CESifo WORKING PAPERS

8944
2021

March 2021

Stable Partial Cooperation in Managing Systems with Tipping Points

Florian Wagener, Aart de Zeeuw

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8944*

Stable Partial Cooperation in Managing Systems with Tipping Points

Abstract

Tipping of a natural system, entailing a loss of ecosystem services, may be prevented by stable partial cooperation. The presence of tipping points reverses the grim story that a high level of cooperation is hard to achieve and leaves large possible gains of cooperation. We investigate a tipping game with constant emissions and a piecewise linear response, and the well-known lake system with concave-convex dynamics and time-dependent emissions. Tipping back, leading to a gain in services, can also be induced by stable partial cooperation, but is harder to achieve. A physically reversible natural system may prove to be socially irreversible.

JEL-Codes: C700, Q200.

Keywords: tipping points, multiple Nash equilibria, stable partial cooperation, ecological systems.

Florian Wagener
CeNDEF, ASE
University of Amsterdam
P.O. Box 15867
The Netherlands – 1001 NJ Amsterdam
F.O.O. Wagener@uva.nl

Aart de Zeeuw
Department of Economics and TSC
Tilburg University
PO Box 90153
The Netherlands - 5000 LE, Tilburg
A.J.deZeeuw@uvt.nl

We are grateful for discussions on this topic with Scott Barrett, Michael Finus and Paolo Zeppini. We have benefitted from presentations and comments of the participants at the EAERE conferences, 2017, 2018, the SURED and CESifo conferences, 2018, workshops in Warsaw, 2017, Thessaloniki, 2018, and seminars in Munich, 2017, Paris, 2018.

1. Introduction

Tipping points are observed in a variety of natural systems. At some point, a small increase in one of the inputs may have substantial consequences, as the natural system will shift to another domain of attraction with a big loss of ecosystem services (Scheffer et al., 2001, Biggs et al., 2012). If this happens, it is very difficult (hysteresis) or even impossible (irreversibility) to restore the original conditions of the natural system. A well-known example is the lake system, where at some point a small increase in phosphorus loading shifts the lake to a bad state, with a big loss of ecosystem services (Carpenter and Cottingham, 1997, Scheffer, 1997). A second example is the coral-reef system, where at some point a small increase in the temperature of the ocean shifts the coral reef to a bad state, with a big loss of coral and fish (Hughes et al., 2003). It is expected that the climate system has tipping points as well (Lenton and Ciscar, 2013).

Economic activities yield benefits but also release emissions on natural systems. At a tipping point, a marginal increase in economic activities, and thus in emissions, leads to a very large increase in costs, i.e. a sudden big loss of ecosystem services. The presence of tipping points is therefore a threat, but it may also have a positive effect. Barrett (2013) has shown that if it is optimal to avoid tipping, a Nash equilibrium may exist that avoids tipping as well, because the incentive to deviate is suppressed by the consequences of tipping. If such a Nash equilibrium exists, the game changes from prisoners' dilemma to coordination game. In a model of using a resource with a tipping point, Diekert (2017) has also considered the question when the first best of staying at the threshold can be sustained as a Nash equilibrium. That paper investigates experimentation beyond the safe point, in case the threshold is uncertain. However, if such a Nash equilibrium does not exist, the question

is up to which level cooperation is needed to avoid tipping, and whether this partial cooperation is stable or not.

The basic idea is to use the two-stage membership game developed in cartel theory (d'Aspremont et al, 1983) and in the theory on international environmental agreements (Hoel, 1992, Carraro and Siniscalco, 1993, Barrett, 1994). In the first stage a coalition is formed, and in the second stage a Nash equilibrium results between the coalition and the individual outsiders. Depending on the size of the coalition, this Nash equilibrium may avoid tipping in case the Nash equilibrium between the individual economic agents does not avoid tipping. However, the coalition has to be stable in the sense that there is no incentive in the first stage to leave or to join the coalition. Stability is achieved if the incentive to cooperate is at least as large as the incentive to free ride. This paper shows that the possibility of tipping increases the incentive to cooperate and decreases the incentive to free ride, so that the size of the stable coalition becomes larger. It follows that partial cooperation often avoids tipping and the ensuing big loss of ecosystem services. Moreover, this paper shows that if it requires a large coalition to avoid tipping, and stability is lost because the incentive to free ride becomes too strong, the welfare loss of tipping is small. It follows that the usual grim story in the literature on two-stage membership games, namely that the size of the stable coalition is small and leaves large possible gains of cooperation, is reversed in the presence of tipping points.

The literature on managing systems in the presence of tipping points is rapidly increasing. Partly this literature focusses on models with concave-convex dynamics (e.g., Brock and Starrett, 2003, Mäler et al., 2003, Wagener, 2003, Crépin, 2007, Kossioris et al., 2008, Heijdra and Heijnen, 2013, Heijnen and Wagener, 2013). Another part of this literature considers hazard-rate models with an event probability of a structural change (e.g., Polasky et al., 2011, Lemoine and Traeger, 2014, Cai et al., 2015, van der Ploeg and de Zeeuw, 2018). This paper mainly builds on Mäler et al. (2003)

who compare cooperation and non-cooperation between the users of a lake. They show that two Nash equilibria exist, one close to the full-cooperative outcome and one where the lake has tipped. Welfare in the good Nash equilibrium is only slightly lower than in the full-cooperative outcome, but welfare drops considerably in the bad Nash equilibrium. It depends on the initial conditions of the lake which Nash equilibrium results. This paper focusses on the situation that the good Nash equilibrium fails to exist, because one user has an incentive to deviate despite the consequence of tipping. Barrett (2013) considers this possibility as well, but we ask the question whether in such a case stable partial cooperation can prevent tipping of the lake. Moreover, our model also allows analyzing the question whether stable partial cooperation can induce tipping back after tipping has occurred. This paper shows that it is generally harder to achieve this than to prevent tipping. This gives another reason to stay away from a tipping point, because it may not be possible to sustain a level of cooperation that is needed to tip back.

In order to clarify the mechanisms, the results are first derived in a relatively simple tipping-point model, with constant emissions and a piecewise linear response function with an upper and a lower threshold. This paper shows when it is first best to avoid tipping and when this can be sustained in a Nash equilibrium. In case it cannot be sustained in a Nash equilibrium, this paper shows when stable partial cooperation avoids tipping. When tipping has occurred but the system is physically reversible, the same results are presented in analyzing the issue of tipping back to the good state of the system. In the sequel, this paper analyzes the same questions in the well-known lake system, with concave-convex dynamics and time-dependent emissions. The lake system can be seen as a metaphor for many natural systems with tipping points. This analysis requires advanced numerical methods, but the pattern of the results is the same as for the relatively simple model.

