BSG Working Paper Series

Providing access to the latest policy-relevant research

Catalytic cooperation

BSG-WP-2018/026

September 2018

Thomas Hale, Blavatnik School of Government, University of Oxford

CATALYTIC COOPERATION

Thomas Hale*
September 2018

Abstract

Scholars typically model the politics of global public goods or common pool resources as a difficult class of collective action problems—the tragedy of the commons. In such cases, theories of international organization aim to explain how institutions can promote cooperation by solving the free-rider problem. Based on an analysis of a quintessential global collective action problem—international climate mitigation—this article challenges both this diagnosis of the problem and the concomitant institutional remedies. The mitigation collective action problem exhibits three key features that depart from the canonical tragedy of the commons model: joint goods, preference heterogeneity, and increasing returns. The presence of these features in climate mitigation creates the possibility for "catalytic cooperation." Under such conditions, the chief barrier to cooperation is not the threat of free-riding, but the lack of incentive to act in the first place. States and other actors seek to solve this problem by creating "catalytic institutions" that work to shift actors' preferences and strategies toward cooperative outcomes over time. While catalytic institutions can be seen in many areas of world politics, the 2015 Paris Agreement on climate change has put this logic of cooperation at its core, raising the possibility that similar catalytic institutions may facilitate cooperation in other areas of world politics characterized by analogous problem structures.

Key words: International relations, global governance, international institutions, climate change, cooperation, Paris Agreement

* Associate Professor, Blavatnik School of Government, University of Oxford https://www.bsg.ox.ac.uk/people/thomas-hale.

For generous comments and feedback I wish to thank: Liliana Andonova, Scott Barrett, Steven Bernstein, Joshua Busby, Jarred Finnegan, Fergus Green, Jessica Green, Matthew Hoffmann, Robert O. Keohane, Mareike Kleine, Michael Oppenheimer, Lauge Poulsen, Charlies Roger, Anne-Marie Slaughter, Duncan Snidal, Harro van Asselt, Johannes Urpelainen, David Victor, and the participants of the London School of Economics International Political Economy Seminar, the Oxford International Political Economy seminar, the Nuffield College International Relations group, and the London School of Economics Grantham Research Institute. workshop.

1. Introduction

The tragedy of the commons is a powerful heuristic for understanding the challenges of providing public goods or preserving common pool resources, marking a particularly difficult class of collective action problems.[1] Cooperation theory, in turn, shows how actors can overcome such challenges through strategies that foster credible commitments and remove incentives to free-ride, such as creating international institutions to monitor compliance, sanction defection, or generate excludable benefits.[2]

While the literature applies this framework broadly to issues of global public goods and common pool resources, scholars have only rarely examined its core assumptions.[3] To what extent do global commons problems actually resemble the tragedy of the commons? And if they depart from it, are different strategies needed to promote cooperation?

This article identifies general features of global commons problems that, to the extent they apply, shift the logic of cooperation from a tragedy of the commons to what I term a "catalytic cooperation" model. [4] The article also shows how international institutions can help create and sustain cooperation in such circumstances not by solving the problems of free-riding and credible commitments, but by mobilizing incremental action that can shift the strategies and preferences of states and other actors over time.

The article grounds its theoretical arguments in an analysis of a quintessential global commons problem, international efforts to mitigate climate change. In 1992, countries agreed the United Nations Framework Convention on Climate Change (UNFCCC), pledging to prevent "dangerous" changes in the Earth's climate. Despite significant diplomatic effort and institution-building, states agreed only one treaty limiting emissions, the 1997 Kyoto Protocol, which required modest cuts for wealthy countries but was rejected by the United States and soon outpaced by the rapid industrialization of emerging economies. Though the Kyoto Protocol was always intended as a first step, states failed to take a second step at the 2009 Copenhagen summit, leading some to question whether the international climate regime could ever succeed, or whether a different approach was needed.

Remarkably, the UNFCCC did not simply sink into stasis. Instead, over the following years, policy entrepreneurs radically changed the logic of the regime.[5] Kyoto, following the model of most other multilateral environmental agreements, took a "regulatory" approach, in which states negotiated a set of shared reductions targets. States are bound to report on their emissions and, should they miss the agreed target, face sanctions (minimal, in this case). The 2015 Paris Agreement, instead, requires each country to put forward its own pledge, or "nationally determined contribution" (NDC). These pledges are then reviewed internationally and ratcheted up every five years. Paris also gives a central role to cities, businesses, provinces and regions, and other sub- and non-state actors, creating institutions to recognize and orchestrate their climate action alongside the national pledges. Both these processes aim toward the ambitious long-term goal of ensuring no more greenhouse gases (GHGs) are going into the atmosphere than are being absorbed—effectively decarbonizing the world economy—by the second half of the 21st century.

[1] Hardin 1968; Olson 1971
[2] Keohane 1984; Axelrod 1985; Ostrom 1990
[3] See Sandler 2004; Ostrom 2009; Urpelainen 2013; Rayner and Caine 2015; Bernstein and Hoffmann 2018
[4] I am grateful to David Victor for suggesting this phrase.
[5] Hale 2017

Can this "pledge and review and ratchet" system work? Existing theories of international organization see the Paris Agreement as a positive but modest achievement for cooperation because it does little to solve free-rider problems and lacks robust enforcement provisions. This article makes two theoretical arguments that invite a different interpretation.

First, it aims to correct a persistent misdiagnosis in international relations (IR) theory, which overemphasizes free-riding as the chief barrier to international cooperation around the global commons. Though the tragedy of the commons represents *prima facie* a compelling interpretation of global climate mitigation, much of the observed political behavior and outcomes over the past two decades departs from its theoretical expectations. Three generalizable patterns emerge: climate policies provide private benefits as well as collective ones ("joint products"); the value actors place on mitigation, as well as the costs and benefits they face, vary enormously ("preference heterogeneity"); and action in the past lowers the costs and raises the benefits of action in the future ("increasing returns").

To the extent these conditions apply, the catalytic cooperation model implies a very different set of barriers to, and strategies for, cooperation than the tragedy of the commons model. In such circumstances, incremental action can lead to cooperation over time. Action by first-movers alters the strategic environment around subsequent action. If early movers reach a critical mass, and if past action has a large impact on subsequent action, cooperative action can become progressively self-reinforcing over time.

Second, the article introduces the concept of catalytic institutions, demonstrating how the Paris Agreement, but also international institutions in areas such as trade, human rights, and global health, attempt to support cooperation not by providing credibility, but through catalytic mechanisms that help build action over time by progressively shifting preferences.

The next section demonstrates how climate mitigation departs from the assumptions of the tragedy of the commons, identifying three conditions that define a catalytic model of cooperation. Section three examines the implications of this model for cooperation. The fourth section then elaborates the logic of catalytic institutions and shows how the Paris Agreement and other international institutions work to shift state preferences and strategies over time. In conclusion, the article considers which other issues in world politics might be productively interpreted through a catalytic cooperation lens, and therefore where else catalytic institutions may support cooperation. It also considers how the problem structure of climate change may shift over time.

2. From the tragedy of the commons to catalytic cooperation

Theories of collective action in political science, sociology, economics, and other fields typically consider the incentive of an actor to contribute to collective action given the actions of others. In the classic model the level of collective action provided is a production function *P* of the sum of all actors' contributions, *A*.[6] For a given actor *i*, the benefit of acting $g_i(A)$ is the value v_i it derives from the collective action *P(A)*, minus the cost c_i of its individual contribution *a*.

The classic model of collective action:

$$g_i(A) = v_i[P(A)] - c_i(a)$$

[6] See, for example Marwell and Oliver 1993

The canonical tragedy of the commons is a special case of the classic model that occurs when no actor has an individual incentive to act unless a sufficient number of others do as well; that is, when on average $c_i(a) > v_i[P(A)]$. This situation creates a familiar and intuitive problem, a large-scale prisoners' dilemma.[7] Bigger actors may have more to gain by acting, because they can do more to affect the problem, but if no actor's contribution is by itself sufficient to gain an overall benefit that outweighs its cost of acting, collective action will not be forthcoming. Furthermore, even if some action should occur, an actor has little incentive to join in because it can more easily free-ride on others' efforts. Absent a credible commitment for all to act, no action is taken, and everyone ends up worse off.

Global climate mitigation is typically seen as a case par excellence of this model. This conventional wisdom seems incontrovertible because the atmosphere is inherently rival (one actor's emissions reduce the absorptive capacity available for others) and non-excludable (emissions anywhere affect the climate everywhere), making the climate a textbook common pool resource. Similar dynamics apply to other global environmental challenges, humanitarian relief efforts, development assistance, refugee policies, and a host of other issues.

But how accurately does the classic model of collective action, and specifically the special case of the tragedy of the commons, capture the politics of climate mitigation or of other global collective action problems? At the extreme, the classical model would expect no one to act in the absence of a credible agreement for collective action. But this has not been the case in global climate mitigation. Over the past two decades, a number of countries, cities, businesses, and others have taken aggressive and costly action on climate change even without any guarantee that others will follow suit.[8]

Indeed, there has been such a proliferation and pluralisation of climate governance, that observers increasingly speak of a "polycentric" climate regime, or a "regime complex" for climate change, in which governance and action occur in myriad ways through many different actors and institutions that range in scale from global to local.[9] Though this decentralized approach to the climate challenge was suggested even before the formation of the UNFCCC, it has now become a reality.[10]

For Ostrom and others, this trend throws into question not only the tragedy of the commons, but the very utility of analyzing the climate challenge as a global collective action problem. Polycentrism argues that climate encompasses many interlocking sub-issues that manifest differently across contexts. It may still be a tragedy of the commons at the global scale, but many actors will engage with specific aspects of the problem at different scales, such as a city focused on transport, a farming community focused on preserving their crops from drought, or a consortium of corporations seeking to make their supply chains sustainable. In many of these settings collective action is either not required or easier to create because problem structures are more benign, or because existing social and political structures facilitate cooperation via social trust and other mechanisms.

