NHH

INSTITUTT FOR FORETAKSØKONOMI

DEPARTMENT OF BUSINESS AND MANAGEMENT SCIENCE

FOR 6 2013

ISSN: 1500-4066

June 2013

Discussion paper

Leadership in Climate Policy: Is there a case for Early Unilateral Unconditional Emission Reductions?

BY

Gunnar S. Eskeland

Norges Handelshøyskole

NORWEGIAN SCHOOL OF ECONOMICS

Leadership in Climate Policy:

Is there a case for Early Unilateral Unconditional Emission Reductions?

Gunnar S. Eskeland[1]

Draft, June 7, 2013

One may misread economic theory on climate policy to provide a warning against unilateral mitigation. While important lessons are drawn from 'global problems require global solutions', these say little about what to do in a phase before or without a global agreement - or with weak ones. In the literature on cooperation and leadership in provision of public goods, early provision may stimulate provision from others. A key to leadership is signaling; an early mover has private information and is motivated in part by knowing that others will follow. Others will follow if they understand that the early mover demonstrates that emission reductions are feasible and adoptable. Our analysis finds that early movers will be cognizant of what they need to demonstrate, and they will be concerned about and act on carbon leakage. Leadership can be deterred by concerns for free riding, but this is more likely for a country or coalition that is large in terms of emissions and face others who are both large and vulnerable to climate change. We suggest leadership is possible early in this century: numbers indicate that few – if any - need find themselves deterred from early action of some sort.

[1] Sparebanken Vest Chair, Department of Business and Management Science, Norwegian School of Economics (NHH), Bergen. The author gratefully acknowledges support from NFR's program Renergi (Celect) and FMR (CenSES), EU (ADAM) and inputs from presentations in Vigo (Fifth Atlantic Conference on Environmental and Energy Economics, 2012), Florence (European University), Oslo (Klimaforum 2012) and Bergen (the BEEER conference), from anonymous referees and especially from Scott Barrett, Michael Hoel and Torben Mideksa.

1: Introduction and Outline

One may accept, as the world's countries have done, that some very significant reduction of greenhouse gas emissions in this century is worth it to mankind. In fact, many of the world's countries, through EU, through climate convention statements (the Copenhagen Accord, for instance), have declared goals that are very ambitious for this century, exemplified by the two degree goal, implying reductions of 80% in emissions by the middle of this century. How can it be done? The difficult part is that this requires cooperation.

Earth is populated by about 7 billion presentliving individuals, by firms, and by about 200 sovereign countries. In terms of consequence, earth also consists of future generations, while carrying the history of past ones. Costs of emission reductions, it appears, are not prohibitive, estimates range from zero to three percent of GDP in this century[2].

In the present paper, we understand mitigation as a global public good, meaning that the benefits thereof are nonrivalrous and nonexclusive. This is the basis for the brief and powerful statement *global problems require global solutions.* Nick Sterns words, in "the economics of climate change (2008), are: "Climate change is global in its origins and in its impacts. An effective response must therefore be organized globally and must involve international understanding and collaboration".

It is possible to conclude that a consequence from "global requires global" is a warning to countries that they should not try to fix this problem themselves. The purpose of the present paper is to scrutinize and set in context the foundation of such a warning. The warning is of course an important one but also one that can be misleading, and we shall emphasize the latter aspect.

The warning against early unconditional unilateral mitigation is attractive not only because it is tempting not to take on mitigation costs unilaterally - of course it is. Also, a unilateral reduction may be partly or fully offset by a response in which other countries (or subsequent generations) emit more than they otherwise would if some have raised their reductions early and unilaterally. One such effect, termed 'carbon leakage' is when mitigation among some countries change world market prices – raise steel prices and reduce coal prices, for instance – so that other countries' emissions rise through increased steel production and coal use. Another effect is the classical 'free rider' response, whereby others' emissions increase because of the reduced threat of climate change or damages. Free riding occurs under noncooperation when there is a public good, as with mitigation of greenhouse gas emissions.

Typically, such free riding is described by a Nash equilibrium in a one-shot game when players move simultaneously, often a "prisoners' dilemma". However, in a variation we shall focus on, if one player is given a chance to move reduce emissions, free riding will be described in the *reaction* – if the first mover reduces emissions others will emit more than

[2] The Stern Review (2007), Heal (2009) and others present estimates in this range, though some include minus one percent.

they otherwise would have, and vice versa. It is in this latter setting we analyse 'early, unilateral, unconditional'.

A reaction of increased free riding cannot – of course – much deter from leadership a country or actions small enough not to change climate prospects. Also, such deterrence of early action will be limited from a country or countries small enough not to change climate prospects, or sufficiently disinterested not to let prospects of climate change influence its emitting activity. Deterrents to leadership – thus – if important at all, will be between countries or groups that are large and sensitive to climate change, and will be important only after some important early mitigation, perhaps akin to what we are seeing in a couple of early decades of this century. Thus, such concerns can probably be contained by ensuing talks, adding weight to the prospects for leadership.

Our notion of leadership

When actions and options arise in a sequence, this allows for a great variation in interpretations and models, and some of these variations qualify the warning by allowing some players to *follow* an action of early mitigation. This possibility exists also in longer sequences, as in so-called repeated games, perhaps akin to a situation where countries periodically talk, perhaps arriving at treaties, then act, talk, and so on. In infinitely repeated games, cooperation can be sustained among forwardlooking players through the threat from noncooperation in the future[3].

We use the term 'leadership' to bring in the possibility - indeed it is used by practitioners – the various possibilities that positive provision *follows* positive provision. We shall define leadership by the fact that someone chooses to follow, and that this is part of what motivates the leader.

We here may think about the term *follow* in three common meanings:

i) after the other, in sequence of time;
ii) caused by the other;
iii) with the same sign as the other. So if you start walking through the forest I head in the same direction after you because you did[4].

So we are asking whether emission reductions from some may be caused by and after emission reductions (or some action in that direction) by another. Importantly, free riding, while often portrayed in a simultaneous move Nash equilibrium, would in this alternative setting typically indicate an effect of the opposite sign – so that emission reductions rasie emissions from others.

[3] That is: if they are sufficiently interested in the future. For any level of cooperation there is a discount rate low enough to sustain it: this type of results is known as the folk theorems of repeated prisoner's dilemma games.

[4] We thereby separate ourselves from Stackelberg leadership in iii) (see below), since a Stackelberg leader often will cause others in the opposite direction, as when leader's increase in output causes reductions from competitors.

It is important to realize that no notion of leadership is included or examined in the general warning founded on the free riding incentive in public goods provision. On the contrary, that warning rather says that absent any structure by which players take clues from each other and may follow, incentives to free ride are such that provision of public goods must be coordinated through a treaty[5].

Of course, the possibility for leadership exists, then early actions may indeed be motivated by the fact or the possibility that others are influenced (i.e. might follow). So a country or a group of countries that acts may consider the consequences for others' actions, in treaty formation or other acts.

Once this is realized, it also becomes clear that not only emissions, or mitigation, or 'quantities', can be on the menu of 'early actions', indeed the menu of early actions may include such items as words, education, R&D expenses, arms purchases, trade or other treaties. Indeed, not only is a game in emitted quantities vastly changed if we acknowledge that 'players' may read something into others' actions. Also, a game about emissions is also vastly changed if we allow early occurrences not only to consist of emitted quantities (this latter reformulation is called strategy space in game theoretic terms).

The 'gobal requires global' statement certainly involves a warning from economists that countries should not act independently but rather act through cooperative agreements. However, economists are also not very optimistic about the prospects for cooperation, even when the cooperative enterprise is deemed productive collectively, as in this case. An obvious point – above - is that free riding can result from not cooperating.

A less obvious point is that the body of literature called cooperative game theory – where parties (countries) may sign binding agreement, is not very optimistic about cooperation, or constructive in how to make it happen. In brief, when there are many potential parties (like three or more), agreements between many of them may fail to be signed because everyone feels that some other 'nonlarge coalition' without her can provide most of the gains from cooperation (in technical terms; the core of the game may be empty even in very productive games, i.e. even in settings where cooperation makes good sense collectively). In cooperative games, while cooperation may happen if all parties know that nothing will happen without all (an example can be if catastrophe happens unless everyone cooperates), in very important settings cooperation will either attract very few or support very little effort (Barrett, 1994).

Outline of the paper

In section 2: "Leadership in public goods provision" we draw from the economics literature on public goods provision, with emphasis on the few contributions – some theoretical, some from experimental and evolutionary settings – illuminating sequences of actions. The economics literature on cooperation is penetrating and ambitious, and may be read as quite pessimistic about cooperation. But a few studies of voluntary contributions and on 'leadership' also show that the economics literature is quite agnostic about how cooperation

[5] Hoel (1991) looks into unilateral mitigation, comparing it to a treaty that is efficient, explicitly ruling out the possibility that unilateral reductions by one country may lead to similar behavior by other countries.

comes about, and especially as a process. Indeed, when there is cooperation and public goods provision in experimental games, it is the extent of cooperation and provision that is puzzling and challenging to theorists, not the breakdown of cooperation and the reduction in provision. Nevertheless, this literature offers several distinct contributions by which one can see a 'positive following', adding nuance to the warning that mitigation from some will raise free riding and emissions from others. Briefly, what leadership entails (i.e. when it causes a following) is a signaling to others that the cooperative venture is an attractive one, and/or that a common understanding of fair burdensharing is possible.

In section 3, "Global problems require global solutions, and the planner's problem" , we detail how present 'lessons' on mitigation really result from a planning model and are relevant only under assumptions of global cooperation where 'the mitigation problem' is solved once and for all. This leads into our reemphasis on *what it is that happens if/when a global or great treaty has not been formed, or is weak*, and *what kind of dependence occurs when actions and choices follow one another?* We could say, today, if some countries and regions press on with emission reductions despite no or very weak treaty commitments: is this plain silly, inviting free-riding from others, or is it possible that its effect on the emissions from others will be to bring them lower than what they otherwise would be, rather than higher?

Section 4, "Pondering the the early problem in absence of a strong global treaty" tries to be concrete about how actions could link between countries and between periods, noting that any early action must be motivated by how it affects global century emissions. Thus, linkages from period to period, from country to country, and carbon leakage all are of interest (unless they can be explicitly ruled out).

Section 5; "Supply of public goods and the free rider problem" presents analysis to the effect – explained above – that for many countries, early actions need not be much deterred by concerns for inducing increased emissions from others.

