

Climate Coalition: the Themis Mechanism

Carl Edward Rasmussen

Cambridge University

Adenauer School of Government, University of Cologne, September 18 2026

Outline

The goals of the talk are to:

- investigate how far principled approaches to climate coalitions can be pushed, in ways necessary for actual implementation.
- separate what we can establish rigorously, and where genuine difficulty remains.

Rough outline:

- Climate coalition principles
- Cap or price?
- Domestic and international pricing
- Governance and Elicitation
- Legitimacy
- Coalition dynamics
- Conclusions

The Carbon Dine and Dash Deficiency

The **carbon dine and dash deficiency**: Climate change is driven by a single structural flaw: the user of fossil fuels enjoys the immediate, local benefit, while leaving the cost for the global community and future generations.

Effectively addressing climate change is a collective action problem. It requires that we

- cooperate
- to remove the separation between climate cost and benefits.

This talk is about

- the principles
- a concrete proposal: the Themis Mechanism

which could govern a climate coalition (or club).

Climate Coalitions

What do we know *must be true* for effective climate coalitions?

Qualification: I am considering here a climate coalition in the narrowest possible sense. Objective to reduce greenhouse gas concentrations.

- The atmosphere is a common pool resource. Everyone benefits from reduced emissions, this creates a free-rider problem; there is pressure on nations individually to do as little as possible. Therefore, membership must be in the self-interest of nations.
- because of the common pool resource property, for climate coalitions big is beautiful.
- unanimous membership is not an appropriate organising principle, because it is a structural guarantee that the least ambitious nation sets the agenda. We must accept non-members in principle.
- coalition commitments must be common; bespoke agreements would be intractable.
- coalition commitments must be conditional (or reciprocal) "I will if you will". This implies simultaneous actions: no-one goes first.

Commitment type: cap or price?

- a cap *seems like* the most direct approach
- because energy requirements are *inelastic*, in practice a cap implies a price (aka cap and trade); the real distinction is which is **explicitly set**.
- caps preclude cooperative behaviour; caps are fixed total budget, they imply zero sum behaviour
- caps can't co-exist with other approaches: caps dominate environmental outcomes
- even in a capped system, the price is the real objective
- caps allow handing out allowances (political convenience)
- in national systems, specifying a price carries political liability

What common commitment?

Common commitments (or rights and responsibilities) must be attractive to nations in widely different circumstances.

Two types of emission pricing: international transfers and domestically retained pricing.

International transfers: pay money into global pot, and distribute according to population. Easily justifiable from equitability point of view.

Either type won't work on their own, for opposite reasons:

International transfers: to be effective at changing behaviour, prices would need to be significant. But this implies very large cash transfers, which are politically impossible.

Domestic pricing: prices would have to be identical across the world to avoid distorting competition. But wealth per person is very different across nations: fees would be easy to meet for rich nations, but crippling for poor nations.

The solution is to combine both.

How would pricing work?

Domestically retained pricing would simply be a (minimum) price p for emitting a ton of CO₂e. This would rule out subsidies.

International transfers: collecting a fixed price t per ton of CO₂e and distributing equally to (member) population is *equivalent* to charging (and paying) only for excess per capita emissions.

In other words: above average (per capita) emitters pay, and below average emitters get paid.

The international transfers balance: no separate finance is needed.

We have created a governance issue: how to agree on the two prices p and t .

Balancing the international transfers in the presence of free-riders

Above average (per capita) emitters pay and below average emitters are paid. But the world average $e^* = 6.4$ tCO₂e per year is not the same as the coalition average $\bar{e}$. We must use the coalition average, otherwise the books won't balance.

Issue: this means that a few large emitters defecting, may cause a nation to switch from being a beneficiary to a contributor, through no fault of their own.

Fix: we use the world average e^* as reference point, but beneficiaries and contributors use different prices t^- and t^+ , with the extra constrain that the books must balance. (This doesn't increase the number of parameters to be determined, because of the constraint.)

This idea was co-developed with Steve Stoft.

Agreeing on prices and membership

This is complex. Prices and membership are interdependent. What price a nation would accept depends on membership, because climate pay-off depends on coalition size. Both need to be determined simultaneously.