The most interesting result is that tipping from the good to the bad conditions of the lake may be avoided by stable partial cooperation and if not, the remaining gains of cooperation are small. In case tipping has occurred earlier, tipping back from the bad to the good conditions of the lake may also be induced by stable partial cooperation, but this is generally harder to achieve. It follows that even if it is physically possible to tip back, it may often not be socially possible. This implies that the full socio-ecological system has more complicated properties than the underlying ecological system. In an ecological system with tipping points, it is costly to tip back, because the level of emissions has to be reduced substantially, but it may still be possible. In this case, the ecological system has hysteresis but is reversible. However, the socio-ecological system may be irreversible, because the level of cooperation that is needed to induce the system to tip back is not stable. For this level of cooperation, the incentive to free ride is too strong. It follows that in this case tipping becomes socially irreversible.

Section 2 analyses the relatively simple tipping game and presents the full-cooperative outcome, the Nash equilibria, and the results on coalition formation, with a short note on uncertainty about the location of the threshold. Section 3 analyses the lake game, and Section 4 concludes.

2. The tipping game

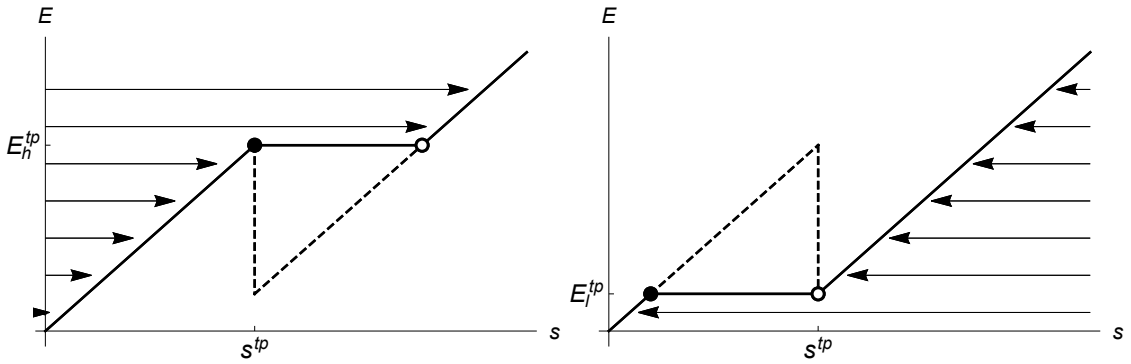
A well-known example of a model with a tipping point is the lake system. The dynamics of the lake is well described by a concave-convex response function to phosphorus loadings (e.g., Brock and Starrett, 2003, Mäler et al., 2003). In case of sufficient curvature of this response function, a threshold occurs so that increasing the phosphorus loadings above this threshold tips the lake into a domain of attraction with a different equilibrium, and with lower ecological services. It may be

possible to tip the system back at a low threshold to the original domain of attraction by decreasing the phosphorus loadings substantially, but the tip may also be irreversible.

The following piecewise linear model is a simple way of representing the characteristics of a model with a tipping point. Suppose that an ecological system is affected by the emissions $e_i, i = 1,2,...,n,$ of n economic agents. The dynamics of the stock of pollutants s (e.g., the stock of phosphorus in the water of the lake or the stock of greenhouse gases in the atmosphere) is given by

$$\begin{aligned} &\dot{s}(t) = E - f(s(t)), E = \sum_{i=1}^{n} e_i, s(0) = s_0, \\ &f(s) = s, 0 \le s \le s^{tp}, \\ &f(s) = s - b, s > s^{tp}, \end{aligned} \tag{1}$$

where s^{tp} denotes the level of the stock where tipping occurs, E the total level of emissions, and b the shift in the equilibrium at a tipping point. Figure 1a shows the dynamics of the system for the initial stock $s_0 = 0$ and for different fixed levels of total emissions E. Figure 1b shows the same for a high initial stock s_0. We fix the high tipping point (E_h^{tp}, s^{tp}) at $(1,1)$ and the low tipping point (E_l^{tp}, s^{tp}) at $(1-b,1)$. Note that for $b > 1$, tipping becomes irreversible, because emissions cannot be negative. A small increase in total emissions above $E_h^{tp} = 1$ shifts the equilibrium from $s^{tp} = 1$ to just above $s = 1 + b$, and the higher stock s implies a big loss in ecosystem services. Furthermore, a substantial decrease of total emissions below $E_l^{tp} = 1 - b$ is needed to shift the equilibrium back from $s^{tp} = 1$ to $s = 1 - b$ (if this is possible, i.e. if $b \le 1$).



(a): Adverse tipping (b): Favourable tipping

Figure 1. Tipping dynamics

Emissions result from economic activities that yield benefits. For example, phosphorus loadings result from agricultural activities, and greenhouse gas emissions result from the use of cheap fossil fuels. On the other hand, the stock of pollutants represents costs in the form of a loss of ecosystem services. Welfare indicators for this trade-off are given by

$$\ln e_i - cs^2, i = 1,2,...,n, \tag{2}$$

where c is a preference parameter that weighs the benefits and the costs. The logarithmic utility is convenient in the analysis below, because it implies that collective optimization is independent of the number of economic agents (e.g., Brock and Starrett, 2003, Mäler et al., 2003). By assuming fixed levels of total emissions E, we can focus the analysis on the steady states that can arise as shown by Figure 1. We will relax this assumption in Section 3 where we analyze the lake system. The question is how the presence of a tipping point affects the cooperative and non-cooperative outcomes.

2.1 The full-cooperative outcome

When the *n* economic agents (such as users of the ecological services of the lake or countries in a negotiation on climate change) cooperate, they maximize the sum of the welfare indicators given in (2). The Lagrangian becomes

$$L = \sum_{i=1}^{n} \ln e_i - ncs^2 + \lambda(E - f(s)). \tag{3}$$

Since the slope of $f(s)$ is 1, the first-order condition yields

$$E = \sum_{i=1}^{n} e_i = -\frac{n}{\lambda} = \frac{nf'(s)}{2ncs} = \frac{1}{2cs}. \tag{4}$$

The first-order condition is a hyperbola in Figure 1a that moves up if *c* decreases. It passes through the high tipping point $(E_h^{tp}, s^{tp}) = (1,1)$ for $c = 0.5$. For $c > 0.5$, it is optimal to stay below the tipping point. However, for $c < 0.5$ two possibilities arise. It is either optimal to stay at the tipping point or to tip and to optimize in the other domain of attraction. It is clear that a value of *c* exists such that the welfare in the tipping point is the same as the welfare for this value of *c* in the optimal point in the other domain of attraction. The tipping point lies on the iso-welfare curve *w* that is tangent to the line $f(s) = s - b$ in this optimal point: see Figure 2. If we denote this value of *c* as $\hat{c}$, for $c < \hat{c}$ it is optimal to let tipping occur, because the costs of tipping are very low. Hence, the tipping point is optimal for $\hat{c} \le c \le 0.5$.