Section 6 asks: Can a leadership perspective inform or explain early contributions. This discussion is informal and tentative, but since one important lesson from the review in section 2 is that leadership involves signaling by leader and inference drawn by others, it follows that countries may look for each other's motives when they look to each other's actions. This has two important implications: one is that not only quantities – emissions or emission reductions – matter, it could also be actions with no (early) impact on emissions, such as R&D, or even words. Another consequence is that such actions as emission reductions will be scrutinized with other criteria than just the emission reductions and their costs, so the criterion of cost effectiveness will not rule the ground alone, or in any case it will rule in a modified fashion[6].

2: Leadership in public goods provision

In this review section we shall find support for voluntary provision – for taking the lead. If one asks how public goods provision comes about, and whether leadership can play a role, insights on leadership can be found in pure theory as well as in experimental literature.

[6] Since we allow for signaling, what one reads into early action can be important. We would not agree that one needs action rather than words. Words can matter positively, but so can other actions.

In our idea of *leadership*, we ask for a setting where

- one player acts before others to provide a public good – early, unilaterally, unconditionally.

Leadership is then much described by the fact that this instigates others to do the same (or something similar), and we ask:

- what instigates this following?

The leadership we look for will often be chosen by the player herself, in the sense that any one of the players could move first, but this is not necessary, and is in fact rare in the literature. Importantly, thinking about sovereign countries, we are not examining a situation of authority – like credible threat of punishment.

In other words, we want following to be voluntary and not due to a threat of punishment or reward. We do not rule out in general situations where following is assured by credible threat of punishment (though this is not our main interest), but we rule out those situations where one does not ask how those threats are credible, or sustained. Here we are in line with important game theoretic treatments (such as the folk theorem of repeated prisoners' dilemma), which rules out cooperation if the incentives that could support it are not credible.

Leadership is given in the outset

We have argued above that the field of economics is pessimistic and agnostic about cooperation in public goods provision, and thin also on leadership in general as well as leadership in public goods provision in particular.

A term established in economics textbooks is Stackelberg leadership, in which one player is by exogenous construct given the first move. The leader's choices follow from her assumption that the other player(s) will take her actions as given. Thus, the existence of followers (in the sense of players which will take her choice as given) is part of what motivates the 'leader', so the slightly paradoxical fact that followers influence the leader is established even in this very simple structure: If a general charges ahead of her troops, she is motivated in part by her knowledge that her troops will follow.

Then, important literature on organizations expands on formal authority – the right to decide - with questions of what it means to lead. Aghion and Tirole (1997), analyses how much authority to delegate to a subordinate, the tradeoff being that delegation allows subordinate initiative but at a loss of control for the principal. Another contribution on the tradeoffs involved in being a leader when subordinates have private information is Prendergast's 'A theory of "yes-men"' (1993), in which subordinates end up 'guessing' or 'mimicking' the assumed choices of their leader, even though the leader wants them to apply their own information and judgment only, and tries to delegate. As with Stackelberg leadership, this literature does not speak so directly to our question, since authority to lead is given, while our question is on how leadership arises among evenly leveled players (like sovereigns). But a theme in delegation we will rediscover in the more relevant studies is that a leader contributes

with information but also is sensitive to information and preferences of those who may follow.

Perhaps with a more direct pointer to our question is Bolton, Brunnermeier and Veldkamp (2008), in which there is a tradeoff between organizational learning and leader resoluteness: the latter is important to coordinate followers, but is restricted by time consistency given that new information is arriving. In their model, as in Aghion and Tirole's, a leader will be more humble (delegating, listening to her subordinates) if her subordinates have valuable information held privately. We shall rediscover this theme below.

Endogenous leadership based on private information

An important theoretical contribution on leadership in provision of public goods is Hermalin, (1998). Hermalin assumes that followers follow because it is in their interest to do so, and turns the table by allowing that the leader has better information of what they – the whole group including herself - should do. The setting is one in which opportunities for public goods are more or less productive (or attractive, in more general terms), and the leader has better information on how productive such a cooperative opportunity is.

In this setting, leadership is about transmitting information, duly accounting for the obvious challenge of transmitting credibly, so that others will choose to follow. Typically, building on the Holmstrom (1982) team model, there are incentives to mislead. Thus, if a leader makes a proclamation that a given public good provision opportunity is a very productive one, she will not be believed.

Hermalin allows a player who has received private information about the productivity of the game to work hard herself, and followers will draw inference from her effort about the productivity of the game (if she is working this hard, it must be something good waiting!). An important result is that a leader who herself obtains a share of the joint product of the game has an extraordinary incentive to work hard *because her effort draws inference* and thereby effort from followers.

Hermalin shows that the act of leadership – the act of credibly signaling productivity in a public goods game – can involve

- leading by example or
- leading by sacrifice.

Both of these shall be important to us in the following.

Two illustrations: It should clear that the image of a general who charges the enemy in front of her troops applies both to 'by example' and to 'by sacrifice': i) the soldiers – the followers – likely can infer from her behavior that she believes they can win and/or is willing to make a sacrifice (this game is a productive one), so they follow; ii) the general's knowledge that they will follow is clearly part of her motivation; iii) bringing in reputation, we may draw on Sun Tsu's "The art of war": whose teachings include: "the only battles I really won were the ones I did not fight". This renowned Chinese thinker and general in this represents that he values

his men's losses very highly (I only fight when I have to, and only when I can win) rendering his men more willing to follow him. Also: he spared enemy soldiers when he could, thereby enticing them to break ranks and *follow*. Finally, to emphasize the 'private information' part of leadership, Sun Tsu writes repeatedly about intelligence and spies, about the value of knowing the terrain, the weather, your enemy, his status (and exploit all of this).

An important feature here is thus that the player becoming a leader has superior information, typically about the productivity of the game, and followers know or understand this. This part of leadership could be that a person is believed better to understand nature. So, you would follow a guide across the mountains because he knows them well. But, as we shall see below, you might also want him to know you and your condition, and perhaps even to emphasize with you.

In many settings, the belief that the leader is better informed is not a mystery or a question. For instance, in a production enterprise, the leader in constructing a bridge is perhaps the engineer with an understanding of bridges and a plan for the bridge. This applies not less, of course, if the bridgebuilding group is individuals who need an opportunity to cross the river, and look for a leader within. In other setting, the leader is a person who both owns the fishing vessel and has the weather forecasts and fishing skills. There are reasons more than authority and pay to follow her orders, in fact she might be the owner *because* she is an excellent hunter, sailor, navigator.

Leadership in an endogenous setting also brings about the possibility that the leader is well informed *about the players*: she may be 'in touch' or 'know their pain', to use Hermalin's words, pointing to Weber, the sociologist. For instance, you may think about a person with great social skills and knowledge of her friends as one who can organize a great party *because* she knows what will make people happy and they all know she does. Such a party, with the positive external effects between guests arriving with high expectations, can be viewed as a public good. Adding a reputation effect from event to event would reduce the costs of signaling good cooperative opportunities, making even better cooperation – i.e. parties - feasible through lower costs of taking the lead.

The idea that leadership can be built on knowing or associating well with your peers – as a union leader, a civic leader, or a religious leader – of course points to the possibility that leadership arises from the ranks – our type of question, arises between nations - and to motivation by other means than through monetary returns. Then, also using Hermalin's idea of a leader who can 'feel their pain' there can be a role for concepts such as empathy and altruism (on norms, see below) both of which can in themselves contribute to cooperation on public goods provision, perhaps including through leadership.[7]

[7] We do not emphasize, here, models that are political or electoral in nature, but may mention Dewan and Squintani (2013), as an example. They build on Machiavelli's "The Prince" and his "*you can assess a prince by his councelors"* to construct a setting where a limited set of advisors who are close ideologically is important to the quality of leadership, thus alluding both to 'feel their pain¨' and to the value to the leader of having access to information that is not widely held. Rotemberg and Saloner (1993), similarly, concludes "We show that

There are of course important questions about whether insights from leadership models for individuals apply to cooperation and leadership among organizations and states. An indication that part of it does, however, would be if what we see in politicians' behavior when countries meet and negotiate that they try to signal that they understand each other and each other's domestic challenges[8].

There are many contributions warning against taking the lead in global environmental policies, an important one being Hoel (1991), with a common feature being no behavioral or other assumption linking responses in other's emissions to actions taken by the leader. Brandt (2004), adds nuance and contrast to this literature by introducing an element like that in Hermalin's (1998) model. Assuming that abatement costs are correlated, he finds that "transmission of information by use of a unilateral move makes sense". Such correlation directly asks for the signaling role that a leader can take, though in Brandt's model a country that takes the lead may also be the one who most optimistically misjudges the costs of mitigation (an effect much like winner's curse, well known in economic theory).

Hoel and Golombek (2008), in a later contribution, demonstrates that positive spillovers from R&D can eliminate the carbon leakage effect of unilateral action and in some settings create positive spillovers in terms of further emission reductions from others. Combining this with Hermalin type leadership expands the effects and thus extent of moving early: If early movers you not only signal feasibility or low costs, but also bring costs down, either through R&D or through learning by doing/scale economies.

Cooperation on public goods provision and leadership in Evolutionary settings

Evolutionary settings were pioneered for a quintessential problem of cooperation, Axelrod's (1981) pursuits with public goods provision in a repeated prisoner's dilemma. Here, a version of limited (or bounded) rationality is specified by a fairly short program code, and Axelrod showed that when behavioral rules played against each other, a very simple reciprocal strategy called 'tit for tat' performs well.

'Performing well' in an evolutionary setting means that your 'profit' gives you or your offspring higher tendency to survival or procreation. The tit for tat strategy in repeated games is simply 'I start by playing *cooperate* and thereafter play as you do (I cooperate if you cooperate and defect if you defect')[9]. When the player with the tit-for-tat program does well, it is a combination of not losing too much when meeting others who do not cooperate (since she switches quickly to not cooperating) and doing quite well with others who are

---

leaders who empathize with their employees adopt a participatory style and that shareholders gain from appointing such leaders when the firm has the potential for exploiting numerous innovative ideas".

[8] A term in the framework convention on climate change and in the Rio Convention of 1992: 'common but differentiated responsibilities' can be interpreted as a signal that countries understand each other's differences, or know each other's 'pain', and and this feeds into a notion of fairness (see below).

[9] Sugden's (1986) book on cooperation is excellently reviewing and discussion what we know of our cooperative abilities through game theory. Sugden highlights both human purpose (efficiency) and convention in the origin of fairness and institutions (such as property and accept of property), and his emphasis on starting out being nice, in a cooperative mode, is similar to our 'early mitigation', which may be just to start out nice for a conditional cooperator.

cooperative. The understanding of a player as 'conditional cooperator' or of 'conditional cooperation' as a strategy results from this line of research.

