

Cooperating to Avoid Catastrophe

Astrid Dannenberg

University of Kassel

Scott Barrett

Columbia University

Knowledge that humans could trigger a regime shift in a vital natural system may help in identifying a goal for collective action, but it is unlikely to spur the degree of cooperation needed to avert a catastrophe. Substantial shifts in human behaviour are more likely to come about by strategically manipulating the institutions that govern human action on the commons.

Human activities are now the major driver of change of Earth's natural systems, ranging from local ecosystems like lakes, estuaries, and forests, to global systems, like the climate, oceans, and biodiversity. In some cases, human pressure can cause a natural system to undergo a large, abrupt, and persistent if not irreversible shift. Many local natural systems are known to be prone to regime shifts¹; globally, there is concern that certain planetary boundaries may be exceeded². Species extinctions, ocean dead zones, and collapse of the West Antarctic Ice Sheet are prominent examples of natural systems that are prone to tipping.

Protection of natural resources often depends on users cooperating to limit their rates of exploitation. Cooperation is needed at the global level, because there is no 'world government.' It is also needed at the local level when governments are either unable or uninterested in regulating use of the commons. Importantly, users not only cause the conditions of the natural resource to change, but the conditions of the natural resource also affect how users behave towards the resource. Because crossing a critical threshold can have disastrous consequences, users have a strong collective incentive to ensure that such thresholds are not crossed. However, individual incentives may conflict with these collective incentives. For example, there is a political consensus of the need to avoid exceeding the 2 °C temperature target for "dangerous" climate change, but countries seem not to be taking the actions necessary to ensure that this target is met.

Laboratory experiments are a uniquely useful way to test whether and under what circumstances knowledge of critical thresholds is likely to change collective action. The advantage of experiments is that they allow behaviour to be observed under controlled conditions, free of confounding factors³. In threshold experiments, players choose how much to contribute to a public good, knowing that they will suffer a collective loss if the aggregate of their contributions falls short of a critical threshold (you can think of "contributions" as abatement of greenhouse gases and the threshold as the total abatement needed to avoid "dangerous" climate change). The threshold is known either with certainty or with uncertainty. If the threshold is uncertain, its realized value is

determined by a random draw from a known probability distribution. For example, if it is known that the threshold for “dangerous” climate change lies somewhere between 350 and 550 ppm CO₂, chance will determine its actual value—a value like 400 or 527 ppm.

Experiments show that the existence of a critical threshold spurs cooperation considerably when uncertainty about the threshold is very small and the expected benefits of reaching the threshold (avoiding “dangerous” climate change) are large relative to the associated costs⁴. The reason is that, when uncertainty about the threshold is very small, the actions of a single player can determine whether or not the threshold is breached, giving every player a strong incentive to play its part in avoiding the threshold. By contrast, when uncertainty about the threshold is large, free riding merely increases very slightly the probability that catastrophe will occur—an incentive that is too small to overcome the urge to free ride⁵. For ecosystems that are common, such as lakes, observations can be obtained about the stresses that cause tipping, allowing uncertainty about the critical threshold to be reduced significantly⁶. However, this opportunity to learn is lost when it comes to unique ecosystems and singular events such as climate change and mass extinctions. In these situations, uncertainty about the location of critical thresholds is to a large extent irreducible.

There also exist large uncertainties about the impact of crossing a critical threshold⁷. For example, it isn’t known precisely how long it would take for the West Antarctic Ice Sheet to collapse, or by how much sea level would rise as a consequence, or how difficult it would be for countries to adapt to this change. However, experiments reveal that this type of uncertainty matters much less for collective action than threshold uncertainty⁸. This is because players’ contributions can reduce the probability that a threshold in nature will be crossed but they can’t reduce the probability that the impact of crossing this threshold will be large *if and when the threshold is crossed*.

Building institutions

If science is unable to shrink uncertainty about the location of a critical threshold, the only way to bring about a change in behaviour is to design and deploy institutions that change the incentives for collective action. Recent experiments show that, under certain circumstances, humans are capable of choosing efficient institutions and achieving high levels of cooperation⁹. Although these experiments simplify the real-life process of institutional formation, they reveal important insights into behaviour by allowing us to explore the choices that individuals and groups make in different institutional settings and whether they choose to change their institutions over time as they learn from past experiences.

Suppose that there are two groups, each facing an internal cooperation problem; that one group allows players to punish each other and that the other one does not: and that individuals can migrate between groups if they wish. The punishment institution will then attract norm-enforcers, individuals who establish a social norm to cooperate and who punish free riders if needed. As this group is seen to earn higher payoffs, it will become a role model for the other players to follow. Given enough time, the efficient institution will thrive and the inefficient institution will disappear⁹.

Unfortunately, in many situations, most notably global cooperation problems, people cannot choose their group. In such situations, learning occurs only over time within the same group,

making it less likely that cooperative norms will emerge. A group may fail to implement an institution, believing that doing so wouldn't help, when doing so might have helped. Or the group may implement an institution but fare badly, because not enough members see the advantage in playing differently under this institution. Repeated failures may even create pessimism and cause groups to accept their fate¹⁰. Notwithstanding these difficulties, some global cooperation problems have been solved in the past, and these experiences offer important lessons for how to address current and future problems.