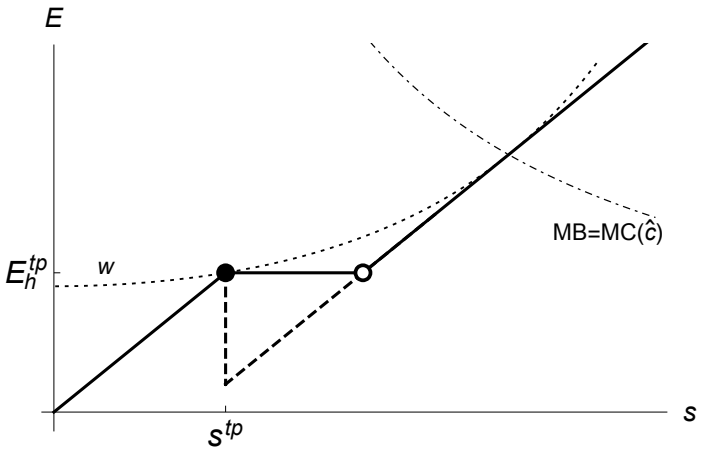


Figure 2. Indifference between tipping and non-tipping if $c = \hat{c}$.

This critical level $\hat{c}$ for the preference parameter c, above which it is optimal to stay at or below the tipping point (E_h^{tp}, s^{tp}), depends on the physical parameter b that indicates the size of the shift in the equilibrium stock s. The first-order condition in the other domain of attraction yields

$$E = \frac{1}{2cs} = s - b \Rightarrow c = \frac{1}{2s(s-b)}. \tag{5}$$

The welfare levels in this point and in the tipping point have to be equal, implying

$$\ln\frac{1}{n} - \frac{1}{2s(s-b)} = \ln\frac{s-b}{n} - \frac{s}{2(s-b)}. \tag{6}$$

From (5) and (6) it follows that the critical level $\hat{c}(b)$ is given by

$$\hat{c}(b) = \frac{1}{2\hat{s}(\hat{s}-b)}, 2\hat{s}(\hat{s}-b)\ln(\hat{s}-b) = \hat{s}^2 - 1. \tag{7}$$

As an example, we take $b = 0.9$. It follows that $\hat{s} = 2.8018$, so that $\hat{c} = 0.0938$. This implies that for $0.0938 \le c \le 0.5$ the tipping point is the optimal solution, and for $c > 0.5$ it is optimal to choose E below the tipping point. The properties of the function $\hat{c}(b)$ from (7) are given in the following proposition: this is a special case of Proposition 2 in the next section, whose proof is given in Appendix B.

Proposition 1. The critical level $\hat{c}(b)$ is continuously differentiable, and satisfies $\hat{c}(0)=0.5$, $\hat{c}(b)<1/(2(b+1))$ for $b>0$, $\hat{c}'(b)<0$ for $b>0$, $\lim_{b\to\infty}\hat{c}(b)=0$ and $\lim_{b\downarrow 0}\hat{c}'(b)=-\infty$.

2.2 Nash equilibria

In a non-cooperative Nash equilibrium, each economic agent *i* maximizes the welfare indicator given in (2). The Lagrangians become

$$L_i=\ln e_i-cs^2+\lambda_i(E-f(s)),i=1,2,...,n. \tag{8}$$

Since the slope of $f(s)$ is 1, the first-order conditions determining the candidate symmetric Nash equilibria yield

$$E=\sum_{i=1}^{n}e_i=-\frac{n}{\lambda_i}=\frac{nf'(s)}{2cs}=\frac{n}{2cs}. \tag{9}$$

This is again a hyperbola in Figure 1a, that moves up if *c* decreases. It passes through the high tipping point $(E_h^{tp},s^{tp})=(1,1)$ for $c=0.5n$. It is clear that for $c>0.5n$, the intersection point below the tipping point is the only Nash equilibrium. However, for $c<0.5n$ it is possible that two Nash equilibria exist, namely the tipping point and the intersection point of the hyperbola and the line $f(s)=s-b.$ The tipping point is indeed a Nash equilibrium if no economic agent has an incentive to deviate. If an economic agent deviates, the system tips. The question is therefore whether a deviating economic agent can achieve higher welfare in the other domain of attraction. This is the same type of problem as in the previous section. If the value of *c* is sufficiently low such that it is better for this economic agent to let tipping occur, the tipping point is not a Nash equilibrium. The first-order condition for the best response of this economic agent yields

$$E=\frac{1}{2cs}+\frac{n-1}{n}E_h^{tp}=\frac{1}{2cs}+\frac{n-1}{n}. \tag{10}$$

It is again clear that a value $\bar{c}$ of c exists such that for this value, the welfare in the tipping point is the same as in the optimal point in the other domain of attraction. Figure 2a applies again, but now specifically to the welfare of a deviator. For $c < \bar{c}$, the best response is to let tipping occur. It follows that only for $\bar{c} \le c \le 0.5n$ the tipping point is a Nash equilibrium. The critical level $\bar{c}$ for the preference parameter c, above which the tipping point (E_h^{tp}, s^{tp}) becomes a Nash equilibrium, depends again on the parameter b but also on the number of economic agents n. The first-order condition in the other domain of attraction yields

$$E = \frac{1}{2cs} + \frac{n-1}{n} = s - b \Rightarrow c = \frac{1}{2s(s-b-(n-1)/n)}. \tag{11}$$

The welfare levels for the deviator in this point and in the tipping point have to be equal, implying

$$\ln\frac{1}{n} - c = \ln\frac{1}{2cs} - cs^2. \tag{12}$$

From (11) and (12) it follows that the critical level $\bar{c}(b,n)$ is given by

$$\bar{c}(b,n) = \frac{1}{2\bar{s}\bar{x}}, \bar{x} = \bar{s} - b - \frac{n-1}{n}, 2\bar{s}\bar{x}\ln(n\bar{x}) = \bar{s}^2 - 1. \tag{13}$$

As an example, we take again $b = 0.9$, and $n = 10$. It follows that $\bar{s} = 2.3616$, so that $\bar{c} = 0.377$. This implies that for $0.377 \le c \le 5$ the tipping point is a Nash equilibrium. The critical level lies below 0.5, so that for $0.377 \le c \le 0.5$ the tipping point is optimal and also a Nash equilibrium (see Section 2.1). For $0.0938 \le c < 0.377$, the tipping point is optimal but not a Nash equilibrium. For $0.5 \le c \le 5$ the tipping point is a Nash equilibrium, but it is optimal to choose E below the tipping point. Note that for $b < 0.65413$ the critical level $\bar{c} > 0.5$, so that the areas where the tipping point is optimal and where the tipping point is a Nash equilibrium do not overlap. In this case, when the tipping point is a Nash equilibrium, it is optimal to choose E below the tipping point. The general properties of the function $\bar{c}(b,n)$ from (13) are given in the following

proposition (with a proof in Appendix B). Note that Proposition 2 implies Proposition 1, because $\hat{c}(b) = \bar{c}(b,1)$.