Such behavior is harder to incorporate in theoretical settings without bounded rationality. As an example, in a repeated game where players could perhaps cooperate in early rounds because of the value in cooperation in future rounds, if there is a final round then cooperation cannot exist in the final, then not in the one before the final, and so on.

In theory, a higher degree of cooperation can be sustained for a deeper and longer punishment of noncooperative behavior. However, with rational and forwardlooking individuals, it is difficult to see how the door for renegotiation can really be credibly and permanently closed. Though the 'folk theorem' for infinitely repeated prisoner's dilemma games establish that any level of cooperation can be attained for a sufficiently low discount rate, the argument is very vulnerable to questions about the time horizon, finiteness, the credibility of punishing with non-cooperation for ever, etc. We could mention here that for a problem which could involve catastrophic loss (Barrett, 2013, for instance), a punishment strategy of noncooperation for ever is a punishment flavored strongly enough to bring everyone into cooperation. But it might also be, by the same token, lacking in credibility[10].

The difficulty in building an argument about cooperation on punishment that is very costly (and/or lengthy), costly perhaps also to the punisher, is part of the pessimism and agnosticism in economics about how cooperation comes about.

An important direction, though, is to bring in the possibility of norms, much along the lines of limited or bounded rationality approaches (i.e. I'll cooperate because I've be taught to, and we'll all cooperate because we've been taught to sanction noncooperators).

The most obvious application is in coordination games, i.e. games with multiple equilibria, in which the question arises how an equilibrium is chosen. A coordination game like 'should we drive on the left or the right side of the road' has two equilibria (left, left, in the UK; right, right, in continental Europe), and either one could be implemented either by a first mover, if visible; or by a focal point or a norm, if known. Axelrod, inspired by and working with a biologist, could start with the premise that ants cooperate to build an ant-hill for the common (ant-people) good because they are programmed to, not because they resist the temptation to shirk. Such short programs can be thought of as rules of thumb, as norms, and they can of course include short programs for jointly punishing, shunning or eliminating anyone not cooperating or not in line with ant-hill code of conduct.

In most coordination games, if a norm regulates behavior by many, then sanctions are typically not needed: there is no general temptation to defy the norm and drive on the right (hand) side in the UK). In some settings a threat of sanctions is needed, and in these one may

[10] Dr Strangelove's remark in Stanley Kubrick's 1963 film: "Are you saying you made a doomsday machine and you did not tell anybody" illustrates two important poInts: i) in principle, creating and communicating an unbreakable unbreakable catastrophic mechanism (like a massive nuclear counterattack) would entice cooperation; ii) the depth of the catastrophe that itself commands cooperation, also dents the credibility in the mechanism. Would not the builder want to cheat?

have norms about punishing those who do not follow the norm (Denant-Beomont et al. (2007) analyses punishment behaviors). Then, coordination games more closely resemble public goods games, since in public goods games cooperation can be sustained by a norm if people act according to the norm (and selection of equilibrium is like choice of norm). As a theory of cooperation, however, it leaves unanswered how cooperation comes about if there is no theory about how norms are formed[11].

Acemoglu and Jackson (2011) studies norms of cooperation in an evolutionary context in a coordination game (i.e. different from a public goods provision game). This setting has multiple equilibria, with "low cooperation" and socially attractive "high cooperation". Players wonder who they meet, and norms are 'frames of reference' that shape how information about the past is interpreted. Some players are equipped with 'prominence' – or visibility, and a prominent player may choose to break a social norm of 'low' cooperation (a socially unattractive equilibrium) to 'high'.

Their model can display stronger or weaker path dependence, so it is a model which can give weight to a few individuals in the evolution of norms and even even in long term history. It is thus an analysis that allows for individuals to make a difference – the way we are looking for – but which does this by exogenously giving certain individuals 'prominence'. As in Hermalin, the inference others make is part of what motivates leaders. While Acemogly and Jackson analyses a coordination game rather than a public goods game, it is not without relevance given that the equilibria distinguish between higher and lower levels of cooperation, and their idea of how norms are shaped by history and leadership. Also, as we shall see below, public goods provision can be thought of as a question of burden sharing and fairness norms, and alternative fairness norms can transform a public goods game to a coordination game (Kosfeld et al, 2009).

Leadership and cooperation on public goods provision in experimental settings

Cooperation on public goods provision is among the topics of modern behavioral economics that has most been studied in experimental settings, often with students in a laboratory. A brief version of this literature is that students in early rounds of play contribute significantly more to provision than what is indicated by Nash equilibrium (in Nash equilibrium you contribute zero or almost zero), but these contributions decline in subsequent rounds as players observe others not contributing fully.

An interpretation of this very general pattern is that individuals are somewhat influenced by social norms: many want to contribute if others do (but perhaps a little less: they are conditional cooperators). Since the average contribution that they observe reflects some low-providers or noncooperators, they lower their contributions as well. Through the course of several rounds of play, average contributions tend downward, coming close to the Nash equilibrium or simply reflecting some unconditional contributors. This literature studies this

[11] Hume, 1739, writes insightfully about norms and morals as human constructs built on purpose: «no action can be virtuous, or morally good, unless there be in human nature some motive to produce it, distinct from the sense of its morality», but is also convinced they rest on convention, tested through slow paced history.

process: how high is average cooperation in early rounds of play, and how slowly does it decline, and what seems to influence it. The general theme is one of free riding, but not in the sense that a high contribution makes others contribute less.

A good example of this literature, is Fischbacher and Gächter (2010), who investigate the role of social preferences and beliefs. They conclude from repeated play: "contributions decline (in randomly composed groups) because, on average, people are imperfect conditional cooperators. After some time, all types *behave* like income-maximizing free riders, even though only a minority is motivated by pure income maximization alone." In their communication of a central lesson, cooperation on public goods provision is inherently fragile, even if most people are not free riders but conditional cooperators.

The role and interpretation regarding norms is important to Frey and Meier (2004), who report that 'people behave pro-socially depending on the pro-social behavior of others': Students contributed more to charity when they were presented with information that others contributed.

Norms about cooperation are important in themselves in facilitating cooperation, but so is their reflection in our behavior, as through emotions such as anger and guilt. Hopfensitz and Reuben (2009) explore the idea that emotions, such as anger, help enforce and thus support cooperation through norms. They show that sanctioning behavior, too, may be deterred by counter-anger from those sanctioned, unless counter-anger is deterred (by guilt). This lends support to cooperation being facilitated by norms, with some relevance to our question, perhaps related to history and historic responsibility.

This literature thus highlights that if norms are to be supportive of cooperative behavior, they have to be reflected in widely (and perhaps deeply) held values. This has motivated studies of fairness beliefs and norms that is largely positive:

- *what is it that people hold as being fair in the sense that they will cooperate and contribute if the burdensharing is considered fair?*

A background for this question can be the formation of entitlements we presently consider quite evident, such as property rights[12]. Cappelen et al. (2010) find evidence through public goods games that individuals attach responsibility more to contributions in terms of effort than to an exogenous variation in the productivity of that effort. Thus, sacrifice is more important to what is considered fair than a luck factor that makes the sacrifice more or less productive. Cappelen et al. (2012) pursues the question of whether motivation for contributions can be traced to intrinsic moral motivations or to extrinsic factors, such as guilt, shame and pride, finding evidence of both.

Gächter and Riedl (2005) argue that entitlements, often based on a history that is no longer relevant (reflecting claims no longer feasible, for instance) can still represent 'moral property rights' that are influential in negotiations. Kosfield et al. (2009) shows that cooperation to

[12] David Hume, previously cited, explores the origin of property rights and law as sensible arrangements from a practical point of view, important through how – when in human beliefs - they influence behavior.

overcome free riding depends on the formation of institutions for sanctioning. In the mixed success with which these form, a problem is that an institutional coalition that forms is small, thus yielding both inequity and low provision. This is very much in line with Scott Barrett's notion of self-enforcing agreements, typically mobilizing only a very small coalition of players (see below).

Two important implications of fairness beliefs in this context are as follows. First, if many are conditional cooperators (they contribute if others contribute with what is seen as fair), then a joint understanding of fairness can help support cooperation. As it happens, though, most studies show such cooperation – when observed – somewhat fragile, not the least when player heterogeneity makes it less obvious what is fair[13].

The second implication, discussed by Kosfield et al. (2009) is both bad news and good news for cooperation. The bad news is that since fairness motives can act as a constraint on equilibria (page 1354), they may also represent risk of failure in the institution formation process. The good news is also related to this constraint. Not only is it the case that constraints that rule out certain equilibria can be useful in a multiple equilibrium context (i.e. in a coordination game) but they can also create equilibria. For instance, if it is considered unfair to let any small coalition carry the burden of provision, then this may facilitate the grand coalition forming the institution, yielding not only greater fairness but higher cooperation as a result. An important example of the bad news part is that in a uniform prisoner's dilemma, an equity based fairness motive would not rule out the low-cooperation Nash equilibrium. An example of the good news is when Kosfield et al. point out that the small coalitions that can form an inefficient cooperative institutions often *are* inequitable and thus that greater efficiency may result when these are ruled out by fairness beliefs[14].

An important question has been leadership, starting with the observation that if a player moves early and contributes, then this raises the contribution from others relative to what they otherwise would have been (see also the importance of prominent players in Acemoglu and Jackson, above).

Further studies have indicated that voluntary leadership in public goods games entails greater following - or 'positive spillovers' - than does exogenously appointed leadership (Rivas and Sutter, 2011), providing further evidence that followers' are thinking about the leader's motives. This has been taken further in studies that ask whether the compensation of a voluntary leader matters. Cappelen et al. (2013 indicates that compensating the leader can raise contributions, including from followers, but that a very high compensation of the leader may 'crowd out' the contributions from 'followers'. Moxnes and van der Heijden (2003) also finds positive effects of leadership in a public good experiment (an environmental one, in their representation). In a follow up study where they vary the costs to the leader of abatement

[13] Reuben and Riedl (2009) examines fairness norms for contribution under heterogeneity.

[14] This is similar to the problem in cooperative game theory highlighted by Barrett (1994) and Chander and Tulkens, (1995), where it is the productivity of a relatively small coalition that prevents the grand coalition from forming, and anything that rules out small coalitions may be conducive of cooperation. See Battaglini and Harstad, 2013, in which contracting problems harm smaller coalitions more than greater coalitions, so contracting problems facilitate cooperation.

actions (van der Heiden and Moxnes, 2012), they find that there is more of a follower effect when the leader contributes despite costs (holding the costs for other players constant).

