state is some *appropriate* aggregate of the welfare of present and future citizens.

So it is important to note that the size of the cooperative gain associated with a PD can vary quite significantly. Thinking about these possibilities helps to sharpen one's views about the climate predicament that we face (based on 'national interest' as roughly defined here). There is reason to think that the status quo falls short even of a non-cooperative climate solution, that things would be different if states really acted in their national interest. On top of that, there is reason to think that the cost of GHG emissions abatement for individual actors is high, and the gains from cooperation also high. That is, it is not at all implausible that the climate predicament involves a Prisoners' Dilemma where the gain from cooperation would be substantial and thus tragic if not achieved. (It would arguably be more convenient if climate change gave rise to a 'mere' coordination game, but this seems wishful thinking.[26]) A further interesting question is whether the fully cooperative solution to the climate PD would effectively be what a global planner would recommend as the optimal mitigation response. This is a topic for further investigation, but the answer surely depends on whether the total welfare calculations of the global planner align with the total welfare calculations that underwrite the national interest of the respective states.[27] With the character of the Pareto improvement in mind, let us now consider whether a climate

[25] See Frigg et al. (forthcoming) on the uncertainties surrounding predictions of *regional* climate change. For the multi-region models of climate economics investigated in Nordhaus and Boyer (2000), it is indeed the case that the expected consequences of BAU climate change for all regions of the world are negative, compared to selected mitigation options. Note too that the mitigation options are costly for the present generation in all regions of the world, compared to BAU. This makes clear that mitigating climate change is better than BAU (on these models at least) only because significant weight is given to the welfare of future generations (enough to counterbalance present costs).

[26] Posner and Weisbach (2010) often seem to assume, through their choice of analogies, that climate change is a coordination game. They do, however, also discuss the possibility of a Prisoners' Dilemma scenario (2010, 181).

[27] For instance, for a utilitarian global planner – as in the analyses of Stern (2007) and Nordhaus (1990) – the fully cooperative PD solution would only line up with that of the global planner if each state's national interest were also a utilitarian aggregate across all citizens in the present and future. This question is relevant to the 'efficiency' argument for IP that was briefly mentioned at the start of Section 2 (see, especially, footnote 2).

treaty that aims for this outcome is really achievable just on the back of ordinary 'self interested' motivations.

3.2 Does 'self-interest' suffice?

If the climate-change predicament were plausibly modelled as a game showing a large mutual gain for the state actors if only they could somehow rise to the cooperative task, then this would seem to be good news for International Paretianism. After all, IP is all about climate proposals that are *feasible* in virtue of appealing only to the cause of self-interest. And surely we can count on actors to merely pursue their self-interest (give or take some misconceptions they might have about what is really in their own self-interest). Even if we grant this simple rule of thumb about motivation, however, there is reason to doubt whether the Pareto improvements suggested by the Prisoners' Dilemma models canvassed above are achievable just on this basis. To begin with, we should be cautious in treating the 'national interest' of states akin to the self-interest of individuals. And secondly, even if states may be persuaded to pursue their national interest, as defined here, this alone does not suffice for overcoming a Prisoners' Dilemma. In what follows, these two issues are discussed in turn.

Many would observe that the analogy between the self-interest of an individual agent, both in explaining and guiding behaviour, and the 'self-interest' of a group agent, is far from tight. The group agents here are nation states. Do states naturally pursue their 'national interest', especially as understood here to involve due concern for the welfare of both present and future citizens? This seems a rather heroic assumption, and yet an important one for IP, because the appeal to 'self-interest' has force precisely because it is supposedly motivating in a straightforward way; something that an agent would quickly recognise as worth pursuing. The obvious worry is that state leaders rather have perverse incentives, when it comes to their domestic and international negotiations, having little to do with the pursuit of the 'national interest'. Moreover, even if democratic institutions provided some assurance that state leaders

were responsive to the welfare of their citizens, it is a further stretch to think that this would somehow incorporate the welfare of future citizens.[28]