Proposition 2. The critical level $\bar{c}(b,n)$ is continuously differentiable, and satisfies $\bar{c}(0,n) = n/2$, $\bar{c}(b,n) < n/(2(b+1))$ for $b > 0$, $\partial \bar{c}(b,n)/\partial b < 0$, $\partial \bar{c}(b,n)/\partial n > 0$ for $b > 0$, $\lim_{b\to\infty} \bar{c}(b,n) = 0$ and $\lim_{b\downarrow 0} \partial \bar{c}(b,n)/\partial b = -\infty$.

If the size *b* of the shift in the stock at the tip is large enough, so that a range in *c* exists where the tipping point is both optimal and a Nash equilibrium, we can formulate the findings as follows.

Proposition 3. Depending on the parameter *b* that indicates the size of the shift in the equilibrium stock *s* in case of tipping, and on the number of economic agents *n*, critical values $\hat{c}(b)$ and $\bar{c}(b,n)$ exist for the preference parameter *c* such that in case $\bar{c}(b,n) \le 0.5$:

- for $\hat{c}(b) \le c < \bar{c}(b,n)$ the high tipping point (E_h^{tp}, s^{tp}) is optimal but not a Nash equilibrium,
- for $\bar{c}(b,n) \le c \le 0.5$ the high tipping point (E_h^{tp}, s^{tp}) is optimal and also a Nash equilibrium,
- for $0.5 < c \le 0.5n$ the high tipping point (E_h^{tp}, s^{tp}) is a Nash equilibrium but not optimal.

Barrett (2013), who also considers a model that features a tipping point, similarly finds a range of parameter values where the tipping point is optimal. This range can be split in a range where the tipping point is also a Nash equilibrium and a smaller range where it is not. He concludes that often tipping can be prevented by coordination on the appropriate Nash equilibrium, and that sometimes cooperation is needed. His model is different from ours in several aspects. Countries decide on abatement or emission reduction. The important parameters are the level of total abatement that

prevents tipping and a fixed loss when tipping occurs. In our model, the economic agents decide on emissions. They still optimize when tipping occurs. The costs of tipping have two parameters: a system parameter that indicates the size of the shift and a preference parameter that weighs costs and benefits. In our model, we have a range of parameter values where tipping is prevented in a Nash equilibrium but where the tipping point is not optimal. Our model can also consider tipping back to the original domain of attraction, as we will analyze below. In Section 3, we extend our model to the lake system, the metaphor for ecological systems with tipping points.

The central contribution of our paper is to investigate the option of stable partial cooperation in the presence of a tipping point. Barrett (2013) does not consider this, but we show in Appendix A how it works out in his model. Using a two-stage membership game, we show there that the size of the stable coalition is larger than in the absence of a tipping point. Furthermore, in a large part of the range of parameter values where cooperation is needed, stable partial cooperation prevents tipping. Only when the loss of tipping is small, the cooperation that is needed is not stable anymore. In the next section, we will develop and generalize this result for our model.

2.3 Coalition formation

Proposition 3 states that for $\bar{c}(b,n) \le c \le 0.5$, the high tipping point (E_h^{tp}, s^{tp}) is optimal and also a Nash equilibrium, but for $\hat{c}(b) \le c < \bar{c}(b,n)$ it is not a Nash equilibrium anymore, although it is still optimal. The question is whether stable partial cooperation can support the tipping point in this last range. As usual in the theory on international environmental agreements, we employ the two-stage membership game (Hoel, 1992, Carraro and Siniscalco, 1993, Barrett, 1994). In the first stage, the economic agents decide whether or not to become a signatory to the agreement. In the second stage, the coalition and the individual outsiders choose the emission levels. Suppose that a

coalition of size $1<k<n$ forms in the first stage. The welfare indicators in the second stage become

$$\sum_{j=1}^{k}\ln e_j - kcs^2, \ln e_i - cs^2, i=k+1,...,n, \tag{14}$$

and the Lagrangians become

$$\begin{aligned} L &= \sum_{j=1}^{k}\ln e_j - kcs^2 + \lambda(E - f(s)). \\ L_i &= \ln e_i - cs^2 + \lambda_i(E - f(s)), i=k+1,...,n. \end{aligned} \tag{15}$$

Since the slope of $f(s)$ is 1, the first-order conditions determining the candidate Nash equilibria between the coalition of size k and the $n-k$ individual outsiders yield

$$E = \sum_{i=1}^{n} e_i = -\frac{k}{\lambda} - \frac{n-k}{\lambda_i} = \frac{kf'(s)}{2kcs} + \frac{(n-k)f'(s)}{2cs} = \frac{n-k+1}{2cs}. \tag{16}$$

Condition (16) is the same as condition (9), but for $n-k+1$ economic agents. For these welfare indicators, the coalition effectively operates as one individual economic agent. This is convenient for the analysis but not essential for the outcomes. It follows that increasing the size of the coalition implies decreasing the number of economic agents in the second stage. This means that the range $\hat{c}(b) \le c < \overline{c}(b,n)$ where the high tipping point (E_h^{tp}, s^{tp}) is optimal but not a Nash equilibrium can be covered by partial cooperation. Using (13) in the previous section, it follows immediately that the tipping point is a Nash equilibrium between a coalition of size k and $n-k$ individual outsiders in the range $\overline{c}(b,n-k+1) \le c \le 0.5$, where $\overline{c}(b,n-k+1)$ is given by

$$\overline{c}(b,n-k+1) = \frac{1}{2\overline{sx}}, \overline{x} = \overline{s} - b - \frac{n-k}{n-k+1}, 2\overline{sx}\ln[(n-k+1)\overline{x}] = \overline{s}^2 - 1. \tag{17}$$

By increasing the size of the coalition k, the critical value $\overline{c}(b,n-k+1)$ decreases step by step. For the last step, in order to reach $\hat{c}(b)$, the grand coalition of *n* economic agents is needed.