While the researchers generally caution about the conclusions that can be drawn for real world environmental policies, they conclude that "it follows from our results and previous studies that taking a leading role will provoke greater reductions in global environmental problems than what follows directly from local abatement".

The literature based on experimental settings, despite the obvious questions of the relevance of the laboratory setting, has become quite influential, much because of the fundamental relevance of such questions as how cooperation in humans come about, and the shortcomings and puzzles left from other research avenues.

We may note that in these studies of leadership's positive effects on followers in the lab, we discover themes both of 'sacrifice' and of 'signaling' as indicated by Hermalin's theoretical model.

3: 'Global problems require global solutions' and the *planner's problem*

The preceding literature review has indicated that support can be found for 'taking the lead', in the sense that early action by some may raise contributions from others, though the settings are admittedly selective, the mechanisms may not all be known well, and are apparently vulnerable.

In this section, we examine the basis for the more well-known warning against early, unilateral unconditional emission reductions. We start by visiting the 'global problems require global solutions' idea, which typically is seen as supportive of countries that are willing work for stronger international cooperation, rather than acting alone.

The planner's problem

A simple version of the problem of tackling climate change is that global emissions (of greenhouse gas emissions, GHG) in this century have to be held to a certain limit[15]

$$\sum_j \sum_s \sum_{t=2000}^{2100} e_{jt} \leq \bar{e}.$$

[15] A focus on an emission budget for the century – or even for its first half, works fairly well for practical purposes. A recent contribution concludes that a centennial budget for six greenhouse gases is a good predictor for the likelihood that climate change beyond two degrees Celsius can be avoided (Anderson and Bows, 2011). Meinshausen et al 2009 uses cumulative emissions until 2050 as a predictor of this likelihood. Two degrees Celsius above preindustrial levels is a widely heralded policy goal, adopted by the international community through the 2009 Copenhagen Accord and for instance by the EU. The main point in our analysis, possible functional dependencies from early actions to later ones, and from some nations to others, does not depend on this simplification: main points will survive extension to more sophisticated models of our impacts on the climate. The most obviously useful extensions of this modeling framework is a depreciation of greenhouse gases in the atmosphere (which would lead to a greater acceptance of emissions early) and a damage to climate change before a presumed limit (which would lead to less acceptance of emissions early). The latter, a damage function, is introduced in section 5. Aaheim is an example of analysis in this directions, as is integrated assessment models such as Nordhaus' DICE (2008).

Here, the summation over countries *j* could be simply North and South or it could be over the world's 190 or so countries, *s* could be two or more sectors or all emitting sources. For instance, summation could be over all cars and homes, all firms and farms, all goats and dumps, all trees, all the twenty billion people that are around in the century if indeed we peak at nine or eleven billion. The summation over the century could be over 100 years as indicated here, over twenty five-year periods to indicate periods of quota commitments, for instance. We shall be simplistic and just use *j=N,S*, for countries and *k* and *l* for two sectors (or technologies, initiatives, projects, policies), where *N* shall typically be something like Norway or Europe or Annex 1 countries (or those that have committed to a cap for emissions for the years 2008 through 2012), and *S* can be 'rest of the world' or 'South' or India or China or Asia or USA.

Reducing emissions will have a cost to mankind[16]. There are benefits too, but these are in this simplistic treatment all represented in our constraint on global century emissions. We may simplify to assume there are no benefits to reductions in climate change within the constraints and very high benefits of a avoiding any breach, so a breach is not considered.

The foundation of the 'cost effective solution', and its implications

A simple formulation may be that the world wants to minimize the net present value of the costs of containing emissions in this century to the limit for emissions, $\bar{e}$:

1: Minimize $\sum_{j=N,S}\sum_{s=1}^{2}\sum_{t=0}^{20}\frac{c_{j,s,t}(e_{jt})}{(1+r)^t}$ subject to $\sum_{j=N,S}\sum_{t=0}^{20} e_{jt} \leq \bar{e}$

Thus, we want to find a set of emissions, or quantities emitted

$$\begin{bmatrix} e_{N,1,0} & \cdots & e_{S,2,0} \\ \vdots & \ddots & \vdots \\ e_{N,1,20} & \cdots & e_{S,2,20} \end{bmatrix}$$

where there are columns as many as (two) sectors times countries, and where rows are time periods, here considered twenty five-year periods in this century (or the seventeen that are left: three rows already have numbers, emissions, fixed).

The solution to problem *1* is a set of quantities satisfying three types of optimality conditions

i) $\frac{\partial c_{N,1,0}}{\partial e_{N,1,0}} = \frac{\partial c_{N,2,0}}{\partial e_{N,2,0}}$,

or equal marginal costs across sectors (also actually more generally also across alternatives within sector, or activity). It applies generally, too, but we could first emphasize how this

[16] The basic idea that such a cost follows from optimization theory is exploited for exhaustable resources in Weitzmann, 1999, who concludes that exhaustable resources are worth about one percent of world GDP. A heterogenous world and large reductions required strengthen the view widely held that costs will be considerable, and rent implications too. Energy subsidies indicate some emission reductions will be cheap.

applies internally within any group of countries acting on emission reductions, to call it *cost effectiveness in the North, now*.[17]

A second implication, then is:

ii) $$\frac{\partial c_{N,s,0}}{\partial e_{N,s,0}} = \frac{\partial c_{S,s,0}}{\partial e_{S,s,0}}$$

This idea, which we could give the name *global problems require global solutions*, suggests that the world should take advantage of cheap emission reduction possibilities everywhere.[18]

Finally, a third condition is:

iii) $$\frac{\partial c_{N,i,1}}{\partial e_{N,i,1}} = (1+r)\frac{\partial c_{N,i,0}}{\partial e_{N,i,0}},$$ and similar for later periods[19].

We might call this condition *we have started early.*

The intuition here is that emission reduction efforts have been introduced rather forcefully, so that only moderate cost increases is expected. This is the Hotelling rule for optimal use of an exhaustible resource, and many will have noticed that the formulation (1) above is describing the atmosphere's ability to accept greenhouse gases simply as an exhaustible resource[20].

Despite the equality sign in iii), we emphasize the 'urgency' part of this condition: the expected cost of emissions should not rise by more than the real rate of interest, and for this to happen it must have risen to a certain level in the outset. If it rises by more in the future, the world foregoes early emission reduction possibilities and thus will have to attain them later at

[17] Such a requirement also can be said to be reflected in instruments such as the European emission trading system, ETS, for instance, as well as the accompanying idea that EU's emitting sources that are not in included in the ETS, should mitigate so as to have the same marginal costs as within ETS. Despite this ideal, in practice, even internally within ETS sectors there will not be equal marginal costs, because of such differentiated supplementary policies as support for renewable energy. See, for example, Böhringer et al. (2009), Eskeland, Mideksa and Rive (2011), and uniformity across non-ETS sectors is an equally elusive goal. We shall argue below that for a subset of the grand table of global century emissions (like ETS emissions between 2012 and 2020) it is difficult to arrive at a firm analytical foundation for a cost effectiveness criterion like this.

[18] This condition may be said to have inspired elements in international negotiations such as the flexibility mechanisms (allowance trade, clean development mechanism, and joint implementation). In that context, the fact that some countries are not or might not for some time be bound by emission caps should not prevent the world from seeking emission reductions in those countries, should reductions there prove to be cheap.

[19] Aaheim (2010) analyses the possibility that the discount rate itself is influenced by the climate problem.

[20] To describe the atmosphere as an exhaustible resource and to do the same for fossil fuels is the same as to state that we either have a problem of limited fossil fuels *or* one of climate change: we cannot have both problems. So acceptance of climate change as a challenge is the same as to say there is room for less carbon in the atmosphere than what we have and might otherwise want to burn of fossil fuels in this century. If the constraint of available fossil fuels were binding the challenges of building institutions would be simpler since, for fuel deposits, 'exclusion' and rivalry in consumption would make adherence incentive compatible. For an atmospheric constraint, building institutions that ensure abiding by the constraint is more challenging, not the least because the atmospheric constraint bears the characteristics of nonexcludability and nonrivalry – of 'commons' or of 'open access'.

This condition, if we expand the model of the climate with a depreciation of carbon in the atmosphere, will be modified so that the shadow price of emissions increases with the real interest rate less the depreciation factor. So, if the real interest rate is 1.5% and the depreciation of greenhouse gases in the atmosphere is 0.5%, the shadow price of emissions should be rising by one percent per year.

an unnecessarily high cost. Not to regret past emission reduction possibilities of course requires that – at the formation of a cooperative arrangement – a treaty - the costs have quickly shot up to a sufficiently high level, so further increases will be moderate.

The three conditions imply that the world should seek emission reductions everywhere so that no low-cost opportunity is foregone when costlier ones are pursued. As we shall argue in the following, in a world of nonglobal cooperation, criteria other than low costs of emission reductions will come into consideration.

4: Pondering the problem of early mitigation in the absence of a strong global treaty

We may start by mentioning a problem of applying the cost effectiveness criterion above if only some countries subject themselves to emission constraints. If in such a setting *all* emission reductions were sought globally, not only could taxes (or quotas) be used, but the tax rates (or quota prices) would be very low. Thus, there would be extremely many 'intervention points' (say, fuel taxes of a third of a cent per liter or kilo everywhere in the world). That would be how to apply the above cost effectiveness solution with a limited 'budget'.

So one might ask: in a world that is initially not very ambitious in terms of global emission reductions, would it all be about getting emission reductions globally at the lowest possible cost as indicated above, or would it (also) be important to demonstrate that larger reductions are feasible, for instance within some countries, some sectors, some settings?

The solution to problem *1* satisfying conditions i) through iii) above, is a table of emissions (a set of quantities) as follows

$$\bar{E}_{G,0\ to\ 20} = \begin{bmatrix} e_{N,s,0} & \cdots & e_{S,s,0} \\ \vdots & \ddots & \vdots \\ e_{N,s,20} & \cdots & e_{S,s,20} \end{bmatrix}.$$

In that formulation of the global problem, the important aspect for the climate is the sum of emissions both across global columns and century rows; the distribution of quantities across columns and rows is chosen only to minimize the discounted costs to mankind of not stressing the climate more than assumed acceptable or possible.

Duration; years arrive in a sequence

We have above described the climate challenge here as that of limiting the sum of emissions in this century to a given constraint. There is no complacency in this; for a given tight limit acceptable to the atmosphere, it could imply emissions were almost eliminated by the middle of the century (important scenarios, for instance the 2 degree ones, demand emission reductions of 80 percent by 2050: if the world economy triples, its emission intensity must fall not by four fifths, but by fourteen fifteens).