This article builds on the polycentric logic's critique of the tragedy of the commons, but seeks to answer two questions polycentrism leaves underspecified. First, though polycentrism describes the cooperation we observe at multiple scales for those actors who wish to address a problem, the framework offers few ideas on how to compel recalcitrant actors outside these communities. While much of the world may be willing to act on climate change, or at least sub-aspects of it, in their own

[7] Hardin 1968; Barrett 2003; Sandler 2004
[8] Roger, Hale and Andonova 2017; Iacobuta, Dubash, Upadhyaya et al. 2018; UNFCCC 2018
[9] Ostrom 2009; Keohane and Victor 2011; Cole 2015; Dorsch and Flachsland 2017; Jordan, Huitema, van Asselt et al. 2018
[10] Gerlach and Rayner 1988

realm, there is no guarantee that enough actors will follow this approach to resolve the overarching problem. Second, the paper specifies how and through what conditions action by some parts of a polycentric system can affect others.[11] In this way the paper shows the continuing value of seeing mitigation, and similar commons issues, through a global collective action lens, albeit a modified one.

A review of mitigation politics over the past decades provides evidence for three features of climate's problem structure that depart from the tragedy of the commons:

1. *Joint goods*: Contributions to collective action can yield private benefits for those that act in addition to public benefits for all.
2. *Preference heterogeneity*: Some actors value collective action highly, others little. Similarly, the costs and benefits of a given contribution to collective action vary across actors.
3. *Increasing returns*: Action in the past can reduce the cost and increase the benefit of action in the future, while also changing how actors value collective action.

Below I explain the logic of each revised assumption and demonstrate its empirical basis in the realm of climate mitigation. To the extent they apply, they shift the classic model of collective action described above to a 'catalytic' model of collective action, which I define below.

2.1 Joint products

The classic model assumes that contributing to a collective good is costly to the contributor. In reality, of course, policy actions typically involve a mixture of costs and benefits. This means that collective goods are often "joint goods" in that actors' contributions provide both a public good to the community and a private benefit to the contributor. Even seemingly pure public goods (e.g. funding basic research to create scientific knowledge) can come with private benefits of some kind (e.g. jobs for researchers).

If the cost of a contribution tends to exceed the private benefit it provides the contributor, which we can separate out as a separate term b_i, then the classic assumption is sensible. But the empirical record shows many cases in climate policy for which this assumption is too strong.[12] After all, a country's net contribution to mitigation is almost never a single policy like a carbon tax, but rather a wide range of regulations and actions across nearly every sector including land-use policy, local zoning laws, transportation initiatives, development assistance, etc. Indeed, many actions relevant for mitigation may not be perceived primarily as climate policy per se.

Given this diversity, it is not surprising that many mitigation actions entail so-called "co-benefits", including reducing local air pollution and improving human health, increasing energy security and reliability, developing new industrial sectors, preserving forests, reducing traffic, etc. In some cases, these co-benefits temper the cost of mitigation actions; in other cases, the other benefits are governments' primary focus, and mitigation is an ancillary result. Indeed, Lipscy suggests that mitigation is more often than not a secondary consideration for policymakers.[13] Even if it is true that, on average, most mitigation actions come with a net cost, some will not be, and these "no cost" actions can allow actors to take significant steps even if they do not hold strong pro-climate views.[14]

[11] Ostrom recognizes the importance of this problem, calling for "methods for assessing the benefits and costs of particular strategies adopted in one type of ecosystem and comparing these with results obtained in other ecosystems." However, she does not specify how this might work or suggest other mechanisms that might diffuse action across the system. Ostrom 2009, 39.
[12] Sandler 2004, 53; Lipscy Forthcoming
[13] Lipscy Forthcoming
[14] Green 2015

Expanding our understanding of actors' preferences to consider non-material costs and benefits creates further possibilities to depart from the cost assumptions of the tragedy of the commons. Even when there are net material costs to mitigation actions, it is not obvious that they always dominate other incentives, including ideational preferences for or against green policy. On average, decision makers certainly prefer to avoid expending money with little immediate return, but many decision-makers face pro-climate stakeholders, including citizens and voters, customers, and investors. For the mayor of a progressive city or the leader of a country whose voters value the environment, or where environmental interest groups are powerful, there may be strong political incentives to reduce emissions even absent global collective action.[15] Indeed, under such conditions, taking a bold leadership position may bring even more votes or customers than a more incremental action. In places where climate policy has become an issue of partisan contestation (such as the United States, Australia, or Canada), these incentives may be particularly strong for politicians from certain parties or from certain jurisdictions. Similarly, companies with a 'green' brand or investors (like public pension funds) with ethical requirements may have strong incentives to take even costly mitigation actions.[16]

Considering the full range of material and ideational private benefits that accompany contributions to collective goods creates the possibility that some actors will strictly prefer acting over not acting. To the extent private benefits exist, we can expect at least some actors to take at least some actions irrespective of others' behavior.

2.2 Preference heterogeneity

If the existence of private benefits creates the possibility for some action to be taken, a key question becomes how the ratio of costs to benefits (both private and public) varies across actors. The tragedy of the commons model assumes relatively symmetric preferences across actors and actions; all derive some benefit from collective action and all must pay some cost to contribute to it, with the latter generally exceeding the benefit from the collective good any one actor can produce by itself. While this simplifying assumption may be accurate on average, again, the empirical literature on climate change provides evidence of significant exceptions. Consider each element of the classical model in turn.

First, the previous section already noted that a wide range of private benefits b_i can be generated by certain mitigation actions, and that these will vary across actors. Similarly, the costs c_i actors face for taking certain actions will vary significantly as well. For example, countries with water resources like Sweden or Costa Rica have already nearly decarbonized electricity production through hydro-power.

In addition, the 'natural' variation in costs actors face is exacerbated by the different decision-making institutions through which they form policy preferences. Comparative studies of national climate policy formation show how regime types and political-economic systems shape countries' climate policies by valuing public goods differently, giving fossil fuel interests groups more or less power, or by giving more or less value to outcomes in the future.[17] These effects can be nuanced. For example, to the extent countries' political institutions are malapportioned, rural interests (which in industrialized countries tend to have higher emissions and to be more averse to climate action) are over-represented in national politics compared to urban interests, limiting climate ambition.[18] But by the same token, malapportionment in democracies can also create incentives to distribute energy production to small scale, rural producers through mechanisms such as feed-in tariffs, which have

[15] Bulkeley, Andonova, Betsill et al. 2014; Dolšak and Prakash 2017
[16] Prakash and Potoski 2006; Hsueh 2017
[17] Dubash, Hagemann, Hohne et al. 2013; Lachapelle and Paterson 2013; Tobin 2017
[18] Broz and Maliniak 2010

been important drivers of renewable energy deployment.[19] The point is that political institutions strongly shape the formation of country preferences for climate mitigation, rendering the "one-size-fits-all" tragedy of the commons model imprecise.

The standard model also misses important differences between nation states and other kinds of actors. There are systematic reasons to expect nation states to be more susceptible to capture by fossil-fuel interests than other types of actors. While publics around the world tend to be supportive of climate action, powerful interest groups stand to lose significantly from the shift away from fossil fuels, limiting climate policy in a classic Olsonian fashion. At the subnational level, or in different economic sectors, such problems can be mitigated. While certain fossil fuel producing regions of a country may indeed be very averse to climate action, such areas are often geographically concentrated, while other regions may have no anti-climate interest groups to contend with. Large, densely populated cities are particularly unlikely to have strong interest group opposition to many climate actions.

Second, consider how actors may value the benefits of collective action v_i differently. Because the climate changes slowly, and with significant variation and uncertainty around localized impacts, the benefits of slowing global warming accrue gradually and unevenly, with the bulk going to people in the future. It is common in the literature to model this innate feature of the climate problem by assigning a discount rate to the future benefits of mitigation. Again, this simplifying assumption can obscure more than it clarifies. For small island states, arid countries near the equator, and coastal cities, the impacts are already severe and will become existential. For more temperate and inland areas, the changes will likely be slower and more moderate, potentially even bringing some benefits in the medium term to certain areas. Differential levels of economic development and state capacity will also mean that some actors will be able to adapt to climate disruptions much more effectively than others. Again, political institutions will interact with these intrinsic sources of variation, further differentiating actors from each other.[20] For these reasons, we can expect v_i to vary dramatically across actors, with some motivated to act even if their action only generates a small level of collective good provision and others unlikely to act even if they could provide significant public goods unilaterally.

2.3 Increasing returns

In the tragedy of the commons, action by others almost always dissuades an actor from acting. If others contribute, the actor can free-ride on their efforts. If others slack, the actor can scarcely hope to solve the problem without them. Only when an actor's own action triggers reciprocal action from others does it have an incentive to act. Moreover, the model assumes, implicitly, that it is useful to conceptualize actors' preferences as if they do not change over time.[21]

In reality, of course, many policy choices reinforce themselves through a variety of feedback loops that generate "increasing returns" to action over time.[22] Moreover, the increasing returns generated in one sphere can also diffuse to other actors. In this way prior action can shifts actors' preferences toward further action over time by creating and diffusing increasing returns.