Moving to the first stage of the two-stage game, the question is how many economic agents will join the coalition. The subgame-perfect Nash equilibrium requires at this stage that a member of the coalition does not want to leave the coalition, and that an outsider does not want to join the coalition. This depends, of course, on the outcome in the second stage. The subgame-perfect Nash equilibrium yields the stable coalition. We only have to check the *internal* stability of a coalition of size k, that is, we have to check if the welfare of a member of the coalition of size k is at least as large as the welfare of an outsider to the coalition of size $k-1$. This may especially be the case if the coalition of size $k-1$ cannot prevent tipping. Therefore, suppose that c lies in the range

$$\bar{c}(b, n-k+1) \le c < \bar{c}(b, n-k+2). \tag{18}$$

This means that the tipping point is a Nash equilibrium in case the coalition has size k, but it is not a Nash equilibrium in case the coalition has size $k-1$. It follows that tipping will occur when a coalition member leaves the coalition, so that in the second stage the outcome will be as a Nash equilibrium for $n-k+2$ economic agents in the other domain of attraction. Thus, the condition for internal stability of a coalition of size k becomes

$$\ln \frac{1}{k(n-k+1)} - c \ge \ln \frac{1}{2cs} - cs^2, \frac{n-k+2}{2cs} = s - b. \tag{19}$$

As an example, we again take $b = 0.9$ and $n = 10$. Suppose that the size of the coalition is $k = 8$. In order to find the range (18), we solve (17) for $k = 8$ and $k = 7$. This yields $\bar{c}(0.9,3) = 0.19662$ and $\bar{c}(0.9,4) = 0.23311$, respectively. Now we can check the condition for internal stability (19). For any $0.19662 \le c < 0.23312$, the equality in (19) yields s and the inequality can be checked. In this case, condition (19) does not hold. Suppose that the size of the coalition is $k = 7$. In order to find the range (18), we solve (17) for $k = 6$. This yields $\bar{c}(0.9,5) = 0.26428$. Now we can check the condition for internal stability (19). For any $0.23312 \le c < 0.26429$, the condition (19) holds.

It follows that the coalition of size $k = 7$ prevents tipping and is stable in the corresponding range for c as given by (18), but the coalition of size $k = 8$ is not stable. In the same way, it is easy to show that the coalitions of size $k = 2,...,6$ prevent tipping and are stable in the corresponding range for c as given by (18), but the coalitions of size $k = 9,10$ are not stable.

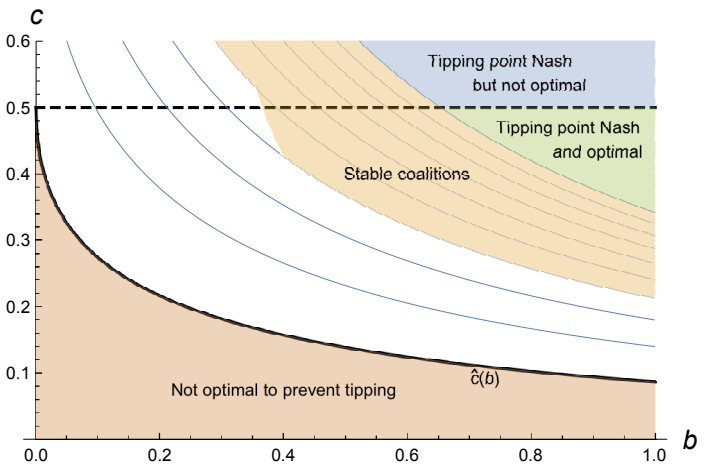


Figure 3. Stable non-tipping coalitions.

It is straightforward to extend the analysis to smaller b for which the area where the tipping point is both optimal and a Nash equilibrium is empty. In Figure 3 we present the results for $0 \le b \le 1$, $0 \le c \le 0.6$, and $n = 10$. For $c > 0.5$ it is optimal to stay below the tipping point (see Section 2.1). The curve $\hat{c}(b)$ indicates the level of c below which it is not optimal to prevent tipping. The area where the high tipping point (E_h^{tp}, s^{tp}) is a Nash equilibrium is split into an area where the tipping point is optimal and an area where it is not optimal. In between are the areas located where a Nash equilibrium between a coalition of size k and $n-k$ outsiders prevents tipping. Coalitions up to size $k = 6$ are stable, and the coalition of size $k = 7$ is stable if b is sufficiently small. In the white area, larger coalitions are needed to prevent tipping, but these coalitions are not stable. Figure 3 is similar to Figure 2 in Barrett (2013), but that figure only has an area with Nash equilibria and a wedge between that area and the equivalent of $\hat{c}(b)$. Barrett (2013) concludes that coordination

on a Nash equilibrium prevents tipping in that area, but that in the wedge cooperation is required. He uses this for the design of experimental work (e.g., Barrett and Dannenberg, 2012). We show that coordination on a Nash equilibrium between a stable coalition and outsiders prevents tipping in a large part of the wedge, as can be seen in Figure 3.

Note that the size of the stable coalition is much larger than what is usually found in the literature on international environmental agreements. The two-stage membership game actually balances the incentive to cooperate and the incentive to free ride. In the regular setting, the incentive to free ride apparently dominates. However, if tipping can occur, the incentive to free ride decreases, so that the size of the stable coalition becomes larger. Figure 3 shows that when the cost parameter c decreases, and the tipping point is not a Nash equilibrium anymore, it is better to form a coalition of size $k = 2$. This prevents tipping, and this coalition is also stable. When the cost parameter c decreases further, a larger coalition is needed to prevent tipping. This process continues until the coalition becomes large, so that the incentive to free ride becomes strong, and the coalition is not stable anymore. However, when stability is lost, the costs of tipping are low (c and b are small), so that the gains of cooperation are low. In general, we arrive at the following proposition (with a proof in Appendix B).

Proposition 4. The range $\hat{c}(b) \le c \le \overline{c}(b,n)$, where it is optimal to prevent tipping but not tipping is not a Nash equilibrium, can be split in the ranges $\overline{c}(b,n-k+1) \le c \le \overline{c}(b,n-k+2), k = 2,3,...,n$, where a Nash equilibrium between the coalition of size k and $n-k$ outsiders prevents tipping. Moreover, a size k^* exists, such that the coalitions of size $k \le k^*$ are stable. For large values of b, the size k^* is limited from above by the largest integer that satisfies $k - 1 + \ln k < n$. For $k > k^*$,

the coalition of size k is not stable, but in that case the costs of tipping are low, so that the gains of cooperation are low.