Nevertheless, almost a century – or even a half – is a lot of time in important ways. First and foremost, it is enough time that if some countries were to make important choices, like emission cuts or power plant investments, they and other countries would have subsequent choices before time is out. Those subsequent choices would no doubt be influenced by that

history. As noted in section 2: if later actions are influenced by early actions, then this will influence early actions. This is what we want to acknowledge.

To make a simple calculation, if world emissions in the business as usual case grows by one percent per annum (one percent improvement in carbon productivity – or carbon intensity- as Stern suggests, combined with two percent economic growth), then in addressing an early five year period in the century, one addresses only three percent of this century's 'business as usual' emissions. If these emissions are cut by five to twenty percent in a region representing a fifth of global total, then the world is reducing this century's global emissions with three hundredths of one percent to two tenths of one percent[21].

The less obvious observation, at the heart of this paper is:

Observation1: Any early emission reductions are of importance if and only if they have consequences for future emissions

Looking at the numbers, this is a rather obvious point, not a very subtle one. Any early period of five to fifteen years is long enough to cause observation, rethinking, and new actions, but anything a subset of countries does in some early years would have its motivation in consequences for future trajectories of global emissions: own and those of others. With a limit for global century emissions, all that matters is shifting from one trajectory of growth to one of substantial global decline quite soon. In a world of underlying growth that can be done only when actions to cut emissions instigate further actions to cut emissions.

A point that follows from observation 1 also is fairly obvious: Early actions other than emission reductions that could have consequence for the trajectory of emissions through this century should be examined with interest equal to early emission reductions. In other words: as soon as we leave the artificial world of one global planner, the focus in formulation 1 on emissions alone is far too restrictive. As an example, teaching sciences, natural and social, in primary schools from Kenya to Kentucky, is on menu of early actions that go beyond only emission reductions.

Participation; no country, no sector is acting in isolation

Early reducers, for instance in Europe, presently represent a minor share of *current* global emissions (a seventh, or a fifth). Still, *their* emissions as a share of global will decline through this century:

- if only because other countries' have higher growth rates, or higher growth rates in emitting sectors, because of population growth, of catching up, of have higher emission propensities, less advanced technologies;
- as a direct consequence of the measures these early reducers take;
- and to the extent that the measures cause carbon leakage (see below).

[21] This could be Europe's reductions in the first commitment period of the Kyoto Protocol, 2008 through 2012, or some version of the Europe's ambition of 20 percent reduction in the eight year period from 2013 to 2020.

Observation 2: Measures among a set of 'early reducers' that is less than global – be they emission reductions or other actions or consequences – are of importance for the climate only if - and in the way that - they influence emissions from other countries / sectors.

Again, this is an obvious consequence of economic development, energy being normal in the sense that a country tends to use more as it gets richer, and fossil energy being sufficiently abundant.

Observations 1 and 2 cut across rows and columns, respectively, in the table of the century's emissions. Present approaches, such as the tradable allowances for Annex 1 countries in the Kyoto Protocol, and also ETS, a system for trading quotas within European firms, can be thought of as windows to allow – within period and between early reducers – some of the gains from cost effectiveness to be played out between columns (Annex 1) and rows (2008 through 2012) in the table.

Our point, here, is that – given *Observation 1* and *2* - we have no result in economics nor in optimization more generally giving a compelling rationale for this equalization of marginal costs within a subset of cells, say of acting countries in a five or ten year period. If some types of cuts or other actions were to have greater consequence for global emissions than others, then those would obviously be favored.

In practice, countries within EU work harder in changing the power sector than in changing the metals sector (the power sector has quota price plus renable support). This could be biases caused by the way actions must be phased in a world with nonglobal cooperation.

Carbon leakage: important phenomena

While one may think of a table of emissions for this century, $\bar{E}_{G,0\ to\ 20}$,like 1) it is not so obvious what one would mean if describing a table of emission $\bar{E}_{N,0,t}$ where countries *N* had restrained their emissions (perhaps internally cost effectively) for some years *t* (say five, or eight, or until 2050) while the rest of the world were stipulated with some 'business as usual' emissions. What is 'usual' for the rest of the world and the rest of the century if part of the world acts forcefully in some way?

Carbon leakage is the fact that if a country reduces its emissions – uses less fossil fuels or produces less cement than it otherwise would - emissions in other countries may increase, though typically by 'less than 100 percent' of the initial reduction[22]:

To detail the mechanisms of carbon leakage, we might use a list of three points as follows:

[22] Hertwich and Peters (2009) provide estimates on carbon leakage through evidence on countries' changes in consumption, production and trade. Bernard and Vielle (2009) quantifies leakage and Frankel (2008/9) discusses remedies. Sinn (2008) and Harstad (2012) analyses supply side measures as alternatives / supplements to demand side measures in a world of nonglobal cooperation. Edenhofer and Kalkuhl (2011) and van der Ploeg and Withagen (2009) analyses the existence/importance of the 'green paradox'. The case of carbon leakage rates greater than 100 percent is possible in theory through energy intensive tradables like metals or cement, if production moves to jurisdictions with more emission intensive practices. In applied analysis, such high leakage rates are not found, more often they are found in the range of five to fifty percenet for energy intensive tradables (Böhringer et al. 2009, Eskeland et al., 2012, are two examples).

i) By lowering world prices for fossil fuels, reductions in demand for fossil fuels among some will lead to increased use among others, as long as fuel supply is not infinitely elastic (which it can hardly be);
ii) Emission-intensive activities – especially production of energy intensive tradable goods such as cement, metals, fertilizer, glass, ceramics, chemical products – may relocate to countries with no emission reduction policies (or costs), without much reduction in global consumption or emissions;
iii) An expectation of future emission reduction policies threatens future fossil reserve rents and creates a race that accelerates present extraction.

Interesting about iii) the "green paradox" – due to Hans Werner-Sinn – is the possibility that *emerging* or *expected* climate policy *increases* emissions in early periods, by causing a race for threatened fossil rents and thus accelerating climate change. Indeed, these instances of carbon leakage are consequences of the cost effectiveness criteria in section 3 being violated. Sinn's green paradox, for instance, follow when a too steep emissions' price is expected.

Observation 3: Realizing that emissions policy in some early years for some countries is motivated by global century emissions, early contributors will not restrict attention to own emissions and their costs or to total emissions and costs in those early years.

First, we should notice that the consequences for the century's global emissions will depend not only on early emissions but also on how early emissions are reduced. Second, global century emissions may be brought down in this century by choices made early – R&D investments for instance - without emission reductions those early years. For both of these, behavioral consequences amongst other countries will be important: these may be dependent on cost effectiveness, but not solely.

Hoel (1996), for instance, asks whether carbon taxes (quotas would be the same) for participating countries should be differentiated across sectors due to carbon leakage, answering that this is not desirable provided one can use tariffs on exports and imports. If such trade policy measures cannot be used, carbon tax differentiation across sectors would be desirable. Also, in a world of nonglobal cooperation, participating countries would address both supply of fossil fuels and demand, while under global cooperation, addressing one side would suffice.

Retracing the argument a little, there are good reasons to think of emissions in some early years, or early emission reductions, as reducing global emissions in this century. Such reductions may be tailored to avoid carbon leakage, or keep it acceptable levels. The point more generally stated is: influence from some nations, some years, on global emissions in this century can in no meaningful way be summarized through emitted quantities in those countries, those years.

The practical part of the *how* argument is simple: Carbon leakage means that a national emission reduction often will only in part result in a global reduction. Leakage can be limited, both by how a country goes about its emission reductions, and by accompanying policies (see Frankel, 2008/2009, for instance, or Fischer and Fox, 2009). On the first, leakage will be

smaller if you reduce emissions in rather protected sectors, like the heating of buildings, domestic transport, and electricity generation, than if you reduce emissions from producing energy intensive tradables, like steel. And leakage from emission reductions in energy intensive tradables will be smaller if production persists – perhaps with help of trade policy measures such border tax adjustments, or free quotas – than if you scale down output[23]. And smaller if you support electricity generation with renewable support than if you let emission costs drive electricity tariffs and output. And smaller if you invest in R&D than if you reduce output of emission intensive tradables. Nevertheless, a part will likely remain which is conveyed through world fossil fuel prices (as Hans Werner-Sinn rightly emphasizes). It is the steepness of the supply curve for fossil fuels that creates leakage, in combination with the low price elasticity of demand. At more general policy levels, leakage can be fought with trade policy measures: if you export emission intensive products – like metals – into participating countries, you'll be paying 'border tax adjustments' that counter your cost advantage. Similarly, as analysed by Harstad (2013), Hoel (1994) and Sinn (2008): participating countries will want not only to address demand but also supply of fossil fuels.

5: Supply of public goods and the free rider problem

An economic perspective is that climate change mitigation – or emission reduction efforts – is to provide a public good, with key contributons by Paul Samuelson (1954), Charles Tiebout (1956), Agnar Sandmo (1972). Samuelson points out that optimal supply requires that the sum across beneficiaries of marginal benefits to equal to marginal costs, Tiebout that exclusion may facilitate provision for local public goods in a market context, and Sandmo (1972) develops the case for collective inputs, which seems to be a relevant case for climate change (impacts are often described as harming agriculture, infrastructure, buildings, etc., see IPCC, for instance, or Stern, 2007).

The standard textbook treatment of public goods provision is that a free rider problem arises: Even if everyone wants the good supplied, individual agents will provide less individually than what together is required for optimal supply:

i) She is enticed to provide only according to her own marginal benefits of her provision, not taking into account the benefits to others (she is, inter alia, not able to charge others for the benefits accruing to them due to her efforts).
ii) In fact, if one agent is providing so as to increase the level supplied of the public good, others might be induced to reduce their provision below what they otherwise would have provided.

Characterizing an equilibrium without cooperation – as a Nash equilibrium in quantities, for instance – i) typically involves each agent taking into account only her own share of benefits.

[23] Free quotas could in theory be nondistortionary – not acting as a subsidy to output – if they were given on the basis of exogenous characteristics, but as a flow they will typically disappear if a steel producer ends being a steel producer. Thus, free quotas may be interpreted as proxying what border tax adjustments would do: limit the carbon leakage that would happen if steel producers had to buy all their quotas. See Harstad and Eskeland, 2010 for an analysis of quotas as a signaling game, recognizing politicians' commitment problems.

For the climate problem, one could think of a country taking into account only that country's (or also only a present generation's) share of total global benefits, so this could mean emission reductions were undervalued to the amount of

- one two hundredth of global benefits if benefits were equal for two hundred equal countries,
- one twelve hundredth if benefits were equal for each of six generations of two hundred countries, or
- a country's present share in global population now if benefits were in proportion to population (a fifth, for India or China, a twentieth, for US)
- or a country's share in global income, agriculture, etc[24].