---

[19] Bayer and Urpelainen 2016

[20] For example, Jacobs notes how different governments value future outcomes differently depending on how domestic institutions a) provide info about long-term consequences, b) stabilize political commitments over time, c) minimize opportunism. Jacobs 2016

[21] Historical institutionalist scholarship has noted the value of relaxing this last assumption, including the idea that changes in preferences may be endogenous to the decision to cooperate in the first place Downs, Rocke and Barsoom 1998; Caporaso 2007; Underdal, Kallbenkken and Hovi 2008

[22] Pierson 2000

The climate literature highlights many examples of increasing returns, and some scholars have recommended that policy be designed to exploit such dynamics.[23] As Bernstein and Hoffmann note, decarbonization in one area can scale and entrench in other areas through normalization, capacity building, and coalition building.[24] Material, informational, and normative mechanisms can be identified, affecting different parts of the model v_i, b_i, and c_i.[25]

First, some climate action can affect the material costs of future action, c_i by changing technology and the economic systems around it.[26] As technologies are developed and deployed, their costs descend down a "learning curve," becoming cheaper as more R&D is conducted and as production and distribution systems "learn by doing" and reach economies of scale.[27] Renewable energy technologies fit this pattern well. For example, the cost of photovoltaic cells dropped 75 percent from 2010-2015, and wind turbines 30-45 percent in the same period; both technologies are now at or below the cost of fossil fuel alternatives in many parts of the world, sharply altering the incentives for taking climate mitigation action.[28]

These cost reductions were only possible because of costly actions by first movers. For example, in the 1980s, California led the world in wind power installation. But the Californian market was served by Danish turbine producers, who began to interest the Danish government in local deployment of the technology, especially after the oil crisis, which led to the adoption of generous incentives. As the cost fell further, Germany began its energy transition in the 1990s, passing California as a world leader in 1996. The size of the German market, driven by its feed-in tariff, created a sharp reduction in cost, and a large market for Chinese producers, who began to dominate global manufacturing. Soon jurisdictions like China and Texas were installing more wind power than traditionally 'green' jurisdictions like Europe or California. This would have been unthinkable without first-movers. Between 1970 and 2000 California, Denmark, and Germany spent $986 million, $709 million, and $1.9 billion, respectively, on research and subsidies for wind power, which fell 85 percent in cost during this time.[29]

In addition to learning curves, many new technologies and business models demonstrate network effects; like telephones or email, the more people have them, the more useful they are, raising the private benefits b_i of such actions. Consider electric cars or solar rooftops. For the first movers, such products are very inconvenient because there are too few charging stations, maintenance technicians, or electric cables and pricing systems that allow homeowners to send power back to the grid. As market penetration increases, however, the enabling environment shifts and later adopters are well supported. Indeed, once network effects are strong enough, they may help to lock in new technologies as default options.

Second, prior action can also generate learning effects. Just as mitigation actions generate new technologies and business models that alter material costs, so too do they produce new knowledge about policy design and implementation through experimentation and diffusion, further reducing

[23] Urpelainen 2011; Levin, Cahsore, Bernstein et al. 2012; Urpelainen 2013; van der Ven, Bernstein and Hoffmann 2017

[24] Bernstein and Hoffmann 2018

[25] I model these effects by positing v_i, b_i, and c_i as functions of A. Other literature has instead considered accelerating versions of the production function P. This approach is less suitable for the current discussion because I am interested less in the rate at which contributions produce collective action and more in how contributions affect preferences. See for example Marwell and Oliver 1993; Heckathorn 1996

[26] Zenghelis, Fouquet and Hippe 2018, 33

[27] Hale and Urpelainen 2015

[28] IRENA 2017

[29] Sawin 2001

costs and enhancing benefits.[30] This effect is significant because many areas of mitigation involve complex policy instruments. Emissions trading systems are a prominent example, where even highly competent regulators like the European Commission have struggled to make their credit allocation and exchange systems operate smoothly. For this reason, the Chinese government has drawn on significant international expertise in the development of its national emissions trading system, and proceeded incrementally by first experimenting at the provincial level. In this way, previous actions generate epistemic resources that improve the ability of followers to emulate leaders.[31]

Third, over time growing actions may change norms around mitigation, altering not just costs and benefits, but also how actors value mitigation itself, v_i. International relations scholars have shown how "norm entrepreneurs" engage in "strategic social construction" by attempting to shift norms toward their policy goals.[32] In the realm of climate, Hoffmann and Bernstein argue that actions in one sphere, such as a city setting a climate target, can "normalize" low-carbon preferences in ways that spill over to other actors.[33] These models follow a logic of increasing returns. As Finnemore and Sikkink argue, norms progress through a life-cycle from emergence, to a "norm cascade" in which they become widely followed in practice, to internalization, in which they are embedded in the beliefs and preferences of most actors. As more action takes place, the more this self-reinforcing logic applies. While decarbonization norms are not currently widespread, some areas of climate mitigation show evidence of norm cascade dynamics, such as divestment from fossil fuel companies.[34]

Finally, reinforcing the above mechanisms, early action can affect the political processes of preference formation for states and other actors by generating new constituencies for action.[35] As new technologies emerge and grow, their producers and consumers develop a distributional interest in their continuance and expansion. At first these new interest groups are unlikely to be able to overcome established incumbents in political contestation. But in economic sectors or geographic regions where incumbents are relatively weak, the new entrants may thrive and, as action spreads, eventually acquire the size and clout to become politically competitive with incumbents in more and more jurisdictions and industries.

Through these various mechanisms, prior actions increase the ratio of benefits b_i to costs c_i for subsequent actions, as well as the value actors place on mitigation v_i. In other words, these parameters are only partially exogenous; to the extent increasing returns exist, they are functions of the previous level of A. Under these conditions actors' preferences shift dynamically in ways not accounted for in the tragedy of the commons framework.

2.4 A catalytic model of cooperation

The above three modifications change the classic model of collective action into a catalytic model of collective action, which can be expressed as:

The catalytic model of collective action:

$$g_i(A) = v_i[P(A)] - c_i(a) + b_i(a)$$

[30] Hoffmann 2011
[31] Biedenkopf, Müller, Slominski et al. 2017
[32] Finnemore and Sikkink, p. 895.
[33] Bernstein and Hoffmann 2018
[34] Green 2016, Ansar, Caldecott and Tilbury 2013
[35] Urpelainen 2013; van der Ven, Bernstein and Hoffmann 2017

Where:

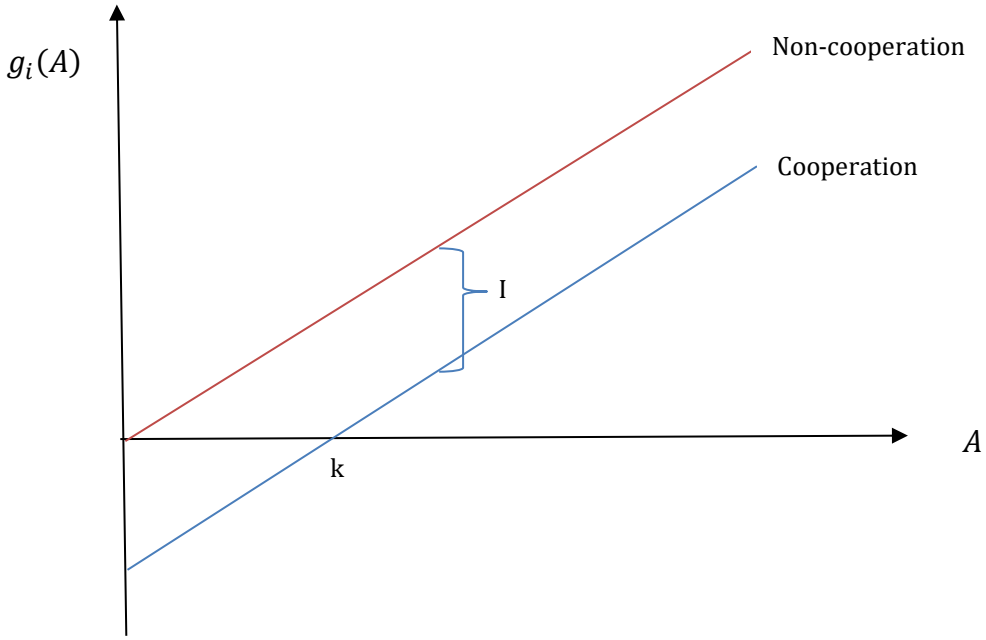
1. Private benefits exist, such that for some actors, $c_i(a) < b_i(a)$
2. Preferences vary such that, for some actors, $v_i[P(A)] > c_i(a)$ or, less restrictively, $v_i[P(A)] + b_i(a) > c_i(a)$
3. c_i, b_i, and v_i are partially functions of *A* such that
   $c_i = C(A)$, C is a decreasing function
   $b_i = B(A)$, B is an increasing function
   $v_i = V(A)$, V is an increasing function

These three revisions to the core assumptions of the tragedy of the commons model, each grounded in an empirical analysis of climate politics, create the possibility for a very different form of collective action.

3. Implications for cooperation

The tragedy of the commons model suggests dim prospects for cooperation, as can been seen clearly in Schelling's classic representation (figure 1).[36] The vertical axis represents the pay-off to actor *i* from collective action as a function of how much action others are contributing, which is measured along the horizontal axis. Actor *i* may either cooperate or not cooperate, and the payoff for each choice is shown on a different curve. Because the non-cooperation curve is always higher than the cooperation curve under the tragedy of the commons model, *i* almost never acts.

Figure 1: Standard cooperation model



An exception occurs if a single actor or coalition of actors can contribute at least *k*, the point where action and non-action yield the same payoff. For example, a hegemonic state or coalition of great powers may have an interest in providing a public good, even if it means other states will free-ride on them, because the actions of the "k-group" are themselves sufficient to generate enough of the collective good to be worth the costs.[37]

[36] Schelling 1978; Snidal 1985
[37] Snidal 1985

This special situation aside, actors will not cooperate under the classic model unless there is some other inducement to do so. This "inducement cost" *I* can be interpreted as the distance between the cooperation curve and the non-cooperation curve. To get actor *i* to cooperate, it would be necessary to offer it some benefit or to impose some cost equal to or greater than *I.* Side-payments, sanctions, issue linkage, or a credible pledge to unlock reciprocal cooperation from others are typically required to induce cooperation in these circumstances. For example, many treaties aim to impose sanctions on non-complying actors to lower the benefit they receive from defection.

How does the catalytic model described above change the prospects for cooperation? Three implications follow from the revised assumptions.

3.1 Unilateral cooperation

Students of collective action have shown how joint products and preference heterogeneity can facilitate cooperation.[38] In the catalytic model, these characteristics make action strictly preferable to inaction for some actors in two ways:

1. v_i may be so high that even a very low level of *A*, potentially just the actor's own contribution *a*, could produce a sufficient benefit to outweigh the cost $c_i(a)$ (heterogenous preferences)
2. The private benefit $b_i(a)$ of acting is higher than the cost $c_i(a)$ (joint products)

Combining these conditions, we can expect cooperation to be strictly preferable to non-cooperation whenever $v_i[P(A)] + b_i(a) > c_i(a)$. In Schelling's representation this "unilateral cooperation" can emerge as follows (figure 2).