Learning from successes

Countries are free to join—or not to join—an international agreement to protect a shared resource as they please. Agreements must therefore be designed in a way that ensures a limited use of the resource *and* that gives countries an incentive to join the agreement and to comply with it. Two examples from history illustrate how this can be done.

The first example is the International Convention for the Prevention of Pollution by Ships, more commonly known as MARPOL. MARPOL establishes a technology standard for oil tankers, ensuring that a tanker's oil cargo is kept physically separate from its ballast water. Previously, most oil pollution in the oceans resulted from tankers flushing out their ballast water mixed with oil. The agreement also allows port states to protect their coasts by restricting entry to tankers meeting the standard. This approach created a “tipping” situation, thanks to shipping being a trading network: As more port states enforced the standard, more ship owners wanted to comply with the standard, and as more ships conformed to the standard, more port states wanted to enforce the standard. The mandated technology-standards approach wasn't cheap—it required more capital and reduced the cargo-carrying capacity—but by changing the incentives of the involved actors, it worked better than the approaches that had been tried before for many years. Today, virtually all oil tankers comply with the MARPOL standard.

The second example is the Montreal Protocol, the international agreement adopted to protect the stratospheric ozone layer. This agreement limits the production and consumption of ozone depleting substances, and bans trade in these substances and products containing these substances between parties and non-parties. The trade ban has had a transformative effect, creating another “tipping” situation: As more countries joined the agreement, the trade penalty facing non-parties increased. Once a critical number of countries participated, the trade penalty to being a non-party exceeded the gain from free riding, giving the remaining countries an incentive to join. Montreal's advantage over MARPOL was that it could create a “tipping” situation more cheaply—a reason perhaps, that Montreal incorporated trade sanctions from the beginning and quickly achieved a high level of participation.

Experiments in which groups choose whether to play a standard cooperation game or a “tipping” game confirm the importance of costs. Groups are quick to choose the tipping game and solve the collective action problem when doing so involves no cost. However, when playing the tipping game is costly and groups could *potentially* earn more when playing the standard cooperation game, groups choose the tipping game only reluctantly, after observing their inability to sustain

much cooperation in the standard game; in some cases, they never even try to play the tipping game despite cooperation in the standard game being low¹⁰.

Climate change, unfortunately, is a greater challenge than either oil pollution by ships or ozone depletion, and so far negotiators have failed to create a “tipping” situation for climate similar to MARPOL or Montreal. This is why participation in the Kyoto Protocol was far from complete and why greater participation in the Paris Agreement was achieved only by allowing parties to pledge voluntary contributions.

Rethinking negotiations

Human behaviour can seem puzzling. When a natural resource is subject to a regime shift, it may seem obvious that humans should act collectively to avoid the critical threshold. Indeed, from the perspective of their collective interests, uncertainty about the location of the threshold should cause players to act even more cautiously compared to a certainty situation, so as to ensure that the threshold is avoided. Unfortunately, evidence from lab experiments shows that humans, when acting independently, respond to such situations in the opposite way; they contribute even *less* to protect the resource when they should contribute *more*. Typically, what occurs in experiments is that the groups that have to act under threshold uncertainty propose that their group should make higher contributions, compared to the threshold certainty case, so as to avoid the collective loss, but in the end contribute much less than the amounts they proposed⁵. We observe similar behaviour in the real world with countries agreeing that temperature change should be kept well below 2 °C and then acting so as to guarantee that the temperature target is exceeded.

Climate change is different from marine pollution and ozone depletion, and can't be addressed in precisely the same ways. One way forward is to take the problem apart, looking for opportunities to create a “tipping” situation for individual sectors and gases separately. A small but important step in this direction is the Kigali Amendment to the Montreal Protocol, which phases down the production and consumption of hydrofluorocarbons, a chemical that does not destroy the ozone layer but that is a powerful greenhouse gas. Another possible route is to exploit countries' cooperation in other areas, such as trade, and make continuation of cooperation in such areas contingent on cooperation on climate change. This approach is bolder, but also riskier; rather than lift up cooperation on climate change, it might only jeopardize cooperation in other areas. Clearly, more research is needed into these and other possibilities, revealing both their opportunities and their risks.

References

1. Scheffer, M., Carpenter, S., Foley, J.A., Folke, C. & Walker, B. *Nature* **413**, 591-596 (2001).
2. J. Rockström *et al.*, *Nature* **461**, 472 (2009).
3. Falk, A. & Heckmann, J.J. *Science* **326**, 535-538 (2009).

4. Dannenberg, A. & Tavoni, A. in *World Scientific Reference on Natural Resources and Environmental Policy in the Era of Global Change* (World Scientific, 2017).
5. Barrett, S. & Dannenberg, A. *Proc. Natl. Acad. Sci. USA* **109**, 17372-17376 (2012).
6. Carpenter, S.R. et al. *Science* **332**, 1079-1082 (2011).
7. Stern, N. *Nature* **530**, 407–409 (2016).
8. Barrett, S. & Dannenberg A. *Nat. Clim. Change* **4**, 36–39 (2014).
9. Güreker, Ö., Irlenbusch, B. & Rockenbach, B. *Science* **312**, 108-111 (2006).
10. Barrett, S. & Dannenberg, A. *J. Eur. Econ. Assoc.* **15**, 910–941 (2017).

Competing Interests Statement: The authors declare no competing interests.