While there is a big question mark as to the motivating force of enlightened 'national interest' in international negotiations, let us nonetheless proceed on the basis that this is at least a persuasive reason for state action.[29] As noted already, it would no doubt be a significant international achievement if states were persuaded to act in their national interest, thus achieving the non-cooperative game solution. As far as securing the further mutual gains associated with the cooperative game solution – the promised Pareto improvement – there are, however, further motivational worries. The problem is that *projects* conforming to IP may require a lot more than the pursuit of national interest to succeed, once initiated. Here the lessons of game theory are important. The notorious problem with Prisoners' Dilemmas is that the fully cooperative solution is not a Nash equilibrium, and worse still, the greater the gains from cooperation, the greater the incentive for individual actors to free ride on the efforts of others, if the project were initiated. So even though the Pareto solution does further all actors' interests relative to the status quo, more than self-interested motivations are required to successfully implement this solution. Actors pursuing self-interest alone will try to free ride on the efforts of others.[30] To use the language of feasibility, if states really were purely concerned with their national interest and the game has a PD structure, the Pareto solution is not feasible.

In many contexts a PD game analysis serves as a plea to change the incentive structure of the game, perhaps via punishments for free riding.[31] But this is arguably not an option in the international setting, given the lack of an overarching authority

[28] In response to this point, Posner and Weisbach (2010) might be tempted to loosen the notion of 'national interest' so that it is more in line with the existing apparent motivations of states. But, as noted already in footnote 13, this would conflict with their claims about IP solving the climate predicament; the climate solution they have in mind depends on an account of national interest that gives due weight to the welfare of both present and future citizens.
[29] Arguably, people are more naturally disposed to care for their own descendants as compared to contemporaries in distant parts of the world (cf. comment to this effect in Posner and Weisbach 2010, 142).
[30] Posner and Weisbach (2010) do acknowledge this point (see, esp., p.181), but they do not explore it in detail and nor do they give it appropriate emphasis.
[31] Recall the earlier reference to the work of Ostrom (1990).

that can facilitate such a change.[32] Economists have otherwise explored the possibilities that may arise from coalition agreements involving only some of the actors (here states) in the game. Solutions involving coalitions may be better than the Nash equilibrium and yet also dynamically stable in the sense that no actor included in the coalition has an incentive to leave, and no actor excluded from the coalition has an incentive to join. Hence these sorts of agreements are dubbed 'self-enforcing' within the public economics literature (see Carraro and Siniscalco 1993, Barrett 1994 & 2005).

Unfortunately, the cooperative possibilities associated with 'self enforcing' agreements are rather limited, as the aforementioned authors have shown. By way of elaboration: A stable or 'self-enforcing' coalition arrangement is one where the individual benefit to all members within the coalition is greater than it would be if they unilaterally defected (since the remaining coalition members would accordingly lower their own cooperative efforts, in this case emissions abatement), and moreover, the individual benefit to all those outside the coalition is greater than it would be if they were to become members (since the extra gains from further cooperation would not outweigh the individual costs).[33] Of course, it may be the case that all states would prefer to be outside the coalition, thus shouldering less of the total emissions abatement, than inside it. States within the coalition cooperate to achieve the greatest total benefit for the coalition, while the remaining states determine their abatement levels according to what is individually optimal, holding fixed the abatement of others. Once a stable coalition has formed, however, no state can benefit from unilateral change.

Across a wide variety of assumptions regarding the precise shape of the benefit and cost functions underlying a Prisoners'-Dilemma game, it turns out that little is achieved over and above the non-cooperative solution by stable coalition arrangements (as shown in Barrett 1994). The first game described above in Section

[32] Although one might argue that a higher authority can at least be approximated in the international setting via multiple agreements, whereby punishments for non-cooperation in one domain (e.g. environmental protection) would take the form of sanctions in another domain (e.g. trade). See, for instance, Barrett (2008) for discussion. Nordhaus (2015) makes a similar point—that a groups of states or 'climate club' can effectively impose external trade-related penalties on states that do not cooperate with respect to emissions abatement.

[33] The notion of a stable coalition is attributed to earlier work on cartels by d'Aspremont et al. (1983).