Democratic voting is not a good option, because its basis is that voters accept the outcome, which won't work well for sovereign nations.

Instead, an *elicitation process* is required, which extracts information from nations about their views on the relationship between price and membership, and a principle to fuse these views into specific choices.

An important issue for elicitation processes is to what extent the process can be manipulated: can nations achieve better outcomes by responding strategically rather than honestly, and how big an effect does this have on fairness?

Unfortunately, game theory offers mostly negative results; and no applicable tight guarantees.

One positive result is: repeated, transparent games have much superior properties to one-off or secret processes. Because it lets partners build trust.

What should nations be asked about?

Free form questions would be impractical to fuse to decisions in a principled way.

It seems natural to ask for information which is most important to national decisions.

This includes obviously the values of p and t , but also who else will be members. Actual membership is complex (combinatorial), this could be simplified to a single scalar, the *coverage* c , the fraction of emissions covered by members.

Because the pay-off from the coalition depends on its size, coverage c could be a minimal representation.

This would mean that nations could express their views about the relationships between the domestic price p , international transfer t and coverage c . But exactly how? We examine two possible choices.

Elicitation via preferences

The simplest elicitation could be to state a **preferred value** for p and t . We could further accommodate coverage c , by asking for p and t conditioned on various choices of c . This would allow nations to allow higher payments the larger, and therefore more impactful, the coalition.

How would the Coalition Authority fuse these preferences? For every possible value of the coverage, compute the highest p or t supported by a coalition of that size. Then select the final configuration as the one which maximises the impact of the coalition; a good proxy for this may be $c \times t$.

A couple of issues:

- we only asked for "preferred" values; perhaps much higher values were acceptable?
- at the end, values were chosen which did not coincide with most nations preferred values. How do we know which nations will join at that value?
- the procedure is not strategy proof.

Perhaps reasoning about asymmetries may help (if a nation prefers p , does that imply that it accepts anything lower?), but that would mean that it's not the preferred price but the maximum (preferred/acceptable) price we've effectively asked about.

Elicitation via acceptable regions

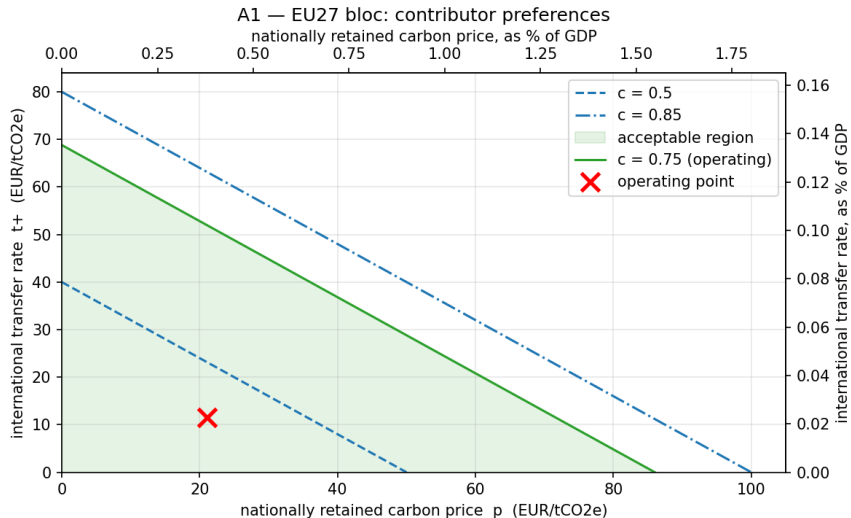
Instead of a point-wise preference, **acceptable regions** could be elicited. In the scalar case, most simply an interval. And working in the coverage c , an interval depending on c (interpolate between intervals at specified value of c).

The Coalition Authority could, much like before, find for each c the largest price acceptable to a coalition of that coverage. But crucially, there would be no need for another confirmation round, since each nation has already specified that exact combination.