In the example for $n = 10$, the largest integer that satisfies $k - 1 + \ln k < 10$ is $k = 8$. Indeed, Figure 3 shows that in the range $0 \le b \le 1$, the size of the largest stable coalition is $k^* = 6$ or $k^* = 7$. For larger *b*, the size of the largest stable coalition can become $k^* = 8$, but a larger stable coalition is not possible. For $n = 100$, the largest integer that satisfies $k - 1 + \ln k < 100$ is $k = 96$. This shows that the largest stable coalition is close but not equal to the grand coalition. Our model also allows to investigate tipping back to the original domain of attraction. This is the topic of next section.

2.4 The inverse tipping game

Suppose that tipping has unfortunately occurred, but that tipping back to the original domain of attraction is possible, because $b \le 1$ (see Figure 1b). We assume that tipping back actually occurs when we reach the low tipping point $(E_l^{tp}, s^{tp}) = (1-b, 1)$, so that the point $(1-b, 1-b)$ results. The first-order condition given by (4) cuts the point $(1-b, 1-b)$ for $c = 1/(2(1-b)^2)$, so that it will be optimal to choose *E* below the low tipping point for $c > 1/(2(1-b)^2)$. The question is when it is optimal to induce tipping back and if so, when tipping back can be realized in a Nash equilibrium or by stable partial cooperation. We call this the inverse tipping game.

The analysis runs parallel to that given above. Because we assume that tipping back occurs in the low tipping point (E_l^{tp}, s^{tp}), the welfare becomes $\ln[(1-b)/n] - c(1-b)^2$. It follows that the critical level $\breve{c}$, above which it is optimal to move down to the low tipping point (E_l^{tp}, s^{tp}), is given by

$$\breve{c}(b) = \frac{1}{2\breve{s}(\breve{s}-b)}, 2\breve{s}(\breve{s}-b)\ln\frac{\breve{s}-b}{1-b} = \breve{s}^2 - (1-b)^2. \qquad (20)$$

As an example, we take $b = 0.9$. It follows that $\breve{s} = 1.3472$, so that $\breve{c} = 0.82985$. This implies that for $c \geq 0.82985$ tipping back is the optimal solution. This critical level is higher than the one we found in (7). Figure 2 applies again, but the iso-welfare curve *w* cuts the line $f(s) = s$ in the point $(1-b, 1-b)$ instead of in the point $(1,1)$, and is tangent to the line $f(s) = s - b$ for a lower *s*: see Figure 4. For *c* in between these critical levels, it is optimal to prevent tipping, but it is not optimal to induce tipping back. It follows that it is generally harder to get out of the bad domain of attraction than to avoid getting in there.

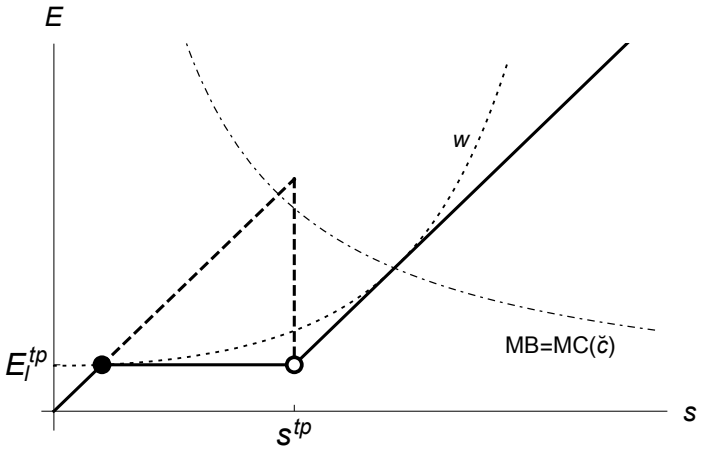


Figure 4. Indifference between tipping and non-tipping at $c = \breve{c}$.

The critical level $\tilde{c}(b,n)$, above which it is a Nash equilibrium to induce tipping back at the low tipping point (E_l^{tp}, s^{tp}), is given by

$$\tilde{c}(b,n) = \frac{1}{2\tilde{s}\tilde{x}}, \tilde{x} = \tilde{s} - b - \frac{n-1}{n}(1-b), 2\tilde{s}\tilde{x}\ln\frac{n\tilde{x}}{1-b} = \tilde{s}^2 - (1-b)^2. \qquad (21)$$

As an example, we take again $b = 0.9$, and $n = 10$. It follows that $\tilde{s} = 1.1876$ so that $\tilde{c} = 2.1306$. This implies that for $c \geq 2.1306$ it is optimal to tip back and this is also a Nash equilibrium, and for $0.82985 \leq c < 2.1306$ it is optimal to tip back but this is not a Nash equilibrium.

In the same way as in the previous section, the solution of (21) for $n-k+1$ economic agents yields the critical level $\tilde{c}(b,n-k+1)$ for the existence of a Nash equilibrium between a coalition of size k and $n-k$ outsiders. Furthermore, the condition for internal stability becomes

$$\ln\frac{1-b}{k(n-k+1)}-c(1-b)^2 \geq \ln\frac{1}{2cs}-cs^2, \frac{n-k+2}{2cs}=s-b. \qquad (22)$$

Suppose that the size of the coalition is $k=9$. Solving (21) for 2 and for 3 economic agents yields $\tilde{c}(0.9,2)=1.1590$ and $\tilde{c}(0.9,3)=1.3786$, respectively. For any $1.1590 \leq c < 1.3786$, the condition for internal stability (22) does not hold, so that the coalition of size $k=9$ is not stable. Suppose that the size of the coalition is $k=8$. Solving (21) for 4 economic agents yields $\tilde{c}(0.9,4)=1.5456$. For any $1.3786 \leq c < 1.5456$, the condition for internal stability (22) holds now. It follows that the coalition of size $k=8$ induces tipping back and is stable. In the same way, it is easy to show that the coalitions of size $k=2,...,7$ induce tipping back and are stable in the corresponding range for c, but the coalition of size $k=10$ is not stable.