In any of these versions, noncooperative provision of the public good would be very far from globally optimal, and on an axis from zero mitigation to optimal, it would be for most countries much closer to zero – indistinguishably from zero, perhaps - than to socially optimal.

Due to this free rider problem in public goods provision, government provision (or coercion, as with regulation and taxes) is typically recommended for such subnational or national public goods as local air quality, rule of law, and national defense.

For a global public good like climate change mitigation, however, even without authoritative global government, sovereign states can achieve provision – mitigation - through a treaty. In a situation with two players, for a public good that is worth providing collectively, an enforceable treaty can ensure optimal provision.

For more players, or countries, as in a world with near two hundred countries and more than six generations of beneficiaries necessary to justify serious dent in emissions[25], many would prefer not to participate in the treaty, rather hoping others to pull the load. Using standard concepts in cooperative game theory, Scott Barrett (in his 2004 book, for instance), demonstrates that in games with multiple players (the number *n* of players three or higher) participation in a treaty will be low, and provision of the public good will be low. As in the case of noncooperative game theory, above, provision of emission reductions will be much below what is optimal for society as a whole, if not actually zero. In Barrett's (2004) words, when the number of players is large, "and the gains to cooperation are large, a self-enforcing treaty cannot sustain a high participation rate and so is unable to make much of a difference".

The problem is, in light of cooperative game theory, quite deep and untractable. If the structure of payoffs is really 'all or nothing' so that it is detrimental not to have global

[24] Hannesson looks into some consequences for treaty formation, selfenforcing treaties, of alternative benefit assumptions, like these.

[25] Stern's review reflects that benefits of mitigation in the second and third century of this millennium are necessary to make a case for mitigation now, requiring a discount rate at 1.4 per cent per annum.

participation, then global participation in provision is facilitated in a self-enforcing treaty[26]. But the more typical structure of underlying economics for public good provision is that some coalitions smaller than the grand coalition can attain a significant portion of the gains to global participation. In this case, since the productivity of smaller coalitions eliminate an "all or nothing" structure, a setting with partial cooperation – perhaps as few as two players – will prevent larger cooperative coaltions. However, since in a game with more than two players there will often be more than one potential coalition that could contribute with emission reductions, even this contribution will be small or late or doubtful[27].

Game theory in general is quite pessimistic about cooperation, and one could say: does not know a lot about how cooperation comes about. Important ideas are reminiscent of trying to convert a game into an 'all or nothing' game, and the logic is simple: Free riding 'on the bus' reflects appreciation of the bus service *and* a belief that the bus will be running independently of whether *I* pay for riding. If it were credibly the case that the bus would not run if I did not pay, then payment is a certainty and free riding is not a problem. A framework for cooperation can thus be described as an attempt to create such all-or-nothing structures[28].

We can use this example, though, to point back to the literature that uses norms and fairness to structure the analysis of cooperation. A Kantian imperative, for instance, can be seen as a norm that you cannot consider not paying on the bus unless you consider nobody paying on the bus. In that case, it is clear to you that if nobody pays on the bus, then the bus will disappear, and you rather pay. Thus, if a Kantian imperative is imposed, then one has converted the bus existence game into an all or nothing game, and the bus can exist. Bus services and other cooperative ventures will thus be more frequently observed on planets were fairness norms (like Kantian imperatives) are regulating behavior of individuals, generations, and countries.

From a player's perspective, an attempt 'to convert the game' can be if she can commit to a strategy of 'I'll cooperate if you cooperate'[29]. Indeed, in a one shot game, if players can

[26] The role of 'all or nothing' type games is discussed in Barrett 'Self enforcing international environmental agreements' (1994). For climate change, all or nothing can be envisaged as an edge or a cliff or a tipping point for catastrophic climate change, see for instance Barrett (2013) and Barrett and Dannenberg (2012).

[27] A good illustration of this is a game with three players with benefits of global cooperation 1 compared to zero for noncooperation and alfa to any twoparty coalition. In this case, if alfa is greater than 2/3, so that a nonglobal coalition is also *quite* productive, cooperative game theory is unable to argue for global cooperation (in technical terms: the core is empty). If nonglobal cooperation is not quite productive, alfa smaller than 2/3, the game is an all or nothing game, and the game theoretic case for global cooperation actually coming about is quite compelling. With a higher number of players, in the analytics of cooperative game theory, cooperation is more difficult to attain. More specifically, for more players, the closer must a game be an all or nothing game for the core not to be empty.

[28] Wood 2010, presents and summarizes results, developing the problem from the perspective of mechanism design. He points out that concepts ruling out smaller coalitions (as with Chander and Tulkens, 1997, see Riedl et al, cited above) may be questionable in terms of credibility.

[29] The US Senate's Byrd Hagel resolution (it will not be the sense of the Senate to ratify an agreement if it will "mandate new commitments to limit or reduce greenhouse gas emissions for the Annex I Parties, unless the protocol .. also mandates .. to limit or reduce greenhouse gas emissions for Developing Country Parties within the same compliance period ") and EU's declaration (to reduce emissions 20 percent by 2020, and 30 percent if important other countries/regions also take part) both appear to communicate something like conditionally matching reductions, much like this line of reasoning in game theory would convey.

commit to such a strategy ('I'll reduce emissions if you do') and a contract can be made enforceable, then cooperative game theory predicts global cooperation. But is such a strategy, such a commitment credible? Many analysts have looked into this, including for repeated games, and it is not easy to see that such a formation of cooperation is credible. Importantly, adherence to cooperation must be sustained by a credible threat of punishment for noncooperators ('defectors', or 'emitters'), and to attain high levels of cooperation, such punishment must be both deep and long and unforgiving. This is indeed very difficult to envisage for a problem such as climate change[30].

For climate change mitigation, it may be that some significant subset of players (US, Europe, and China, say, or two of these three) in terms of early action could achieve quite a significant part of what a global coalition could achieve. According to cooperative game theory, this fact makes cooperation difficult to achieve – at least early in this century - simply because nonimportant and important players want other important players to pull the load[31]. In Barrett's (1994) framework, since a coalition of two often will be the largest feasible irrespective of the total number of players, cooperation is hard to envisage even between these three prominent players[32]. Our argument turns this around: since it is clear that even a quite big coalition with early action can do very little with the climate problem unless almost everyone joins, early action cannot itself much harm global cooperation.

Unilateral and unconditional emission reductions by a country or a small coalition

Whether a small coalition forms or not (one may argue that a country or Europe or the US or the North Eastern United States are such coalitions), an important question is whether a consequence of its unconditional unilateral emission reductions would be free riding in the explicit sense that other countries then make *less* of an emission reduction effort in response to this. In this discussion, we shall abstract from 'carbon leakage', dealt with above. We shall simply examine free riding in response to early unilateral unconditional emission reductions motivated by reduced (threat of) climatic change.

As a starting point, one may assume that another country *j* acts according to a one-period noncooperative condition, in the sense of taking other countries' quantities of emissions as given.

[30] In infinitely repeated games, high levels of cooperation can be sustained through such threats if discount rates are low. This observation builds a bridge between the low discount rates that are required for mitigation in global cost benefit analysis (key contributions are the Stern Review and Weitzman, 2007), and those required at the individual or country level to attain cooperation. Weitzman (2007) argues an insurance perspective rather than the usual one in savings and investment models to render mitigation benefits sufficient weight. Tol and Yohe (2006) and Sterner and Persson, 2008, discuss the discounting in the benefit cost analysis of the climate problem.

[31] Hannesson explores Barrett's type of question with heterogeneously sized nations, still finding mitigating coalitions small, and especially so if countries large in emissions and not the same as those large in benefits from slowing climate change. These findings are supportive of our argument below.

[32] Battaglini and Harstad (2012) analyse an aspect of climate change agreements when a holdup problem relates to noncontractible investment. The noncontractability harms smaller coalitions more than larger, and can thus – counterintuitively - facilitate more cooperation.

To make things simple, assume emissions are the same as the use of fossil fuels, a tradable. A country *j* for this given period maximizes its welfare π with respect to fuel use *q* in this period less its share *v* of damages *d* from climate change. These damages should be thought of as the sum of discounted damages in the future – so emissions Q=$\sum_j \sum_t q_t^j$ are summed over countries and periods in this century:

$$\pi^j = f^j(q^j) - pq^j + v^j d(\sum_j \sum_t q_j^t)$$

Here, *p* is the world market price for fossil fuel, and abstracting from carbon leakage means we assume *p* a constant. If fuel is supplied infinitely elastically, then there is no leakage if a country reduces its fuel use. Damages $d(Q)$ are damages from climate change, so first and second derivatives are both negative, possibly catastrophic, so *d*'s first derivative may approach minus infinite for large finite Q. We can think of the model used earlier, with a constraint for total emissions (and not damages within the constraint) as being a simplified version of this model, or alternatively as an implementation formulation meant to avert risk if damages *d* reflecting risks like catastrophes are the motivating ones (see Barrett, 2013, on how catastrophic risks is relevant to cooperation). Alternatively, the correspondence between the two models could go the other way: if there is an absolute limit, but *d(Q)* could represent the probability of not satisfying the limit, thus making a continuous and differentiable version of the absolute limit used in problem *1*.

A simple setting is one in which country *j* maximizes π^j in this period taking emissions from other countries and in other periods as given. The first order condition for *j*'s optimal emissions (or fuel use) is then

$$(1)\ \frac{\partial f^j}{\partial q^j} - p + v^j d' = 0$$

The case with the third term equal to zero represents an important reference point. A country *j* that does not reduce its own emissions beyond the point where marginal productivity of fuel equals its cost, $\partial f^j/\partial q^j = p$, could be one that

a) does not have, perceive or pay attention to future damages from climate change, or
b) one that considers its share v^j of these damages to be zero, or
c) one that considers its influence on climate change through own emissions to be zero.

c) could be because its share or present share in global century emissions are very small. Alternatively, somewhat outside the formulation used here, it could be a country that considers other countries will eventually make up for its emissions, whether they are raised or reduced, but this we shall leave to the discussion of alternative behavioral assumptions.

In all these cases (a,b,c), country *j* uses fossil fuel to the point where its marginal product equals its world market price:

$$(2)\ \frac{\partial f^j}{\partial q^j} = p.$$

Let us notice that it is only thorough the third term, $v^j d'$ - the term that is zero in cases a) through c) that *j*'s optimal emissions can possibly be influenced by other countries' emissions[33].