3.1.2, for instance, involving constant marginal cost and benefit functions, unsurprisingly does not support any self-enforcing coalition agreement. The second game, involving a concave benefit function and a convex cost function, supports only a limited advantage in coalition formation. For instance, when the number of stakeholders, *N,* equals 100, and the cost to benefit parameter ratio, *c*/*b*, equals one (so there is potentially much to gain from cooperation), the self-enforcing agreement involves only three states; these signatories abate considerably more than they would under the non-cooperative arrangement, while non-signatories abate just a little less. The large number of non-signatories means that the free-riding effect is substantial and so little is gained over and above the non-cooperative solution (see Barrett 1994, p. 882).[34] In fact, to the extent that a coalition in this class of cases has a large number of signatories, there is little to be gained from cooperation (*c*/*b* is either very large or very small). Similar results apply to other functional forms: generally speaking, where there is much cooperative gain at stake, only coalitions with 2 or 3 signatories are stable, and thus not much cooperative gain can actually be realised in a 'self-enforcing' manner.[35]

4 The feasibility of IP climate deals

The assumption that states pursue their national interest may yield a PD climate game with a sizeable Pareto improvement, but we see that 'self-interest' alone may not get us very far in realising that Pareto improvement. IP proponents would thus be better served by a more nuanced story regarding the association between feasibility and self-interest. In fact, proponents do try to sell IP on moral 'other-regarding' grounds as well as 'self-interest' grounds. Throughout their book, Posner and Weisbach, for instance, appeal to the happy confluence of gains in national interest and moral gains under an IP climate treaty. This suggests a hidden commitment to the moral standing

[34] Note that these results are obtained via simulation. For other cost and benefit functional forms, analytic results can be derived.

[35] One might wonder whether the situation changes if we allow asymmetry, i.e., when cost and benefit functions for the *N* states are not identical. There has been some investigation of this case in the literature. While Barrett (1997) derives similarly pessimistic results for heterogeneous as for homogeneous parties to an agreement, McGinty (2007), on the other hand, finds that, assuming transfers are permitted within the coalition, heterogeneity can significantly increase the percentage gain from full cooperation, even when this gain is substantial. It seems that further investigation is required to establish the robustness of this result.

of a project being relevant to its motivational appeal and thus feasibility, even if the moral gains are presented as if they were a mere by-product of the pursuit of self-interest. Recall too that the pursuit of 'self-interest' in the context of IP is already morally demanding. It is hardly the norm for state leaders to act in the interests of present and future citizens, suitable balanced. So IP proponents appeal in various ways to moral reasons for climate action, without necessarily advertising this fact.

Posner and Weisbach make an even larger concession regarding the need to appeal to moral reasons for climate action. They state that motivations beyond 'self-interest' are required to secure Pareto improvements where states have an incentive to free-ride (p. 169):

> '… A key for a climate treaty is what we have called International Paretianism—nations must believe that they are better off with a treaty than without. But the obligation to achieve a broad, deep, and enforceable treaty imposes a serious ethical duty on rich and poor nations alike—the obligation to cooperate. In our view, it is unethical for a nation to refuse to join a climate treaty in order to free-ride off of others.'

While it has an air of authority, this statement of ethical duties has no clear justification.[36] The question, when it comes to the feasibility merits of IP, is whether a climate treaty relying on this extra commitment to secure a Pareto improvement would be *more or less motivationally compelling*, for all actors concerned, as a treaty relying just on the pursuit of national interest.

Let us briefly consider this question. Refraining from free-riding involves a sacrifice in national interest. For this to be motivationally compelling, it would need to be offset by some other, say moral, gain. The question is then: Does an IP climate deal have the required moral merit? Does it make an obvious step forward in moral terms? On certain ways of calculating national interest and the global good, the IP climate solution may well accord with the optimal amount of global climate change mitigation. But note that this is consistent with the present generation of poor countries paying deeply for this mitigation so as to ensure climate stability for their descendants. (The alternative, of course, would be for other nations to pay more for

[36] As acknowledged by Posner and Weisbach (2010, 180), although they do not allow that this lack of justification severely undermines the feasibility case for IP.