First, as costs c_i fall and private benefits b_i rise, the cooperation curve shifts upward vertically, potentially falling above the non-cooperation curve. Under such conducive circumstances action will emerge readily. Even in the perhaps more common situation that costs remain higher than benefits, the catalytic model suggests that heterogenous preferences and joint products can reduce actors' inducement costs *I*.

Second, when actors value the benefits of mitigation more, higher values of v_i, say v_i', increase the slope of both the cooperation and non-cooperation curves. This shift rotates the curves counter-clockwise, which reduces the size of the k-group required to make cooperation attractive.

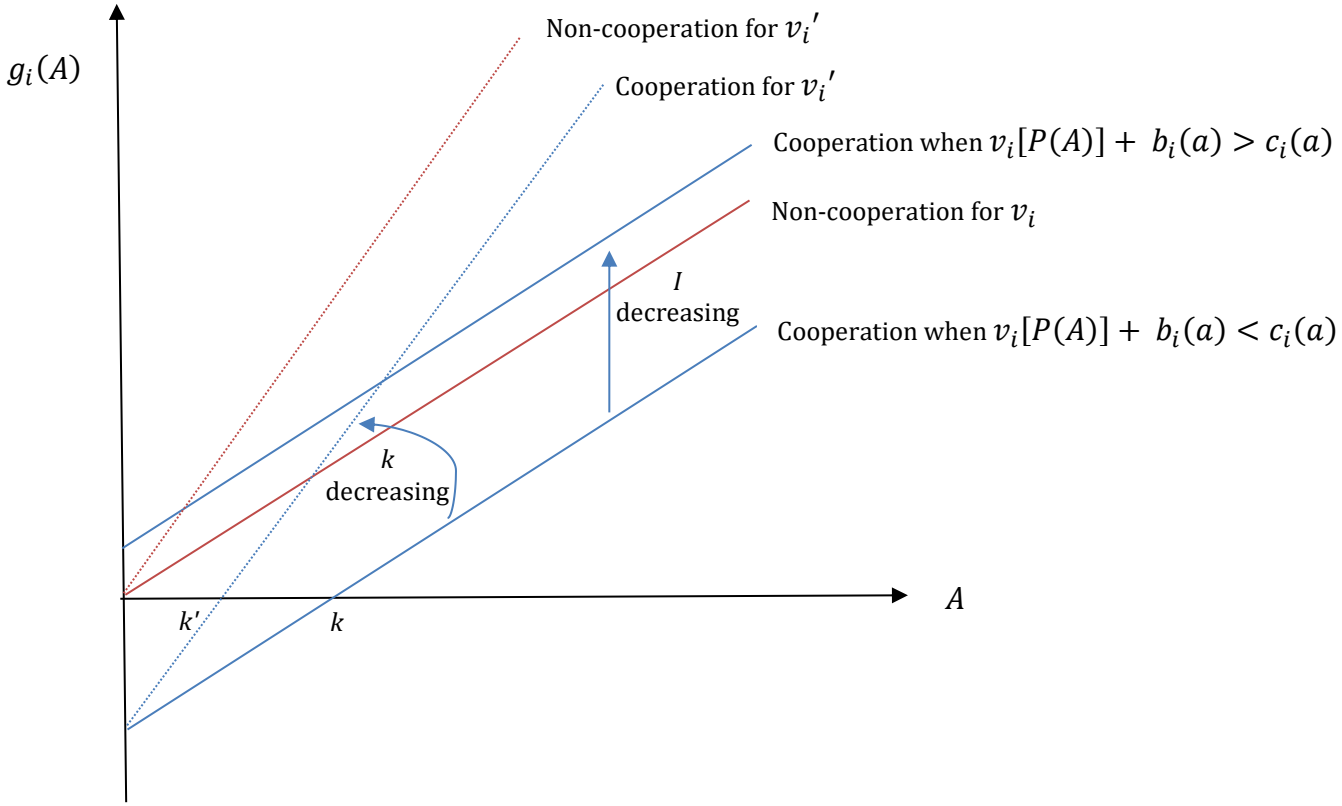
The inverse situation, when v_i and b_i are low and c_i is high, is also of theoretical interest. The standard approach to tragedies of the commons suggests that some sort of collective agreement is needed to create a certain k group or to generate the level of inducement (through reciprocal cooperation or sanctions) necessary to make cooperation superior to non-cooperation. But if v_i is very low, actors may struggle to create a k-group even if they could contribute quite a lot on their own. And if costs are very high relative to benefits, the level of inducement required may exceed the maximum that can be created through collective action. That is, if v_i is low enough, $v_i[P(A)] + b_i(a)$ may be less than $c_i(a)$ even at very high levels of *A*. Collective action is just not "worth it."

In the realm of climate change, this situation is likely applicable at least in part to a number of significant actors, such as corrupt regimes with large oil reserves or fossil fuel-dependent companies focused on quarterly earnings reports. Unlike many other commons, the benefits of mitigating climate change are distant; they are damage avoided in the future, not fish, lumber, or grazing enjoyed in the present. As argued above, the actual benefit this provides, upon which actors in the present might free-ride, is highly variable across actors, and in many cases likely quite small. This

[38] Hardin 1982; Marwell and Oliver 1993

means that even if collective agreements to reduce emissions can be reached, they will be of little value to many actors, and will therefore struggle to incentivize cooperative behavior.

Figure 2: How heterogenous preferences and joint products can induce unilateral cooperation



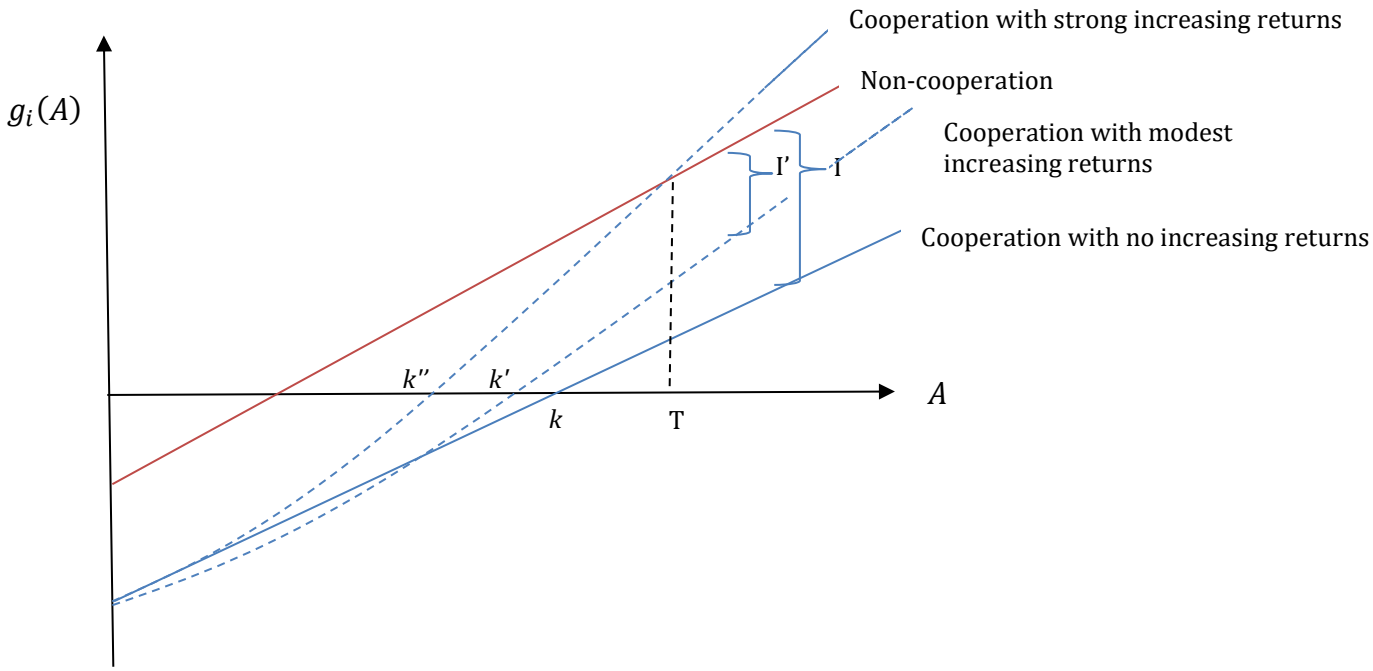
3.2 Increasing returns and tipping

If some actors are willing to act unilaterally, the dynamic nature of the catalytic model invites us to consider how this effect will vary over time. The key question is the extent to which increasing returns make v_i, b_i, and c_i partially endogenous to *A,* with the value of collective action and private benefits rising and costs falling as action expands. This change can be represented as an upward shift in the slope of the cooperation curve as action accumulates. In figure 3, as increasing returns strengthen, *k* shrinks to *k'* or even *k''*, and *I* shrinks to *I'*.

To the extent increasing returns exist, a different pathway emerges for achieving cooperation: unilateral cooperators can become "first movers" who reduce the costs and increase the benefits of acting for followers. As Marwell and Oliver note, "initial contributions lower the necessary interest for subsequent contributions."[39]

[39] Marwell and Oliver 1993, 88

Figure 3: Catalytic cooperation with increasing returns



Similar ideas inform "cascade" or "threshold" models that have been used in political science to describe norm diffusion, protests and revolutions, racial segregation of neighborhoods, treaties that require a minimum critical mass to go into effect, and other phenomena, and have been used in experimental settings to analyze the Paris Agreement.[40] But these applications tend to describe coordination games of various kinds, in which actors are trying to converge on an equilibrium or shift from one equilibrium to another. For example, Schelling's foundational study of neighborhood racial segregation showed how potentially even a small increase in the number of non-white residents could induce first the most anti-integration whites to move away, then the next-most racist ones, etc., potentially setting of a chain reaction that could change the racial composition of the entire neighborhood.[41] In such models, actors' preferences remain constant (though note that preference heterogeneity plays a key role); their choices change as others' behavior changes. In the catalytic model developed here, in contrast, actors' very preferences alter as a consequence of other's actions.