Now, if we allow for the (free riding) possibility that q^j responds to other countries' change in emissions dQ^{-j} , then $q^j\left(Q^{-j}\right)$ would respond such as to restore (1) above:

$$(2)\ \frac{\partial^2 f^j}{\partial q^{j2}}\frac{\partial q^j}{\partial Q^{-j}}dQ^{-j} + v^j\frac{\partial^2 d}{\partial Q^2}\left(1+\frac{\partial q^j}{\partial Q^{-j}}\right)dQ^{-j} = 0, \text{ or}$$

$$(3)\ \frac{\partial q^j}{\partial Q^{-j}} = \frac{-v^j\frac{\partial^2 d}{\partial Q^2}}{\frac{\partial^2 f^j}{\partial q^{j2}}+v^j\frac{\partial^2 d}{\partial Q^2}}$$

Let us remember from above, *a*,*b* and *c* the important cases when there is no response in emissions from *j*, so $\partial q^j/\partial Q^{-j}$ is zero. (3) highlights that in the case when it is not zero, it is strictly negative (the numerator is positive by assumption, the denominator negative by the 2nd order condition for *j*'s optimum), meaning that an emission reduction from other players instigates an emission increase from *j*, and vice versa. According to (3), such a case exists only for a country *j* acknowledging damages to itself from climate change (the numerator) *and* seeing itself able to influence climate change. Morover, *j*'s response in emissions has minus *1* as a limiting case if the marginal productivity of emissions to *j* is constant ($\partial^2 f^j/\partial q^{j2} = 0$). This argument holds also if *j* is a coalition of countries (to go through it, it simplifies to think in terms of just two players or groups).

In the situation we depict here, we exclude carbon embodied in trade, and then the important use of fuel (emissions) is for 'domestic' purposes, such as transport – our cars -, buildings – our homes and offices -, and electricity generation – for our light bulbs and fridges. In those uses, the marginal productivity (or the value of the marginal product) of fuel is strictly declining, and a free riding response in the sense of (3) will be strictly smaller than 100 percent ($\partial q^j/\partial Q^{-j} > -1$), and equal to zero in cases a, b and c, above.

We now conclude with

Observation 4: In response to early, unilateral and unconditional emission reductions, a country (or several) may free-ride and raise its emissions, but only if

i) those emission reductions are large enough that it perceives the prospects for climate change are reduced, AND

[33] Hoel, 1991, formulates 'unilateral action' by explicitly changing the country's reaction function (country acts as if v^j is higher, in our formulation), showing that global emissions will typically decrease if a lowcost country makes a small unilateral move from the Nash equilibrium base, and also under cooperation in the Nash bargaining solution case. Obviously, unilateral action will be costly and in some cases raise emissions relative to a treaty, which is assumed efficient. Our concern here is different for two reasons: i) we are concerned about what happens before or without a treaty, and what can influence a treaty being entered, and ii) we consider the case Hoel excludes "I have ignored the question of whether unilateral emission reductions by one country may lead to similar behavior by other countries'.

ii) the country itself perceives and pays attention to climate change and damages to itself, AND

iii) considers that it has through its own emissions influence on climate change.

In the case all of these three are satisfied for one or more responding countries, the increase in emissions will jointly be limited not to exceed the initial reduction, and will typically be strictly smaller[34].

An informal way of putting observation *4* is that a country or a coalition to be deterred from early unconditional unilateral emission reductions through concerns for inducing free riding would have both to think about itself as being big in terms of emissions *and* about potential free riders as big in terms of their emissions *and* vulnerable to climate change. This is of course not an impossible scenario, though it is a fact that many nations who are vulnerable are not big in terms of emissions now (vulnerable nations are poor nations, according to IPCC), and – as argued earlier - most who consider emission reductions now are not really big enough in terms of centennial emissions that they come anywhere close to meeting the climate challenge on their own.

Along these lines, it would be countries or coalitions that are big in terms of global emission reduction potential in this century and considering to take the lead that could worry about other's free riding responses. It is possible, thus, that nonlarge countries would have an edge in leadership, simply because they need not consider that others see them as fixing the climate problem, while they could see them as interesting test cases.

Speculation might be easier using names, here, but in terms of methodology a question is whether energy taxes, emission costs or other domestic policy variables could serve as an indicator of a country's own concern for climate change impacts. Clearly this can be misleading: a) a country may have no or weak climate policy not because it is not interested but for other reasons (as the Byrd Hagel resolution indicates, for instance) and b) because a free riding response in the sense we analyze here could be in terms of reduced *future* climate policy. Still, if early action is leadership, leadership could perhaps lead into stronger treaties, limiting the deterrence that lies in concerns about induced freeriding further into the future.

A concern for inducing free riding is not farfetched, and not farfetched as a deterrent to early unilateral unconditional action. But we should notice that our analysis here indicated that such concerns would be less among smaller countries, so it it perhaps the case that EU in this regard, if it is an early mover, is small or acts as a bundle of small countries (its share in century global emissions is, in fact, not large).

We have, however, argued above that it is possible through behavioral models to envisage responses in the opposite direction; that early movers through norms or signaling can stimulate rather than threaten the subsequent emission reductions from others. Reading the two sections together, we might conclude either that smaller countries or coalitions in so case may have an advantage in taking a leadership role, since they have less reason to worry about

[34] The 'strictly smaller' part of this statement applies also if one includes carbon leakage in the response.

inducing increases in free riding, or – but this is clearly a judgment case and more speculative – that even big countries and groups as large as Europe or USA are not big enough to worry or to be worried about free riding responses induced by early action in a few early decades of this century.

6: Discussion: can a leadership perspective apply to early contributions?

Is climate policy about leadership?

Unilateral and unconditional emission reductions in an early phase without a treaty ensuring global participation (or under a weak treaty) involves difficult questions. We have argued that countries moving early may be motivated directly by *leadership,* i.e. that others will be enticed to emission reductions because of those early actions.

A leader that signals through her early actions likely conveys her beliefs that emission reductions are feasible, not prohibitively costly, are adoptable more broadly, and are worth it. We believe the latter part – worth it – is not the most important part of signaling in the sense of private information, but it obviously a part of what goes on. Hermalin, who describes the private information held by the leader in public goods provision as information about the productivity of a public goods project, finds that signaling can be through example or through sacrifice.

Fairness beliefs and norms also can govern public goods provision, including early action. If players are conditional cooperators, they will be observing others' actions and can be moved positively to contribute more than they otherwise would if there has been early mitigation.

Both in the case of signaling and in the case of fairness beliefs, the possibility of instigating a following motivates leadership, and shapes it.

Three lines of cautioning should not be lightly dismissed.

First, early movers will be concerned about carbon leakage, seeking to limit it, accepting it only to some degree. Leadership is, after all, about making others support your direction of emission reductions, not about inviting others to undermine them.

Second, early mitigation might be deterred by concerns for instigating more free riding from others. We argue that for many countries such concerns will be minor if present at all, the argument being that the remaining burden of mitigation is large – not markedly smaller – even if countries quite large in emissions were to be quite successful and quite aggressive quantitatively for a few decades. Countries or regions that represent a fifth or a twentieth of global income, of global population, of global emissions have in common that they hardly can – by early mitigation in a few decades of this century – give other countries reason to be less willing to engage in mitigation.

An example calculation can be that even if early movers represented half of world's emissions and eliminated them completely, climate scientists recommend global reductions of 80% by 2050, so a tremendous and growing job would still be left for 'the others'.

We show that concerns for increased freeriding would be greatest for someone who is great in terms of emissions and face others who are both great in terms of emissions *and* vulnerable to climate change. Even in this case, this concern would shape the actions of a country or a coalition wanting to act early: spending more on demonstration, R&D and on reducing costs, showing adoptability, less perhaps on reducing emissions. A country showing leadership would be thinking ahead, on the goals and behavior of others, and on how to support future reductions and the formation of a treaty.

Third: The fact that leadership perspective can shape early action should not exclude but rather be informed about the urgency of pressing for stronger global participation.

How does the leadership perspective inform early action?

We might invoke the ideas by hypothesizing:

Say that in a situation without a treaty (or a pretty weak one), a small country does something to reduce emissions in one to three decades early in this century. Then, a big country (or several):

- observes this,
- asks some questions, say in 2020 or 2040 and
- concludes: We shall do something like that, too.

What could it be that the small country had done, and what would the big country's questions be?

Our suggestion is that the 'following' country would ask:

i) have you reduced your own emissions?
ii) did it cost you a lot?
iii) can I do it?
iv) why should I?

These suggestions are based on ideas from literature on leadership and cooperation. We may notice that climate negotiations and policy till now largely has focused on question iv), how to provide incentives for emission reductions. One may conclude from our analysis that i) through iii) has received too little attention.

Somewhat speculatively, building on the above discussion, tentative answers to these four questions, would be

i) Yes, I have. Come over and take a look. I have emission free cities, buildings, cars, fertilizer. I am even working on an emission free coal fired power plant, and aviation fuels.
ii) Well it has cost me. I would not advise you to go to this task empty handed. But as you can see, none of these measures threatens my lifestyle or my competitiveness. In some areas, you can see the solutions are getting cheaper, and this will accelerate as others join. In other areas, we have invested in technological change. We chose

fertilizer, cement and carbon capture and storage – because we felt some obstacles to global cooperation had to be tested. And in some areas you'll see we earned some other benefits, like cleaner air, less traffic jam. In some areas, we have made little progress.

iii) Much of it, surely, applies directly to you. And we also have done some nice projects together, in areas where adoptability in your setting was not so obvious. Like in tropical forests, agriculture and solar thermal.

iv) Well, look around. We are in this together. It has to be done everywhere. Others are joining. It makes sense. Nobody wants to be outside, and we'll find ways to share the burden fairly. And we'll assist each other, trade, create new technologies, make them accessible. Surely, we'll look into incentives and ways to help each other out.

We may notice that there is very little of cost effectiveness here, there is quite a lot of signaling. Costs are respected; nothing will work if prohibitively costly, but it is also the case that once broad adoption is possible, solutions may be feasible that are out of the question either at a smaller scale or for a country acting alone in a globalized world. This is also why R&D can be as important as emission reductions in an early phase.

R&D can bring future costs down, or show feasibility, as for instance with carbon capture and storage. If it can be shown not to be prohibitively costly, it will likely have an important effect. Given that windmills and solar can be in the range of twice as expensive as low-cost (carbon intensive) electric power generation, prohibitive would have to mean more in the range of three to times more expensive. Estimates for the costs of CCS (see IEA, for instance) are in the range of adding fifty to hundred percent to the cost of power, so presently the expected costs are not as deterring as is the fact that it is not demonstrated.