the same mitigation effort.) So we see that, in solving one global problem – climate stability – another is exacerbated, namely, inequalities (not to mention the failure to meet historical responsibilities). Does it make sense for nations on the losing side of this equation to accept an imperative to participate in an IP climate deal rather than hold out for an alternative mitigation proposal that better promotes equality? It is far from clear that IP has the upper hand here, in terms of being motivationally compelling.[37]

More precisely, the problem for IP proponents is that, once we allow for other moral motivations to weigh against the pursuit of national interest, the feasibility case for IP is much shakier. This is exacerbated when the gains from emissions abatement are not symmetric and when states have more complex moral concerns (say, they care about equality and/or historical responsibility). In this case, states may have very different views about the moral merit of an IP climate treaty, which may well affect the overall appeal of the proposal for them. For rich, high-emitting states, an IP treaty is good for national interest and it may look pretty good morally-speaking as well, when one compares to the business-as-usual baseline; thus rich states may have both national interest and moral reasons to commit to an IP treaty. But poor states may rather see an IP treaty as morally deficient, compared to other possible treaties that would better promote equality and compensation for past wrongdoing. In other words, they may see little reason to treat business-as-usual as the *moral baseline*, and thus have little moral reason to commit to an IP treaty. In that case, there would be little reason for them to refrain from free riding or defecting from such a treaty.

Let us then conclude on a cautionary note. While concern for the feasibility of climate treaties is important and timely, there is no simplistic way to make such assessments. Feasibility is simply the likelihood, from the deliberator's perspective, that a proposed project or multi-stage plan will be successfully implemented, once initiated in a specific way in a given context. When it comes to climate treaties, the deliberator (whether a member state or an outside official) must assess whether other actors,

[37] See Brennan and Sayre-McCord (2016) for discussion of how moral facts can affect the feasibility of a political proposal. They make the point that the position of those advocating a mutually beneficial or 'no sacrifice' climate treaty on feasibility grounds is inconsistent, at least when coupled with the claim that the broader problem of global injustice can always be addressed later (as per remarks of Posner and Weisbach, e.g. p.92).

namely states, will be motivated to follow through on their part in the project, if it were initiated as well as possible by the deliberator. The more reliable these motivations, the more likely the project would succeed, i.e., the greater its feasibility. We must conclude then that it is an open empirical question as to whether states' motivations would likely align with an IP climate treaty; indeed it is plausible that states' motivations would not be so aligned, given that self-interest alone may be self-defeating, and states may well perceive the moral situation rather differently.[38]

Bibliography

D'Aspremont, Claude, Alexis Jacquemin, Jean Jaskold Gabszeweiz and John A. Weymark. 1983. "On the Stability of Collusive Price Leadership." *Canadian Journal of Economics* 16: 17–25.

Baer, Paul. 2013. "Who Should Pay for Climate Change? "Not Me"." *Chicago Journal of International Law* 13(2): 507–525.

Barrett, Scott. 1990. "The Problem of Global Environmental Protection." *Oxford Review of Economic Policy* 6(1): 68–79.

Barrett, Scott. 1994. "Self-Enforcing International Environmental Agreements." *Oxford Economic Papers* 46: 878–894.

Barrett, Scott. 1997. "Heterogeneous International Environmental Agreements." In *International Environmental Negotiations: Strategic Policy Issues*, edited by C. Carraro. Cheltenham: Edward Elgar.

Barrett, Scott. 2005. "The Theory of International Environmental Agreements." In *Handbook of Environmental Economics Volume 3*, edited by Karl-Goran Mäler and Jeffrey R. Vincent, 1457–1516. Elsevier B.V.

Barrett, Scott. 2008. "Climate treaties and the imperative of enforcement." *Oxford Review of Economic Policy* 24(2): 239–258.

[38] Many thanks to Christian Barry, Richard Bradley, Fergus Green, Jeremy Moss, Nicholas Southwood, Kai Speikermann, and all three editors of this collection for very helpful comments on earlier drafts of this chapter.

Barrett, Scott and Astrid Dannenberg. 2012. "Climate negotiations under scientific uncertainty." *Proceedings of the National Academy of Sciences* 109(43): 17372–17376.