If increasing returns are sufficiently strong, the cooperation curve may intersect and rise above the non-cooperation curve at a certain level of *A*. If this happens, a "tipping point" *T* occurs at which action becomes strictly preferable and self-reinforcing.

The key question thus becomes, are the increasing returns strong enough, given the distribution across actors of their willingness to act at a certain level of A, to induce enough further action to create a tipping effect? And if it does not, does the cooperation curve at least bend close enough to the non-cooperation curve to reduce *I* or *k* to a manageable level?

3.3 How the number of actors affects cooperation

[40] Schelling 1971; Granovetter 1978; Kuran 1989; Finnemore and Sikkink 1998; Barrett 2003; Culpepper 2003; Barrett and Dannenberg 2017
[41] Schelling 1972

In the classic collective action approach, a larger number of actors makes cooperation more difficult. If actors are relatively similar, a larger number of them means that each one's individual contribution will diminish relative to the whole, making it harder to assemble a k-group. Moreover, the transaction costs of negotiating and implementing a collective agreement to induce cooperation are higher.[42]

The catalytic cooperation model, in turn, helps us identify conditions under which increasing the number of actors can instead facilitate cooperation. Building on Olson and Hardin, Marwell and Oliver show that "when groups are heterogenous and a good has high jointness of supply, a larger… group can have a smaller critical mass."[43] This seemingly heterodox logic is deceptively simple. Assume, conservatively, that preferences are distributed normally across actors. Following the logic of joint products and heterogenous preferences, raising the number of actors therefore increases the likelihood that at least a few actors will hold very pro-action preferences, because we become more likely to reach the "tails" of the distribution of preferences. Under a collective bargaining framework, this would do little to increase cooperation, because pro-action actors would, on average, be balanced by anti-action actors. But in a catalytic cooperation context where increasing returns apply, having a continuous chain of actors along the full preference spectrum—and especially at the pro-action extreme—is critical. This distribution of preferences increases the likelihood of at least one or more actors having a strong enough preference to act even when it is costly to do so.

This logic is of course even stronger if the newly added actors hold pro-action preferences. In climate mitigation, where sub- and non-state actors are now formally incorporated into the intergovernmental architecture, these favorable conditions apply. Cities, states and provinces, private businesses, investors, and other actors can hold not just different preferences on climate change, as argued above, but can substantially affect the problem through their own actions. Under these conditions, we can expect the number of first movers to be larger, and for their actions to be more consequential, helping generate the catalytic dynamics than can drive cooperation.

Furthermore, having more actors in the system makes it easier for increasing returns to set off a chain reaction. As Granovetter's threshold model shows, a chain reaction can occur when actors' thresholds for action (*k* or *T* in the graphs above) are sufficiently close together, such that each new action is enough to reach the next-most pro-action actor's threshold.[44] Assuming actors are normally distributed, increasing the number of actors means that it is less likely a large gap between thresholds will stall the effects of increasing returns.

Note that the magnitude of these effects is conditioned by what Sandler calls the "aggregation technology," which can be interpreted here as the nature of the production function P.[45] When collective action is "summative," i.e., when every contribution helps create the public good, an actor can always get a bit of public benefit from its own contribution, no matter how small. But when a certain threshold of contributions must be reached before any meaningful collective benefit is created, actors, especially small ones, will not obtain any broader benefit v_i form unliteral action.

3.4 Catalytic strategies for cooperation and the conditions under which free-riding does and does not matter

Each of the three mechanisms described above shows how cooperation might emerge, given the assumptions of the catalytic cooperation model. While each individually highlights a potentially

[42] Olson 1971
[43] Marwell and Oliver 1993
[44] Granovetter 1978; Macy 1991
[45] Sandler 2004, 68

sufficient mechanism for collective action, they are also additive and complementary, together comprising a catalytic pathway to cooperation:

1. Joint products and preference heterogeneity mean that action will be strictly beneficial for some actors, creating a set of first-movers.
2. Increasing returns make action more attractive, over time, for actors with less pro-action preferences, reducing the size of the k-group and the inducement cost, and potentially "tipping" further actors into unilateral cooperators.
3. Increasing the number of actors increases the pool of potential first movers needed to kickstart the catalytic effect, particularly if new actors hold strong pro-action preferences, and allows increasing returns to diffuse more efficiently.

To the extent these conditions apply, free-riding is not the central barrier to cooperation.[46] When costs fall and benefits rise such that the cooperation curve rises above the non-cooperation curve, *I* falls to zero or below. Action or inaction by others will have no effect on an actor's decision to act or not. Similarly, as the value of collective action v_i rises, *k* falls. It becomes increasingly likely that an actor's own action will create enough collective benefits to justify the cost to them. And to the extent increasing returns obtain, both these effects get stronger as action builds, meaning even recalcitrant actors could be "tipped" into unilateral cooperation.

Just as the catalytic strategies complement each other, they can also potentially combine with "traditional" approaches to cooperation, such as collective agreements to ensure that the benefits of collective action are realized, sanctions against defectors, provision of exclusive benefits, etc. Indeed, such strategies will be needed if there is not a sufficient mass of unilateral cooperators, or if increasing returns are too weak, or if the distribution of actors remains so lumpy that a k group is not created or a tipping point is not reached. The implication is that strategies to increase private benefits, reduce costs, and increase the value of collective benefits may be as useful for creating cooperation as solving the free-rider problem, even if they cannot eliminate it.

Note that, in this conception, the meaning of cooperation expands slightly beyond Keohane's strict definition of cooperation as mutual policy adjustment.[47] The catalytic model focuses attention on how some actions that contribute to common goals may emerge unconditionally, how such actions may influence each other and grow over time ("mutual" in an indirect sense), and how these changing circumstances will affect the prospects of (intentional) mutual policy adjustment going forward. In other words, the effect of the catalytic model is on actors' propensity to engage in cooperative behavior. If the catalytic effect is large it may obviate the need for Keohanian cooperation (if, for example, a tipping point is reached for enough large actors), or it may create conditions under which cooperation through mutual policy adjustment becomes more likely.

4. Catalytic institutions

What role can international institutions play when commons problems are characterized by the catalytic cooperation model? The standard institutional solution corresponds to the classic model of collective action, a collective agreement that specifies how much each actor must contribute such that $v_i[P(A)] > c_i(a)$. Typically, this requires a common agreement on how much action will be taken by each actor, monitoring to detect defection, and credible inducements to avoid of defection (such as withdrawing cooperation, or providing sanctions or side payments). Indeed, these features describe the basic contours of the Kyoto Protocol in the climate realm and many other agreements to regulate the global commons.

[46] Marwell and Oliver 1993, 182
[47] Keohane 1984

As Schelling wrote following the Bush Administration's "un-signing" of the Kyoto Protocol, such an approach has not proven workable in climate mitigation.[48] The Paris Agreement, however, is markedly different, allowing actors to set their own targets and not imposing penalties for failing to meet them (the monitoring requirement is however similar in both treaties). If we see climate only through a tragedy of the commons lens, we likely interpret Paris as a relatively modest achievement. For example, Keohane and Oppenheimer argue that Paris uses "discretion and vagueness" to make countries comfortable with making commitments.[49] Future commitments will therefore be determined through domestic politics. In a similar vein, Victor notes that "flexibility offers a way to get started and build confidence that, in time, will beget more confidence and a willingness to do more," but that "eventually a much more integrated global treaty will be needed to make major cuts in the greenhouse gases."[50] Bang et al. are more critical, noting that "Paris does little to restructure states' incentives so as to avoid free riding," though they hope actors will become more pro-climate over time.[51] Barrett and Dannenberg present experimental evidence that pledge and review systems can help increase contributions, but not to the threshold required to reach the goal they have set.[52] Following these arguments, we can see Paris is a positive but ultimately modest step.

The catalytic model offers a different interpretation of the Paris architecture. This section shows how, when the conditions specified by the catalytic model are met, institutions can drive collective action by:

1. Stimulating first-movers and incremental action through flexibility
2. Iteration of commitments
3. Increasing the effect of prior action on subsequent action via:
    a. Material transfers to alter future preferences and capacities
    b. Transferring experiences to shape the informational environment
    c. Normative goal-setting and benchmarking
    d. Domestic constituency-building

Below I lay out the causal logic of each mechanism, illustrating with examples from both the new architecture of the climate regime and other international institutions. Many regimes include catalytic institutions alongside institutions that follow other approaches as well. The Paris Agreement is distinctive, however, in that it places the catalytic logic at its core, constituting not just a catalytic institution but perhaps the first major catalytic regime.

4.1 Stimulating first-movers and incremental action through flexibility

In the catalytic cooperation model, the most important challenge is to reach a critical mass of action. Catalytic institutions therefore seek to stimulate first-movers to come forward and to encourage small steps even from recalcitrant actors.

Flexible, unconditional commitments provide an important way to do this. By lowering the cost of action to what actors are willing to do, flexible commitments can help create the initial moves required to initiate increasing returns. If the bar for cooperation is too high (e.g. a binding commitment of the kind required for a regulatory solution), i.e. a minimum a is required, these "small steps" might never occur.[53] Moreover, by explicitly allowing variation in these initial moves, flexible commitments allow the most pro-mitigation actors to put forward ambitious commitments,

[48] Schelling 2002
[49] Keohane and Oppenheimer 2016b, 10
[50] Victor 2015
[51] Bang, Hovi and Skodvin 2016
[52] Barrett and Dannenberg 2016
[53] Victor 2011; Urpelainen 2013

instead of limiting themselves to a least common denominator negotiated outcome. Under a flexible framework, we can expect actors to take actions when $v_i[P(A)] + \ b_i(a) > \ c_i(a)$.