Signaling that a low-carbon society is feasible can be required not only in technical but also social, behavioral and political spheres. Whether a low-carbon city is with low transportation intensity or with low-carbon transport, whether it is dense, whether internet takes over, all of these questions may be raised and answered.

To represent leadership, a low-carbon city has to be surprisingly liveable, politically feasible: it has to be an eye-opening experience. Obviously, nobody can guarantee to make such a demonstration, but exploring that frontier is what is entailed in leadership.

In the course of this century, mankind could get three times richer, or fifteen (try 3 percent income growth, and UN population projections, indicating population culminates in this century at plus 50%), so it is possible to afford mitigation costs even if they were feared to be five percent of world income. Choosing costly power could cost us one percent. If it can be done, it has to be shown that it can be one. That taking the lead is self-defeating by inducing free riding, is hard to imagine.

An element of leadership not discussed here that could be important is rewards in terms of winning technologies and competitiveness. We shall not claim that this 'Porter hypothesis' is not important - it may be.

Our emphasis has been to point out that the present literature have failed to recognize that early action may be motivated by the leadership perspective: certain early acts are meaningful – productive – because others may follow, meaningful not because you'll sell winning technologies to them, but because you instigate large reductions (even if uncertain, of course).

In a time of weak or no global agreement, it may not be much deterrence to early action in the threat that others will respond with increased free riding. Importantly, early action should certainly not detract from efforts to form treaties. Finally; early, unilateral, unconditional action will be more than just cutting emissions at least cost. An appropriate question to a leader is: exactly what are you trying to prove?

References

1. Aaheim, H. Asbjørn (2010). The determination of optimal climate policy. *Ecological Economics*, 69 (3): pp. 562-568.
2. Acemoglu D. and Jackson M.O (2011): History, Expectations, and Leadership in the Evolution of Social Norms, *NBER Working Paper Series* No. 17066.
3. Aghion and Tirole (1997) "Formal and Real Authority in Organizations". *Journal of Political Economy* , Vol. 105, No. 1 (Feb., 1997), pp. 1-29.
4. Anderson, K., and A. Bows, 2011. Beyond 'dangerous' climate change: emission scenarios for a new world. Philosophical Transactions of the Royal Society A, 369, 20-44.
5. Axelrod, R. The emergence of cooperation among egoists. American Political Science Review, 75, pp. 306-318.
6. Barrett S.(2013): Climate treaties and approaching catastrophes, *Journal of Environmental Economics and Management*
7. Barrett S. and Dannenberg A. (2012): Climate negotiations under scientific uncertainty, *Proceedings of the National Academy of Science of the United States of America* 109 (43) 17372-17376.
8. Barrett, S. (2003) Environment and Statecraft: The Strategy of Environmental Treaty Making. Oxford U. Press.
9. Barrett, S. (1994). Self-enforcing international environmental agreements. Oxford Economic Papers, New Series, V 46, pp 878-894.
10. Bernard, A. and M. Vielle (2009). Assessment of European Union transition scenarios with a special focus on the issue of carbon leakage. *Energy economics*, 31, s274-s284.
11. Bolton P., Brunnermeier M.K. and Veldkamp L.(2013): Leadership, Coordination, and Corporate Culture. *Review of Economic Studies* 80, 512-537.
12. Böhringer, C. et al. (2009). EU climate policy up to 2020: an economic impact assessment. *Energy economics* 31, s295-s305.
13. Brandt, U.S. (2004). Unilateral actions, the case of international environmental problems. *Resource and Energy Economics*, 26, 373-391.
14. Cappelen A.W., Sørensen E.Ø. and Tungodden B. (2010): Responsibility for What? Fairness and Individual Responsibility, *European Economic Review* 54, 429-441.
15. Cappelen, A. et al (2013a). Face-saving or fair-minded: What motivates moral behavior? Discussion paper, Dept o Eccs, 5 2013, NHH.

16. Cappelen, A. et al. (2013b). Leadership and incentives. Discussion paper Dept o Eccs 10 2013. NHH. Also presented at Norwegian Conference of Economics Researchers, Stavanger, January 2013.
17. Chander, P. and H. Tulkens (1995). A Core-theoretic solution for the design of cooperative agreements on transfrontier pollution. International tax and public finance, 2, 279-293.
18. Denant-Beomont, L. et al. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. Economic Theory, 33, 145-167.
19. Dewan T. and Squintani F. (2013): On Good Leaders and Their Associates.
20. Edenhofer, O. and M. Kalkuhl (2011). When do increasing carbon taxes accelerate global warming? A note on the green paradox. *Energy Policy*, 39, 2208-2212.
21. Eskeland, G. S. and A. Harrisson (2003). Moving to greener pastures? FDI and the pollution haven effect. *Journal of Development Economics.*
22. Eskeland, G. S., T. Mideksa and N. Rive (2012). European climate goals and the electricity sector. Energy Policy, 41.
23. Fischer, C. and A. K. Fox (2009). Comparing policies to combat carbon leakage. RFF discussion paper # 09-02-REV.
24. Frankel, J. A. (2008/2009). Addressing the leakage/competitiveness issue in climate change policy proposals. Brookings trade forum, pp 69-91.
25. Gächter S. and Riedl A. (2005): Moral Property Rights in Bargaining with Infeasible Claims. *Management Science* 51(2), 249-263.
26. Hermalin B.E. (1998): Toward and Economic Theory of Leadership: Leading by Example, *American Economic Review* 88(5) 1188-1206.
27. Harstad, B. and G. S. Eskeland. Trading for the future: Signaling in permit markets. Journal of public economics.
28. Harstad B. (2012): Buy Coal! A Case for Supply-Side Environment Policy, *Journal of Political Economy* 120(1) 77-115.
29. Heal G. (2009): Climate Economics: A Metal-Review and Some Suggestions for Future Research. *Review of Environmental Economics and Policy* 3(1), 4-21.
30. Hertwich E.G. and Peters G.P. (2009): Carbon Footprint of nations: A Global, Trade-Link Analysis. *Environmental Science & Technology* 43(16), 6414-6420.
31. Hoel, M. (1991). Global environmental problems: The effects of unilateral action taken by one country. *Journal of environmental economics and management*, 20, 55-70.
32. Hoel, M. (1994). Efficient climate policy in the presence of free riders. *Journal of environmental economics and management*, 27, 259-74.
33. Hoel, M. and R. Golombek (2008). Endogenous technology and tradable emission quotas (with Rolf Golombek), Resource and Energy Economics 30, 197-208.
34. Hoel, M. (1996). Should a carbon tax be differentiated across sectors? *Journal of environmental economics and management*, 59, 17-32.
35. Hopfensitz A. and Reuben E. (2009): The importance of Emotions for the Effectiveness of Social Punishment. *The Economic Journal* 119, 1534-1559.
36. Hume, D. (1793). A Treatise on Human Nature. Electronically at ebooks.adelaide.edu.au

37. Konow J. (2000): Accountability and Cognitive Dissonance in Allocation Decisions. *The American Economic Review* 90(4), 1072-1091.
38. Kosfeld M., Okada A. and Riedl A. (2009): Institution Formation in Public Goods Games. *The American Economic Review* 99(4) 1335-1355.
39. Levati M.V., Sutter M. and Van der Heijden E. (2007): Leading by Example in a Public Goods Experiment with Heterogeneity and Incomplete Information, *Journal of Conflict Resolution*, 51(5) 793-818.
40. Lomborg, Bjorn *The skeptical environmentalist.* Cambr U Press, 1998.
41. Meinshausen, M. et al. (2009). Greenhouse-gas emission targets for limiting global warming to 2^C. Nature, v. 485, april 30, pp: 1158-1163.
42. Moxnes E. and Van der Heijden E. (2003): The Effect of Leadership in a Public Bad Experiment, *Journal of Conflict Resolution*, 47(6) 773-795.
43. Moxnes E. and Van der Heijden E. (2013): Leading by Example to Protect the Environment; do the Cost of Leading Matter? *Journal of Conflict Resolution*, Forthcoming.
44. Nordhaus, W. (2008). A Question of Balance: Weighing the Options on Global Warming Policies. Yale University Press.
45. Prendergast, C. (1993). 'A Theory of "Yes Men"', *The American Economic Review*, Vol. 83, No. 4, pp. 757-770.
46. Reuben E. and Riedl A. (2009): Enforcement of Contribution Norms in Public Goods Games with Heterogeneous Populations. *IZA Discussion Paper Series* No. 4303.
47. Rivas M.F and Sutter M. (2011): The Benefits of Voluntary Leadership in Experimental Public Goods Games. *Economics Letters* 112(3) 176-178.
48. Rotemberg J.J and Saloner G. (1993): Leadership Style and incentives. *Management Science* 39(11), 1299-1318.
49. Samuelson, Paul A. (1954). The pure theory of public expenditure. Review of Economics and Statistics, 36, 387-89.
50. Sandmo, Agnar (1972). Optimality rules for the provision of collective factors of production. Journal of public economics, 1, 149-57.
51. Sinn H-W. (2008): Public Policies Against Global Warming: a Supply Side Approach, *International Tax Public Finance* 15, 360-394.
52. Stern N.P. (et. al.) (2007): *Stern review: The Economics of Climate Change*, London: HM Treasury. Cambr U Press, jan 2007.
53. Stern N.P (2011): The Economics of Climate Change. *American Economic Review*, 98(2), 1-37.
54. Sterner, T. and U. M. Persson (2008). An Even Sterner Review: Introducing relative prices into the discounting debate. Review of Environmental Economics and Policy, v 2, pp. 61-76.
55. Sugden, R. (1986). The economics of right, cooperation and welfare. Basil Blackwell Publishers, Oxford, UK.
56. Tol R.S.J. and Yohe G.W (2006): A review of the Stern Review, *World Economics* 7, 233-251.
57. Van der Ploeg, F. and C. Withagen (2009). Is there really a green paradox? CESifo Working paper 2963.

58. Van der Heijden E. and Moxnes E. (2012): Leading by Example to Protect the Environment: Do the Cost of Leading Matter? *Journal of Conflict Resolution* 00(0), 1-20.
59. Weitzman, M. 1999. Pricing the Limits to Growth from Minerals Depletion. *The Quarterly Journal of Economic,* Vol. 114, No. 2 (May, 1999), pp. 691-706
60. Weitzman, M. (2007). A Review of the Stern Review on the economics of climate change. Journal of Economic Literature, Vol XLV, pp. 703-724.
61. Wood P.J. (2010): Climate Change and game Theory: a Mathematical Survey. *Environmental Economics Research Hub*, Research Report No. 62.