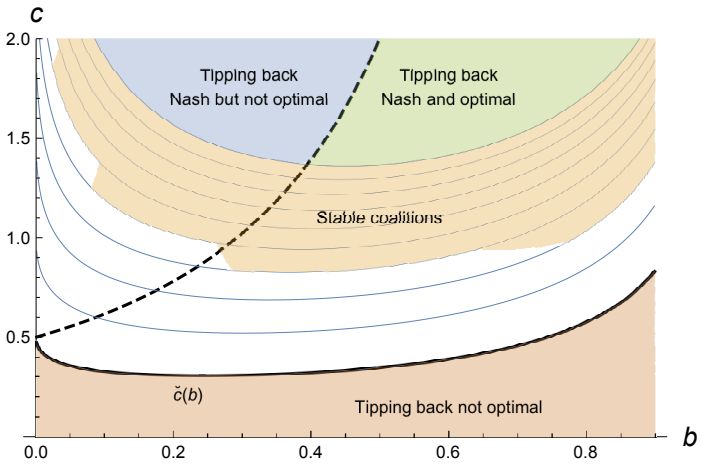


Figure 5. Stable tipping coalitions

In Figure 5, we present the results for $0 \leq b \leq 0.9$, $0 \leq c \leq 2$ and $n=10$. Above the dashed curve, it is optimal to move below the tipping point. The curve $\check{c}(b)$ indicates the level of c below which

it is not optimal to induce tipping back. The area where tipping back at the tipping point (E_l^{tp}, s^{tp}) is a Nash equilibrium is split into an area where this is also optimal and an area where it is optimal to move below the low tipping point. In between are the areas located where tipping back is a Nash equilibrium between a coalition of size k and $n-k$ outsiders. In the white area, larger coalitions are needed to induce tipping back, but these coalitions are not stable.

The results are similar to the results for the tipping game in Section 2.3. The main difference is that it generally requires a higher cost parameter *c* to make it beneficial to tip back. However, it turns out that the stable coalition can be a bit larger. We have assumed that the system is physically reversible, and we find that it is optimal to induce the system to tip back in the area above the line $\breve{c}(b)$. However, we also find that at some point, the level of cooperation that is required to achieve this may not be stable anymore. This means that in such a case, tipping is physically reversible but socially irreversible, because the level of cooperation that is needed breaks down. In general, we arrive at the following proposition. Its proof, given in in Appendix B, relies on showing that the tipping game and the inverse tipping game are equivalent.

Proposition 5. The range $\breve{c}(b) \le c \le \tilde{c}(b,n)$, where it is optimal to induce tipping back but this is not a Nash equilibrium, can be split in the ranges $\tilde{c}(b,n-k+1) \le c \le \tilde{c}(b,n-k+2), k = 2,3,...,n,$ where a Nash equilibrium between a coalition of size k and $n-k$ outsiders induces tipping back. Moreover, a size k^* exists, such that the coalitions of size $k \le k^*$ are stable. The size k^* is limited from above by the largest integer that satisfies $k - 1 + \ln k < n$. For $k > k^*,$ the coalition of size k is not stable, but in that case the costs of not tipping back are low, so that the gains of cooperation are low. However, the original tipping is socially irreversible in that case.

2.5 Uncertainty

In the previous sections, we have assumed that the location of the high tipping point (E_h^{tp}, s^{tp}) is known. This is a reasonable assumption for the lake system, because it has been studied extensively (Carpenter, 2003), but in other cases the threshold is quite uncertain. For example, Rockström et al. (2009) indicate nine so-called planetary boundaries, such as the climate tipping point, but each of those have a zone of uncertainty. This implies that we have to consider that the tipping point is located between a lower bound s_1^{tp} and an upper bound s_2^{tp}, where this zone of uncertainty has a probability distribution. As the lower bound, we take $s_1^{tp} = 1$. For a uniform distribution function, the uncertainty changes the response function into a certainty-equivalent, continuous, piecewise linear function, given by

$$\begin{aligned} &f(s) = s, 0 \le s \le 1, \\ &f(s) = \frac{(s_2^{tp} - 1)s + b}{s_2^{tp} - 1 + b}, 1 < s \le s_2^{tp} + b, \\ &f(s) = s - b, s > s_2^{tp} + b, \end{aligned} \tag{23}$$

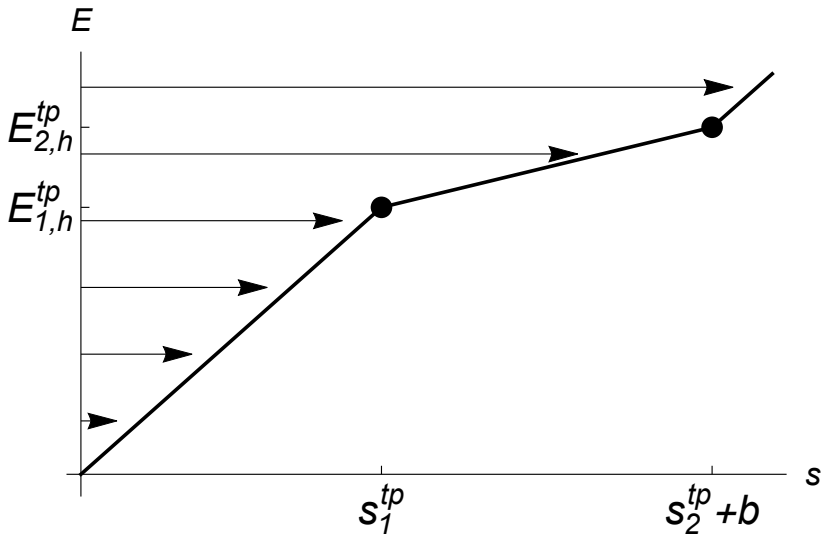


Figure 6. Certainty-equivalent response function.

Figure 6 depicts the new situation. The analysis in Sections 2.1 - 2.2 remains the same if the iso-welfare curves through the point $(1,1)$ that are tangent to the line $f(s) = s - b$ lie above the middle segment of (23). Comparing the slope of the lowest iso-welfare curve *w* in the point $(1,1)$ in Section 2.2 with the slope of the middle segment of (23) yields an upper limit for s_2^{tp}. This implies that if the zone of uncertainty is sufficiently small, so that s_2^{tp} does not exceed this upper limit for *b* and $\overline{c}(b,n)$, the point $(1,1)$ is still optimal and also a Nash equilibrium for $\overline{c}(b,n) \le c < 0.5$, according to Proposition 3. The analysis becomes more complicated in case the zone of uncertainty is larger or if a different probability distribution is considered. However, we expect our main point to hold up: even in the presence of some uncertainty, whatever its precise nature, it is possible to coordinate on a Nash equilibrium that is close to the optimal outcome. We leave a full analysis of the modalities under which this coordination breaks down to a future paper. Instead, in the next section we show that our results carry over in a straightforward manner to a more realistic model that shares basic characteristics with our tipping game model.