Voluntary, flexible commitments are often seen as weak tools of global governance. They do not give actors incentives to do what they would otherwise not, but rather codify "business as usual." Still, even outside a catalytic context, they may be helpful for promoting at least modest cooperation by helping to build trust.[54] Indeed, "soft law" commitments and pledge and review systems are a growing phenomenon across nearly all spheres of world politics.[55] While the pledges delivered at Paris were not sufficient to achieve the Agreement's goals, they represent the single largest commitment to mitigation ever made, perhaps enough to lower global temperature rise by 1-2C if implemented.[56] Critically, for the first time, all major emitters made a commitment.[57]

In addition, the Paris Agreement seeks to spur first-movers amongst cities, businesses, and other actors, explicitly inviting them to align with the long-term goals. Again, this is facilitated by the flexible nature of the Paris commitments, which allows for a variety of commitments and actions at varying degrees of ambition to be made. As argued above, increasing the number of actors in this way helps lubricate the catalytic mechanisms by increasing the number of unilateral cooperators and helping increasing returns diffuse by shrinking the gaps between actor's thresholds for action.

4.2 Iteration of commitments

Even though flexible commitments capture only what actors are willing to do, with increasing returns what actors are willing to do changes over time, allowing the functions $C(A)$, $B(A)$, and $V(A)$ to drive down costs and increase benefits. Catalytic institutions therefore create an ongoing process to stimulate and capture updates in actors' preferences over time. As Macy notes in "cascade" models of collective action, it much easier to achieve cooperation when choices are serial rather than parallel.[58]

Iteration of commitment making is, of course, a common feature of international institutions. The global trade regime has developed through progressive trade rounds, and the "convention and protocol" approach to global commons issues like the ozone regime typically involves a series of increasingly stringent negotiated agreements. While these agreements rely on collective agreements enforced with sanctions, the broader regimes also have catalytic elements in the sense that past cooperation alters future preferences. In trade, for example, the expansion of multinational companies enabled by early trade rounds re-shaped the domestic politics of economic openness in major economies by creating powerful new constituencies for deeper openness.[59] Paris replicates this logic but with individual (for both parties and non-party actors), as opposed to negotiated,

[54] Keohane and Oppenheimer 2016a
[55] Abbott and Snidal 2000
[56] UNFCCC 2015
[57] Flexible commitments also enhance cooperation by making action more resilient in the face of backsliding. In a tragedy of the commons framework, or under a regulatory or club arrangement, one actor's cooperation is conditional on others' cooperation. Defection is therefore toxic to future action. With flexible arrangements, such as those in polycentric or catalytic arrangements, this is not the case. Since countries commit to doing what they would do anyway, other countries' failure to deliver on their commitments does not create an incentive to shirk. This dynamic has been observed in countries' (and other actors') reactions to the Trump Administration's announced withdrawal from the Paris Agreement. After the announcement in 2017, every major emitter reiterated its commitment to Paris, and the G7 and G20 issued exceptional "G6" and "G19" statements outlining further steps to implement the Agreement.
[58] Macy 1991
[59] Milner 1987; Bailey, Goldstein and Weingast 1997

commitments, which allows it to capture "updates" in countries' preferences more quickly and easily than would be possible with successive negotiated protocols.

4.3 Increasing the effect of prior action on subsequent action

As discussed above, many of the mechanisms through which increasing returns accrue fall outside the realm of international institutions. For example, reductions in the price of clean technologies are transmitted through markets and social norms diffuse through a wide range of media. It is important to recognize that international institutions can also play a number of complementary roles in these processes.

Material transfers to alter future preferences and capacities

Many international agreements involve resources transfers, for example payments from one country to another to host a military base. Such transfers are catalytic when, above and beyond providing a direct side-payment, they increase the capability of the recipient to cooperate by affecting v_i, c_i and b_i. For example, aid that helps pay for a renewable energy project can be seen as a side payment to induce cooperation. How much is needed will depend on the value of *I* for the country in question. In contrast, capacity building support that improves a country's ability to run effective tendering processes for energy procurement reduces c_i, making subsequent actions easier as well. Similarly, technical support that gives a country a better sense of what impacts it will suffer from climate change may increase the value v_i it places on mitigation efforts.

Indeed, while the Paris Agreement did not contain any firm commitments to increase funding for developing countries, it did create a Paris Committee on Capacity Building that will propose a new set of institutional arrangements to support developing countries' ability to formulate and implement national climate policies. It remains to be seen, however, whether this process will actually generate resource flows that have a catalytic effect. At the same time, many of the transnational networks of sub/non-state actors aim at peer-to-peer capacity building that can also have catalytic effects.[60]

Sharing experiences to shape the informational environment

Many international institutions seek to generate and transfer information about state behavior and the underlying policy dilemma in order to enhance compliance. As discussed in section three, such information is catalytic when it creates demonstration and learning effects that boost actors' ability and willingness to undertake mitigation actions, potentially reducing c_i and boosting b_i.

Several mechanisms in the Paris Agreement help to generate and diffuse these epistemic resources. First, flexible commitments, in addition to stimulating first-movers and incremental action, can help drive the policy innovation through experimentation.[61] Because countries are free to develop their own mitigation plans, reflecting their own conditions and interests, the NDCs contain a very diverse array of policies and measures. As the difficult work of implementing these plans progresses, countries will gain useful experience of what works and what does not.

A second important way through which catalytic institutions generate and diffuse information is via review processes.[62] These are common features of many international regimes, and are often oriented toward compliance, gathering information about countries' behavior in order to make sure

[60] Bulkeley, Andonova, Betsill et al. 2014
[61] Sabel and Victor 2015
[62] Victor, Raustiala and Skolnikoff ; Sabel and Victor 2015; Abbott 2017; Aldy 2018

they are implementing commitments. But review processes can also have a learning effect on states and other actors. Through review, states and outside experts gather information about state behavior vis-à-vis an international obligation, generating and transmitting knowledge about how best to approach the problem. Peer-to-peer transgovernmental networks have been shown to employ particularly influential versions of this type of review.[63] In the environmental realm, the technical committees of the Montreal Protocol's provided an expert-oriented version of implementation review that was seen as particularly effectual.[64]

The Paris Agreement includes various review processes, including review of individual countries' implementation of NDCs (Art. 13), review of aggregate progress toward the long-term goal in a global stock-take (Art. 14), and a "non-punitive" enforcement review designed to troubleshoot barriers to NDC implementation (Art. 15). At the time of writing, these mechanisms remain under negotiation, so it is difficult to know how catalytic they will ultimately be.[65]

Third, including sub- and non-state actors in the intergovernmental regime is specifically intended to increase the generation and transmission of epistemic resources. The large number and diversity of sub- and non-state actors makes them excellent laboratories for climate policy. Moreover, information diffusion is a major function of transnational climate governance initiatives, as cities, companies, and other actors experiment and seek common ways to advance their agendas.[66] The Paris system does not just celebrate and encourage sub- and non-state climate action, but tries to maximize its epistemic benefits.[67] As it does for countries, it creates structures to review and extract lessons from non-state actors, including the NAZCA [68]platform, the climate action events at Conferences of Party (COPs), and an annual Yearbook of Climate Action. These review and tracking tools serve as soft accountability mechanisms, but also seek to draw out the epistemic benefits of sub- and non-state actors' innovations and experimentation. Moreover, it explicitly links these tools to the technical examination process that is meant to provide countries with concrete policy options to enhance their own climate action. Still, as Abbott notes, there is potential for the UNFCCC and other actors to play a more active role in enhancing the catalytic effect of these transnational elements of the regime.[69]

Normative goal-setting and benchmarking

Section two noted how increasing action can normalize itself, potentially generating norm cascades, and therefore altering actors' preferences in favor of further action. Catalytic institutions seek to drive such processes through goal setting and benchmarking.

Countries often declare their collective intent to eradicate a disease, uphold human rights, or provide other global public goods, such the Millennium Development Goals and Sustainable Development Goals. Article 2 of the Paris Agreement reaffirms countries' commitment to limit temperature change

[63] Slaughter 2004
[64] Victor, Raustiala and Skolnikoff 1998
[65] The experience of previous mechanisms along these lines in the UNFCCC such as the technical examination process, the technology mechanisms, and the review under the Kyoto Protocol have not been shown to be particularly efficacious for learning.
[66] Roger, Hale and Andonova 2017
[67] Van der ven et al. 2017 term these "orchestration platforms."
[68] NAZCA is a climate action tracking platform captures the commitments to climate action by companies, cities, subnational, regions, investors, and civil society organizations. See http://climateaction.unfccc.int.

[69] Abbott 2017

to "well below" 2C, aiming toward 1.5C, which, Article 4 notes, requires making sure no more GHGs are going into the atmosphere than are coming out by the middle of the 21st century.

As many observers have noted, absent concrete plans and accounting, goal-setting can be cheap talk or even intentionally dissembling.[70] Nonetheless, scholars have identified various conditions under which, and mechanisms through which, goal-setting can affect political behaviour. First, a goal provides a focal point around which actors can converge. This mechanism is unlikely to sway actors who do not wish to cooperate, but it can enhance efficiency and enable benchmarking amongst first-movers.[71] As Young notes, goal-setting can have a significant effect on actors guided by a logic of appropriateness, which, it was argued above, a non-trivial number of actors in the climate realm are.[72] Second, international goals can enhance the leverage of pro-cooperation constituencies in domestic politics, intra-firm deliberations, or other spheres of political contestation. To the extent states or other actors are sensitive about reputational critiques, explicit goal-setting raises the costs of non-compliance, and gives pro-compliance groups a "hook" for their arguments.[73] In this way goals can bolster norm diffusion by codifying and legitimating certain policy preferences and giving them the imprimatur of international consensus.[74]

Once a normative goal has been set, it becomes possible to review actors' progress toward that goal. Benchmarking systems—which grade states against some standard and compare their compliance to others—are now common tools of multilateral institutions (like the World Bank "Doing Business" indicators), NGO advocacy (like the Transparency International corruption index), and bilateral diplomacy (like the US human trafficking scorecard).[75] When combined with a long-term goal, as in the Paris Agreement, such tools can increase the reputational benefits for leaders and, commensurately, the costs for laggards. To the extent actors are motivated by their reputations, such rankings can have a catalytic effect.