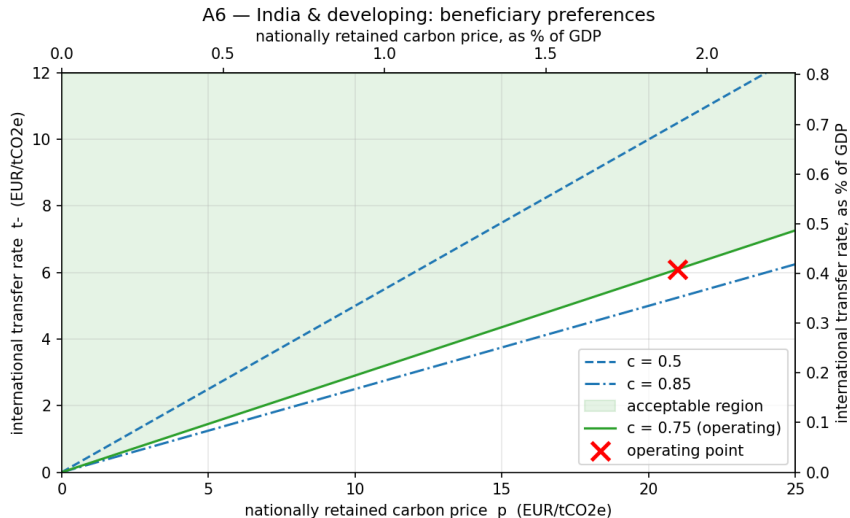
This both simplifies the elicitation procedure and probably makes strategic voting less attractive, since the first vote binds.

In practice, we need to generalise from scalar price, to prices with two components, p and t .

Elicitation via acceptable regions: contributor example



Elicitation via acceptable regions: beneficiary example



Elicitation via acceptable regions

Beneficiaries specify regions in p and t^- space, and contributors in p and t^+ space. These are combined seamlessly via the balanced book constraint. The computational work is trivial (non combinatorial), even for 200 nations.

The optimal point, will have one beneficiary and one contributing nation on the "pivot", ie whose region boundary defines the operating point.

One may fear that this means that these nations hold power to manipulate the operating point in their favour. Perhaps so. But they didn't choose to be at the pivot, and with many nations, the size of the region where they stay on the pivot is small.

But attempted manipulation would not be risk free: if too small regions are specified, then you risk being left out of something that was actually attractive. If too large regions were specified, the risk is getting something you don't want, or having to renege, and lose reputation.

Seven archetype mock outcome: 90% world population membership, 75% coverage, $p \simeq 21\text{€}$, $t^- \simeq 6\text{€}$, $t^+ \simeq 12\text{€}$, and total international transfer 98B€.

Legitimacy

How can this approach find legitimacy?

It rests on many sound principles. But it also makes some choices, which could perhaps have been made differently. For example, to carve out and separate ongoing from historical emissions and details about the elicitation process.

How can such choices, even in principle, be broadly legitimised?

The answer, it seems to me, has to be, that the effects of following the proposals are much better to a large majority, than not following them.

It's not that Themis is absolutely fair; it's just much better than what we have at the moment. Since carbon dine and dash is so poor, there is much scope for improvement.

As a proposal gains acceptance, that intrinsically makes it more attractive; exploit this virtuous dynamic.

Annual cycle guarantees reinforcement of trust, flexibility, no open-ended commitments.

For any novel approach, it is wise to start small.

But for a climate coalition, the pay-off depends directly on size.

How are these opposite requirements reconciled?

Luckily, there are two ways to start small: small membership or small prices. Initial small prices may be a more convincing way to get off the ground, than small initial membership.

Starting large inherently captures the simultaneousness requirement implied by the central "I will if you will" insight, that no-one goes first.

Conclusions

I've attempted to add more specificity to the ground work from "Price Carbon – I will if you will".

These principles include conditionality or reciprocity, and common commitments, and have detailed preferences for prices over caps.

We've been more specific about having to accept non-members, and more about domestic and international payments. And advocated for the general superiority of repeated elicitation compared to one-off processes.

To come even closer to a process which could actually be implemented, we've had to become a little bit more speculative. Elicitation processes are exceptionally complex, and full confidence is difficult to achieve.

None the less, only a narrow range of option seem viable. The bar for process must not be set too high to dissuade ideas which have definite merit over the status quo.