Climate Cooperation: the Themis Mechanism

Carl Edward Rasmussen

Cambridge University

Dyson School of Design Engineering, Imperial College London, Nov 12 2026

Outline

- motivation
- our climate crisis
- about cooperation
- what must coalitions be like?
- conclusions and what next?

Motivation

Key idea: to derive a response to climate change from the constraints of the problem.

Synthesise established insights from climate science, economics, game theory, behavioural economics, psychology, philosophy and institutional design.

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).

Incentives to climate action

Governance based on altruism is fragile and impractical. Therefore the real challenge is to promote **governance which aligns self-interest with the common good**.

When a nation, comprising 1% of the world, like the UK, makes an investment to limit greenhouse gas emissions:

- 1% of the climate benefits manifest in the UK.
- 99% of the climate benefit manifest elsewhere.

It's not that investment towards emission reduction doesn't pay off, it just mostly pays off elsewhere.

This poor national return on investment makes nations stall climate action.

Three illustrative games

The **independent commitment game:**

1. players can invest an amount of their choice in a communal pot
2. the pot doubles in size
3. the pot is shared equally between all players

Let's say there are 100 players, each with €1.00. How would you play?

Investment is good for everyone: if all 100 players invested €1.00, everyone would end up with €2.00.

But what would be your return on investment? Compare not investing with investing €1.00. The difference (irrespective of what others do) in payout is €0.02: that's -98% return, terrible!

It leads to defection or free-riding.

Three illustrative games

The **fixed budget game**:

1. players can invest an amount of their choice in a communal pot, subject to the condition that the total budget must be met
2. the pot doubles in size
3. the pot is shared equally between all players

There are 100 players, each with €1.00. The budget is €50.00. How would you play?

Because the budget is fixed, the payout is fixed at €1.00. Players become adversarial. It's a zero sum game. No cooperation can take place.

Three illustrative games

The **common commitment game:**

1. players can pledge an amount of their choice towards a communal pot; only the part of the pledge which is *common to all* is invested (the rest is returned)
2. the pot doubles in size
3. the pot is shared equally between all players

There are 100 players, each with €1.00. How would you play?

There is no reason not to pledge the full amount. You cannot be exploited by free-riders.

The common commitment is key.

Three game conclusions

The games are instructive ...

- the ”pot is shared equally” is the crucial aspect of ”common-pool resource” (CPR) problems
- ”independent commitment” games get sunk by free-riders
- ”fixed budget games” are zero-sum. They turn players against each other (adversarial)
- ”common commitment” games change everything: encourages cooperation, or alignment of incentives

... but simplistic ...

- there is no notion of time
- it assumes that agents are equivalent. But real nations differ. So how should we interpret ”common commitment”?
- it is not practical that a single nation could prevent a successful coalition

Climate Coalitions (or Clubs)

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

- The atmosphere is a common-pool resource (CPR). 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, big is beautiful for climate coalitions.
- 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. Non-members must be sanctioned.
- 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 best approach, but pricing is better.

Because energy requirements are *inelastic*, in practice a cap implies a price (aka cap and trade); the real distinction is which is **explicitly set**.

Pro price:

- caps preclude cooperative behaviour; caps are fixed total budgets, 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
- carbon pricing is often said to be socially regressive: depends entirely on how proceeds are used!

and cons:

- 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: domestically retained pricing and global fee and dividend.

Global fee and dividend: 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 composite pricing, comprising a bit of both.

How would pricing work?

Domestically retained pricing would simply be a (minimum) price p for emitting a ton of CO_2e. This would rule out subsidies.

International transfers: collecting a fixed price t per ton of CO_2e 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.

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.

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 fairer than what we have at the moment. Since carbon dine and dash is so po eor, 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.

The Themis Mechanism

The *Coalition Authority* (CA) administers the Themis Mechanism in an annual cycle:

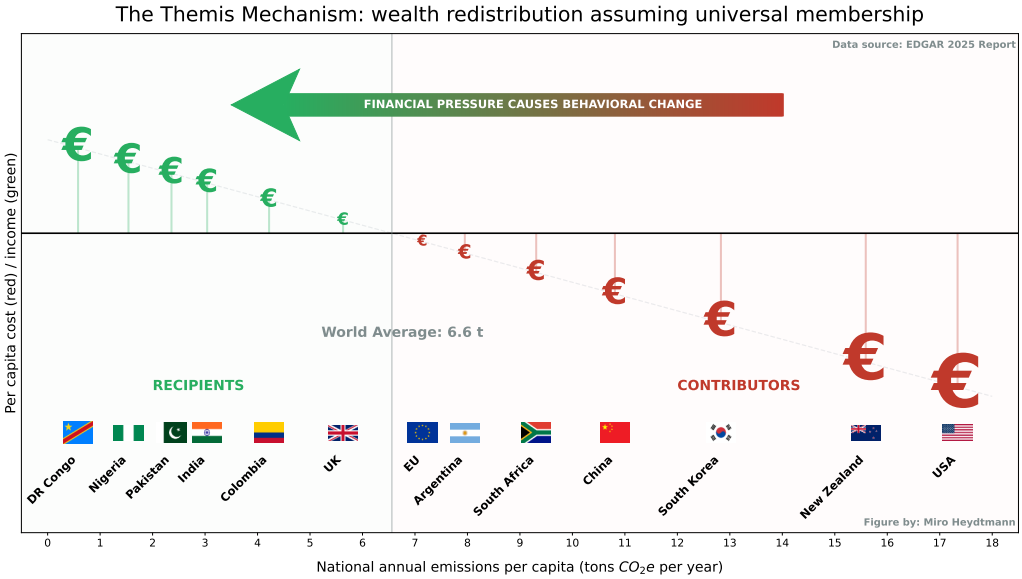
1. all nations are invited to join the coalition at predefined price p and international transfer t
2. members enforce the domestic price p, report their emissions (using the established UNFCCC reporting rules), and pay (or get paid) t times their excess emissions to the coalition
3. next year's p and t are determined by a member elicitation process

In the inaugural year, the prices are set to zero.

The only decision nations make is: to join or not? There is nothing to negotiate.

Themis immediately creates economic pressure on all nations to emit less. The annual cycle creates trust.

The Themis Mechanism



Conclusions and Next Steps

Climate change is a tractable problem – one we can choose to solve.

The solution

- must price carbon to eliminate the carbon dine and dash deficiency,
- does not require altruism, but agreement between self-interested parties,
- is a voluntary coalition based on conditional common commitments which sanctions non-members.

Next steps

- better understanding of cooperation,
- new ideas take time,
- we need nations to commit reciprocally: ”I will if you will”, actually a small ask.