While some advocates proposed creating a grading system for NDCs in the Paris Agreement as part of the enhanced transparency framework, countries balked at exposing their "nationally determined" contributions to the collective judgement of their peers or others (suggesting the power of such ranking systems). Still, Article 13 of the Paris transparency framework requires countries to report on their progress toward their NDCs, and thus creates the conditions under which third parties can compare and rank national ambition.[76]

Domestic constituency-building

Catalytic institutions also shift preferences by reaching into the black box of domestic politics and building new constituencies for greater climate action. While orchestration is widespread across many areas of world politics, the climate regime is perhaps unique in the extent to which it has aimed to catalyse non-state action, as key policy entrepreneurs engaged in "webcraft" to orchestrate new actors to complement intergovernmental diplomacy.[77] While sub- and non-state and transnational governance of climate change had been building over the course of the regime, in the lead up to Paris the UN Secretary-General, the UNFCCC Secretariat, and the Peruvian and French hosts of COP20 and COP21, respectively, took a much more purposeful approach, mobilizing dozens

[70] Downs, Rocke and Barsoom 1996
[71] Urpelainen 2013
[72] Young 2017
[73] Thomas 2001
[74] Young 2017
[75] Kelley 2017
[76] van Asselt 2016
[77] Hale and Roger 2014; Hale 2016; Slaughter 2017

of initiatives that ultimately came to include over 10,000 actors.[78] The Paris Agreement and COP22 in Marrakech then codified this engagement, creating two "Climate Champions" to orchestrate further sub- and non-state action, an annual segment of the COP for sub- and non-state actors to report on progress and make new announcements, a tracking system to monitor implementation, and explicit links between sub- and non-state climate action and national policy options in the technical examination process.

As described above, in addition to affecting climate mitigation directly, and transmitting material and informational resources, orchestration may also have broader catalytic effects by influencing the process of national preference formation. The more sub- and non-state actors in a country take ambitious climate action, the more they increase the likelihood of both their peers and the national government adopting strong climate policies as well.[79]

5. Conclusion: where and when can catalytic institutions work?

The article has advanced two theoretical arguments regarding international cooperation. First, it identified three features of international climate mitigation—joint products, preference heterogeneity, and increasing returns—that depart substantially from the logic of the tragedy of the commons. These aspects comprise a model of catalytic cooperation. The argument is not that global climate mitigation has no resemblance to a tragedy of the commons, but rather that choosing this, and only this, analytic model delivers a very partial understanding of global climate politics.

Second, the article has argued that, under such conditions, catalytic institutions like those in the emerging climate regime can help drive international cooperation. Such institutions seek to initiate and stimulate early action, to iterate commitments over time, and to maximize the positive effect of early action on the strategies and preferences of actors in later rounds.

Of course, it is not yet clear how effective the catalytic properties of the Paris Agreement will be. At the time of writing, many of the details of the pledge, review, and ratchet process remain under negotiation. Still, the fundamental shift in the logic of the regime, after many years of gridlock, raises two questions of scope. Could such a catalytic approach work in other issue areas? And, given that the climate regime has shifted its logic of cooperation once, is it likely to evolve into something else in the future?

The theoretical arguments above suggest that the use of catalytic institutions could be increased in many areas where international cooperation dilemmas exhibit characteristics of the catalytic cooperation model, i.e. joint products, preference heterogeneity, and increasing returns. As section three discussed, these features imply conditions under which catalytic cooperation becomes possible:

1. When there is high variation in actors' preferences, with at least some willing to act unilaterally (in other words, on average, the costs and benefits of free-riding are low).
2. Prior action alters the costs, benefits, and preference formation processes around subsequent actions, potentially creating tipping points
3. Many different actors can affect the problem

Clearly, many areas of world politics do not fit this description. For example, the cost/benefit of free-riding in the realm of nuclear proliferation is high. If one country agrees to take in a number of refugees, it does not lower the cost for other countries to take in additional refugees in the future.

[78] GGCA 2016
[79] Urpelainen 2009; Bromley-Trujillo, Butler, Poe et al. 2016; Cao and Ward 2017

And while many actors are implicated by trade policies, only sovereign states can raise or lower tariffs.

Still, many other areas of world politics, and especially the realms of development, environment, and social welfare, satisfy these conditions. For example, of the 17 Sustainable Development Goals, areas like health, gender equality, water and sanitation, food security, energy, urbanization, and responsible consumption and production all are relatively insulated from free-riding concerns and affected by a wide range of actors with highly variable preferences. These tend to be, like climate change, classic "intermestic" issues that have domestic political dimensions but also spill across borders.

In addition to variation across issue areas, it is interesting to consider how the problem structure of a single issue area may shift over time. While some features of an issue are innate (e.g., the principles of atmospheric chemistry that cause GHGs emissions anywhere to affect the climate everywhere), many are subject to change, such as the kind and number of actors' whose behavior matters, the balance of power and interests between them, and the preferences they hold. Such factors are not just changeable, but potentially endogenous. Indeed, the entire purpose of a catalytic regime is to shift actors' preferences over time in favor of cooperation. In other words, if the Paris Agreement succeeds, it will do so by (eventually) altering parts of the problem structure of climate mitigation itself. Put another way, to the extent catalytic cooperation succeeds, climate change will resemble a tragedy of the commons less and less.

Various outcomes are possible in the medium- to long-term. Most optimistically, the result could be a harmonious world in which every actor of consequence "tips" and pro-climate preferences diffuse globally. Under such conditions, the problem structure would change into a coordination game, and we would expect actors to create institutions to provide focal points and facilitate implementation.[80]

Alternatively, or in a previous phase, we may see some particularly recalcitrant actors, such as states whose political economies depend heavily on fossil fuels, fail to shift toward the Paris goals, even as the majority of other actors do. This could lead to a problem structure not dissimilar to that of nuclear proliferation or state support for transnational terrorist groups, where action by a few "rouges" imposes externalities on others. In this more conflictive scenario, a large bulk of states could have an incentive to develop more coercive club institutions to exclude and penalize the recalcitrant emitters, for example through carbon tariffs or other economic sanctions (equal to *I* in the models above). In this way, a catalytic logic may lead over time to exactly the conditions needed for the emergence of a climate club that could credibly use negative incentives to enforce cooperation.

Of course, it may also be the case that the Paris Agreement will fail to catalyze sufficient action to force a critical mass of actors to raise their mitigation ambition. Under this scenario, the climate regime will look much as it did after Copenhagen, characterized by a flexible multilateral process and increasing pluralism. As the polycentric logic suggests, this will involve many different approaches across different actors reflecting their individual preferences and strategies, but may not add up to a global solution.

While the article has identified how climate mitigation fits a tipping-point problem structure and ways through which catalytic institutions seek to respond to this problem structure, the theory does not a priori predict the extent to which they will succeed. However, it provides a conceptual framework that gives institutional designers an additional strategy to solve collective action problems under conditions of catalytic cooperation.

[80] Barrett 2003, Chap. 9

Bibliography

Abbott, K. W. (2017). "Orchestrating Experimentation in Non-State Environmental Commitments." Environmental Politics **26**(4): 738-63.

Abbott, K. W. and D. Snidal (2000). "Hard and Soft Law in International Governance." International Organization **54**(3): 421-56.

Aldy, J. E. (2018). Policy Surveillance: Its Role in Monitoring, Reporting, Evaluating and Learning. Governing Climate Change: Polycentricity in Action. A. Jordan, D. Huitema, H. van Asselt and J. Forster. Cambridge, Cambridge University Press.

Ansar, A., B. Caldecott and J. Tilbury (2013). Stranded Assets and the Fossil Fuel Divestment Campaign: What Does Divestment Mean for the Valuation of Fossil Fuel Assets? Oxford, Stranded Assets Programme: University of Oxford.

Axelrod, R. (1985). "The Evolution of Cooperation."

Bailey, M. A., J. L. Goldstein and B. R. Weingast (1997). "The Institutional Roots of American Trade Policy: Politics, Coalitions, and International Trade." World Politics **49**(3): 309-38.

Bang, G., J. Hovi and T. Skodvin (2016). "The Paris Agreement: Short-Term and Long-Term Effectiveness." Politics and Governance **4**(3): 209-18.

Barrett, S. (2003). Environment and Statecraft. Oxford, Oxford University Press.

Barrett, S. and A. Dannenberg (2016). "An Experimental Investigation into 'Pledge and Review' in Climate Negotiations." Climatic Change **138**: 339-51.

Barrett, S. and A. Dannenberg (2017). "Tipping Versus Cooperating to Supply a Public Good." Journal of the European Economic Association **15**(4): 910-41.

Bayer, P. and J. Urpelainen (2016). "It Is All About Political Incentives: Democracy and the Renewable Feed-in Tariff." Journal of Politics **78**(2): 603-19.

Bernstein, S. and M. Hoffmann (2018). "The Politics of Decarbonization and the Catalytic Impact of Subnational Climate Experiments." Policy Sciences **51**(2): 189-211.

Biedenkopf, K., P. Müller, P. Slominski and J. Wettestad (2017). "A Global Turn to Greenhouse Gas Emissions Trading? Experiments, Actors, and Diffusion." Global Environmental Politics **17**(3): 1-11.

Bromley-Trujillo, R., J. S. Butler, J. Poe and W. Davis (2016). "The Spreading of Innovation: State Adoptions of Energy and Climate Policy." Review of Policy Research **33**(5): 544-65.

Broz, J. L. and D. Maliniak (2010). "**Malapportionment, Gasoline Taxes, and Climate Change**." APSA 2010 Annual Meeting Paper.

Bulkeley, H., L. B. Andonova, M. Betsill, D. Compagnon, T. Hale, M. Hoffmann, P. Newell, M. Paterson, C. Roger and S. D. VanDeveer (2014). Transnational Climate Change Governance. Cambridge, Cambridge University Press.

Bulkeley, H. and M. Betsill (2003). Cities and Climate Change: Urban Sustainability and Global Environmental Governance. London, Routledge.

Cao, X. and H. Ward (2017). "Transnational Climate Governance Networks and Domestic Regulatory Action." International Interactions **43**(1): 76-102.