Brennan, Geoffrey and Geoffrey Sayre-McCord. 2016. "Do Normative Facts Matter… to what is feasible?" *Social Philosophy and Policy* 33(1–2): 434–456.

Brennan, Geoffrey and Nicholas Southwood. 2007. "Feasibility in Action and Attitude." In *Hommage a Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, edited by T. Ronnow-Rasmussen, B. Petersson, J. Jonefsson, and D. Egonsson. url: www.fil.lu.se/hommageawlodek, only published online.

Broome, John. 2010. "The most important thing about climate change." In *Public Policy: Why Ethics Matters*, edited by Jonathan Boston, Andrew Bradstock and David Eng, 101–116. Canberra: ANU E Press.

Carraro, Carlo and Domenico Siniscalco. 1993. "Strategies for the international protection of the environment." *Journal of Public Economics* 52: 309–328.

Estlund, David. 2011. "Human Nature and the Limits (if Any) of Political Philosophy." *Philosophy and Public Affairs* 39: 207–37.

Frigg, Roman, Leonard A. Smith, and David A. Stainforth. Forthcoming. "The Myopia of Imperfect Climate Models: The Case of UKCP09." *Philosophy of Science*.

Frisch, Mathias. 2012. "Climate Change Justice." *Philosophy and Public Affairs*, 40(3): 225–253.

Gilabert, Pablo and Holly Lawford-Smith. 2012. "Political Feasibility: A Conceptual Exploration." *Political Studies* 60: 809–825.

Heal, Geoffrey and Howard Kunreuther. 2011. "Tipping Climate Negotiations." *Working Paper 16954, National Bureau of Economic Research*.

Jamieson, Dale. 2013. "Climate Change, Consequentialism, and the Road Ahead." *Chicago Journal of International Law* 13(2) 439–468.

Lawford-Smith, Holly. 2013. "Understanding political feasibility." *Journal of Political Philosophy* 21: 243–259.

List, Christian and Philip Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents.* Oxford: OUP.

Maslin, Mark and Patrick Austin. 2012. "Climate models at their limit?" *Nature* 486: 183–184.

McGinty, Matthew. 2007. "International Environmental Agreements Among Asymmetric Nations." *Oxford Economic Papers* 59(1): 45–62.

Nordhaus, William D. 1990. “To Slow or Not to Slow: The Economics of the Greenhouse Effect”, mimeo, Department of Economics, Yale University.

Nordhaus, William D. 2015. “Climate Clubs: Overcoming Free-Riding in International Climate Policy.” *American Economic Review* 105(4): 1339–70.

Nordhaus, William D. and Joseph Boyer. 2000. *Warming the World: Economic Models of Global Warming.* Cambridge, MA: The MIT Press.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Ostrom, Elinor. 2009. “A Polycentric Approach for Coping with Climate Change.” *Working Paper 5095, World Bank.*

Posner, Eric A. and David Weisbach. 2010. *Climate Change Justice*. Princeton, NJ: Princeton University Press.

Schokkaert, Erik and Johan Eyckmans. 1998. “Greenhouse negotiations and the mirage of partial justice” In *Global Environmental Economics*, edited by Mohammed H. Dore and Timothy D. Mount, 193–217. Oxford: Basil Blackwell.

Shue, Henry. 2013. “Climate Hope: Implementing the Exit Strategy.” *Chicago Journal of International Law* 13(2): 381–402.

Skyrms, Brian. 2004. *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.

Southwood, Nicholas. 2016. “Does “Ought” Imply “Feasible”?” *Philosophy and Public Affairs* 44(1): 7–45.

Stern, Nicholas. 2007. *The Economics of Climate Change: The Stern Review.* Cambridge: Cambridge University Press.

Tavoni, Allesandro. 2013. “Building up cooperation.” *Nature Climate Change* 3: 782–783.

Victor, David G. 2011. *Global Warming Gridlock: Creating More Effective Strategies for Protecting the Planet* Cambridge: Cambridge University Press.

Wiens, David. 2015. “Political Ideals and the Feasibility Frontier.” *Economics and Philosophy* 31: 447–477.