Caporaso, J. A. (2007). "The Promise and Pitfalls of an Endogenous Theory of Institutional Change: A Comment." West European Politics **30**(2): 392-404.

Cole, D. H. (2015). "Advantages of a Polycentric Approach to Climate Change Policy." Nature Climate Change **5**.

Culpepper, P. D. (2003). Creating Cooperation: How States Develop Human Capital in Europe. Ithaca, Cornell University Press.

Dolšak, N. and A. Prakash (2017). "Join the Club: How the Domestic Ngo Sector Induces Participation in the Covenant of Mayors Program." International Interactions **43**(1): 26-47.

Dorsch, M. J. and C. Flachsland (2017). "A Polycentric Approach to Global Climate Governance." Global Environmental Politics **17**(2): 45-64.

Downs, G., D. M. Rocke and P. N. Barsoom (1996). "Is the Good News About Compliance Good News About Cooperation?" International Organization **50**(2): 379-406.

Downs, G., D. M. Rocke and P. N. Barsoom (1998). "Managing the Evolution of Multilateralism." International Organization **52**(2): 397-419.

Dubash, N. K., M. Hagemann, N. Hohne and P. Upadhyaya (2013). "Developments in National Climate Change Mitigation Legislation and Strategy." Climate Policy **13**(6): 649-64.

Finnemore, M. and K. Sikkink (1998). "International Norm Dynamics and Political Change." International Organization **52**(4): 887-917.

Gerlach, L. and S. Rayner (1988). Managing Global Climate Change: A View from the Social and Decision Sciences. Oak Ridge, Oak Ridge National Laboratory. **ORNL/6390**.

Granovetter, M. (1978). "Threshold Models of Collective Behavior." The American Journal of Sociology **83**(6): 1978.

Green, F. (2015). "Nationally Self-Interested Climate Change Mitigation: A Unified Conceptual Framework." Center for Climate Change Economics and Policy **Working Paper No. 224**.

Green, F. (2016). "Anti-Fossil Fuel Norms." Forthcoming.

Hale, T. (2016). ""“All Hands on Deck”: The Paris Agreement and Nonstate Climate Action." Global Environmental Politics **16**(3): 12-22.

Hale, T. (2017). Climate Change: From Gridlock to Catalyst. Beyond Gridlock. T. Hale and D. Held. Cambridge, Polity Press.

Hale, T. and C. Roger (2014). "Orchestration and Transnational Climate Governance." Review of International Organizations **9**(1): 59-82.

Hale, T. and J. Urpelainen (2015). "When and How Can Unilateral Policies Promote the International Diffusion of Environmental Policies and Clean Technology?" Journal of Theoretical Politics **27**(2): 177 - 205.

Hardin, G. (1968). "The Tragedy of the Commons." Science **162**(1243).

Hardin, R. (1982). Collective Action. Baltimore, Johns Hopkins University Press.

Heckathorn, D. D. (1996). "The Dynamics and Dilemmas of Collective Action." American Sociological Review **61**(2): 250-77.

Hoffmann, M. (2011). Climate Governance at the Crossroads: Experimenting with a Global Response after Kyoto. Oxford, Oxford University Press.

Hsueh, L. (2017). "Transnational Climate Governance and the Global 500: Examining Private Actor Participation by Firm-Level Factors and Dynamics." International Interactions **43**(1): 48-75.

Iacobuta, G., N. K. Dubash, P. Upadhyaya, M. Deribe and N. Höhne (2018). "National Climate Change Mitigation Legislation, Strategy and Targets: A Global Update." Climate Policy.

IRENA. (2017). ""“Wind” and “Solar”." Retrieved December 27, 2017, from www.irena.org.

Jacobs, A. M. (2016). "Policy Making for the Long Term in Advanced Democracies." Annual Review of Political Science **19**(1): 433-54.

Jordan, A., D. Huitema, H. van Asselt and J. Forster (2018). Governing Climate Change: Polycentricity in Action. Cambridge, Cambridge University Press.

Kelley, J. G. (2017). Scorecard Diplomacy: Grading States to Influence Their Reputation and Behavior. Cambridge, Cambridge University Press.

Keohane, R. O. (1984). After Hegemony. Princeton, Princeton University Press.

Keohane, R. O. and M. Oppenheimer (2016a). Paris: Beyond the Climate Dead End through Pledge and Review? Discussion Paper 2016-85. Harvard Project on Climate Agreements, Belfer Center for Science and International Affairs, Harvard Kennedy School.

Keohane, R. O. and M. Oppenheimer (2016b). "Paris: Beyond the Climate Dead End through Pledge and Review?" Politics and Governance **4**(3): 142-51.

Keohane, R. O. and D. G. Victor (2011). "The Regime Complex for Climate Change." Perspectives on Politics **9**(1).

Kuran, T. (1989). "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution." Public Choice **61**(1): 41-74.

Lachapelle, E. and M. Paterson (2013). "Drivers of National Climate Policy." Climate Policy **13**(5): 547-71.

Levin, K., B. Cahsore, S. Bernstein and G. Auld (2012). "Overcoming the Tragedy of Super Wicked Problems:

Constraining Our Future Selves to Ameliorate Global
Climate Change." Policy Sci **45**(123-152).

Lipscy, P. Y. (Forthcoming). The Institutional Politics of Energy and Climate Change.

Macy, M. W. (1991). "Chains of Cooperation: Threshold Effects in Collective Action." American Sociological Review **56**(6): 730-47.

Marwell, G. and P. Oliver (1993). The Critical Mass in Collective Action. Cambridge, Cambridge University Press.

Milner, H. (1987). "Resisting the Protectionist Temptation: Industry and the Making of Trade Policy in France and the United States During the 1970s." International Organization **41**(4): 639-65.

Olson, M. (1971). The Logic of Collective Action: Public Goods and the Theory of Groups. Cambridge, Harvard University Press.

Ostrom, E. (1990). Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge, Cambridge University Press.

Ostrom, E. (2009). A Polycentric Approach for Coping with Climate Change. Washington, DC, World Bank Policy Research Working Paper No. 5095.

Pierson, P. (2000). "Increasing Returns, Path Dependence, and the Study of Politics." American Political Science Review **94**(2): 251-67.

Prakash, A. and M. Potoski (2006). The Voluntary Environmentalists: Green Clubs, Iso 14001, and Voluntary Environmental Regulations. Cambridge, Cambridge University Press.

Rayner, S. and M. Caine, Eds. (2015). The Hartwell Approach to Climate Policy. London, Earthscan.

Roger, C., T. Hale and L. B. Andonova (2017). "The Comparative Politics of Transnational Climate Governance." International Interactions **43**(1): 1-25.

Sabel, C. F. and D. G. Victor (2015). "Governing Global Problems under Uncertainty: Making Bottom-up Climate Policy Work." Climatic Change.

Sandler, T. (2004). Global Collective Action. Cambridge, Cambridge University Press.

Sawin, J. L. (2001). The Role of Government in the Development and Diffusion of Renewable Energy Technologies: Wind Power in the United States, California, Denmark and Germany, 1970–2000. PhD, Tufts University.

Schelling, T. C. (1971). "Dynamic Models of Segregation." The Journal of Mathematical Sociology **1**(2): 143-86.

Schelling, T. C. (1972). The Process of Residential Segregation: Neighborhood Tipping. Racial Discrimination in Economic Life. A. H. Pascal. Lexington, Lexington Books.

Schelling, T. C. (1978). Micromotives and Macrobehavior. New York, Norton.

Schelling, T. C. (2002). "What Makes Greenhouse Sense?" Foreign Affairs **81**(3): 2-9.

Slaughter, A.-M. (2004). A New World Order. Princeton, Princeton University Press.

Slaughter, A.-M. (2017). the Chessboard and the Web: Strategies of Connection in a Dangerous World. New Haven, Yale University Press.

Snidal, D. (1985). "The Limits of Hegemonic Stability Theory." International Organization **39**(4): 579-614.

Thomas, D. C. (2001). The Helsinki Effect: International Norms, Human Rights, and the Demise of Communism. Princeton, Princeton University Press.

Tobin, P. (2017). "Leaders and Laggards: Climate Policy Ambition in Developed States." Global Environmental Politics **17**(4): 28-47.

Underdal, A., S. Kallbenkken and J. Hovi (2008). "Exploring the Dynamics of International Cooperation." International Studies Association Annual Convention.

UNFCCC (2018). Yearbook of Climate Action. Bonn, United Nations Framework Convention on Climate Change.

Urpelainen, J. (2009). "Explaining the Schwarzenegger Phenomenon: Local Frontrunners in Climate Policy." Global Environmental Politics **9**(3): 82-105.

Urpelainen, J. (2011). "A California Effect for International Environmental Externalities." International Interactions **37**(2): 170-89.

Urpelainen, J. (2013). "A Model of Dynamic Climate Governance: Dream Big, Win Small." International Environmental Agreements **13**: 107-25.

van Asselt, H. (2016). "The Role of Non-State Actors in Reviewing Ambition, Implementation, and Compliance under the Paris Agreement." Climate Law **6**(1): 91-108.

van der Ven, H., S. Bernstein and M. Hoffmann (2017). "Valuing the Contributions of Nonstate and Subnational Actors to Climate Governance." Global Environmental Politics **17**(1): 1-20.

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

Victor, D. G. (2015). "Why Paris Worked: A Different Approach to Climate Diplomacy,." Yale360 2015.

Victor, D. G., K. Raustiala and E. B. Skolnikoff (1998). The Implementation and Effectiveness of International Environmental Commitments. Cambridge, MIT Press.

Young, O. R. (2017). Conceptualization: Goal Setting as a Strategy for Earth System Governance. Governing through Goals: Sustainable Development Goals as Governance Innovation. N. Kanie and F. Biermann. Governing through Goals: Sustainable Development Goals as Governance Innovation, MIT Press.

Zenghelis, D., R. Fouquet and R. Hippe (2018). Stranded Assets: Then and Now. Stranded Assets and the Environment: Risk, Resilience and Opportunity. B. Caldecott. London, Taylor and Francis.