3. The lake game

In the previous sections, we assumed constant emissions $e_i, i = 1,2,...,n$, and could therefore focus on a steady-state analysis. This was the simplest representation of the typical tipping-point model. In the sequel, we will extend the analysis to a real and fully dynamical model. We will analyze the well-known lake model (Carpenter et al., 1999, Brock and Starrett, 2003, Mäler et al., 2003, and Wagener, 2003) which can be seen as the metaphor for ecological systems with tipping points. The response function in the lake model is not a piecewise linear function with a shift at the threshold but a concave-convex function. The dynamics of the accumulated stock of phosphorus *s* in the water of the lake is given by

$$\dot{s}(t) = E(t) - f(s(t)), E = \sum_{i=1}^{n} e_i, s(0) = s_0,$$
$$f(s) = bs - \frac{s^2}{1+s^2}, \tag{24}$$

where E denotes the total level of phosphorus loadings on the lake, and b a parameter that differs across lakes. The non-linear term in (24) is called a Holling type III functional response, and it is typical for many ecological models with tipping points. In this case, it represents the release of phosphorus from the bottom of the lake in case of an increase in the stock s. It is easy to show that for $0.5 < b < 3\sqrt{3}/8$, the curve $f(s)$ has similar tipping points (at the local maximum and at the local minimum of f) as the simple model in Section 2. Note that in this model, tipping becomes physically irreversible for $b < 0.5$.

Phosphorus loadings result from agricultural activities that yield benefits. On the other hand, an increase in the stock of phosphorus causes a loss of ecosystem services, such as clean water, fish and leisure. Infinite-horizon discounted welfare indicators for this trade-off are given by

$$\int_0^{\infty} [\ln e_i(t) - cs^2(t)] e^{-rt} dt, i = 1, 2, ..., n, \tag{25}$$

where c is again a preference parameter that weighs the benefits and the costs, and r denotes the discount rate.

The problem (24) - (25) is an extension of the simple tipping-point problem (1) - (2). In the simple problem, the economic agents choose fixed emissions $e_i, i = 1, 2, ..., n$, but in this extended problem, the economic agents choose time paths for the phosphorus loadings $e_i(t), i = 1, 2, ..., n.$ The problem (24) - (25) is a so-called differential game (Basar and Olsder, 1982). The choice of these time paths means that we will search for so-called open-loop solutions. It is well known that different types of Nash equilibria arise depending on the available information and on commitment. If emissions

$e_i, i = 1,2,...,n$, are only a function of time and the initial stock s_0, the open-loop Nash equilibrium results (Mäler et al., 2003, for the lake game). However, if emissions $e_i, i = 1,2,...,n$, are a function of time and the observed current level of the stock s, the feedback Nash equilibrium results (see Kossioris et al., 2008, and Dockner and Wagener, 2014, for the lake game). In this paper, we add the two-stage membership game for coalition formation. As first step, we consider only one round of coalition formation in which the economic agents commit to a path of emissions, leading to the open-loop Nash equilibrium. The question is again how the presence of a tipping point affects the cooperative and non-cooperative outcomes.

3.1 The full-cooperative outcome and Nash equilibria

When the *n* users of the lake cooperate, they maximize the sum of the welfare indicators given in (25), subject to the dynamics of the system given in (24). The Hamiltonian becomes

$$H = \sum_{i=1}^{n} \ln e_i - ncs^2 + \lambda(E - f(s)). \tag{26}$$

which yields the necessary conditions

$$E = \sum_{i=1}^{n} e_i = -\frac{n}{\lambda}, \tag{27}$$

$$\dot{E}(t) = -[r + f'(s(t))]E(t) + 2cs(t)E^2(t). \tag{28}$$

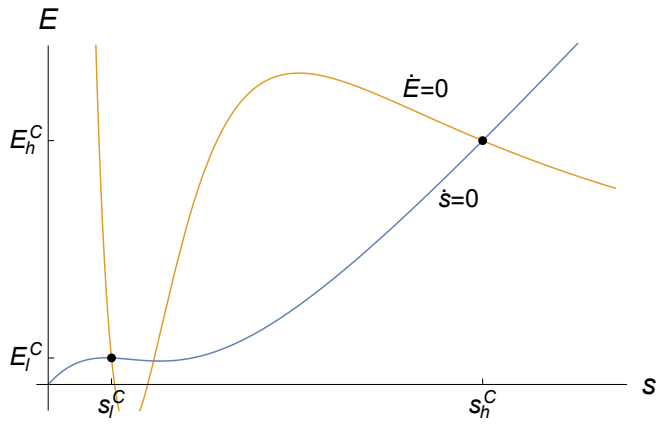


Figure 7. Phase diagram

Figure 7 shows the phase diagram of this modified Hamiltonian system (24) and (28). The isocline of (28) cuts the isocline of (24) in one or three steady states (Brock and Starrett, 2003). If there are three steady states, the middle one is unstable and the other two are saddle-point stable. The low steady state is indicated by (E_l^C, s_l^C), and the high steady state by (E_h^C, s_h^C). Wagener (2003) shows that a value $\hat{c}$ of c exists such that for $c > \hat{c}$, starting from a low initial stock s_0, it is optimal to let the system converge to the steady state with a low stock of phosphorus. For $c < \hat{c}$, it is optimal to let the system tip and converge to the steady state with the high stock of phosphorus. As in Section 2.1, this critical value $\hat{c}$ depends on the parameter b, but it also depends on the discount rate r and the initial stock s_0.

For a symmetric non-cooperative Nash equilibrium, the Hamiltonians become

$$H_i = \ln e_i - cs^2 + \lambda_i(E - f(s)), i = 1,2,...,n, \tag{29}$$

which yields the necessary conditions for the candidate Nash equilibria

$$E = \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \frac{1}{\lambda_i}, \tag{30}$$

$$\dot{E}(t) = -[r + f'(s(t))]E(t) + 2\frac{c}{n}s(t)E^2(t). \tag{31}$$

The phase diagram of this modified Hamiltonian system (24) and (31) is the same as the one for the full-cooperative outcome. The only difference is that the parameter c in (28) changes into c/n in (31). Mäler et al. (2003) show that for $b=0.6$, $c=1$, $n=2$ and $r=0.03$, the full-cooperative outcome has one steady state in the low-phosphorus area, whereas two Nash equilibria exist, one with a steady state in the low-phosphorus area and one with a steady state in the high- phosphorus area. The good Nash equilibrium is close to the full-cooperative outcome, so that the welfare loss is low. However, if the users of the lake cannot reach the "good" Nash equilibrium and are stuck in the "bad" Nash equilibrium (Grass et al., 2017), the welfare loss is substantial. With two users, only full cooperation will get the users out of the bad state. We extend this analysis by investigating the existence of Nash equilibria and the possibilities of stable partial cooperation in general.

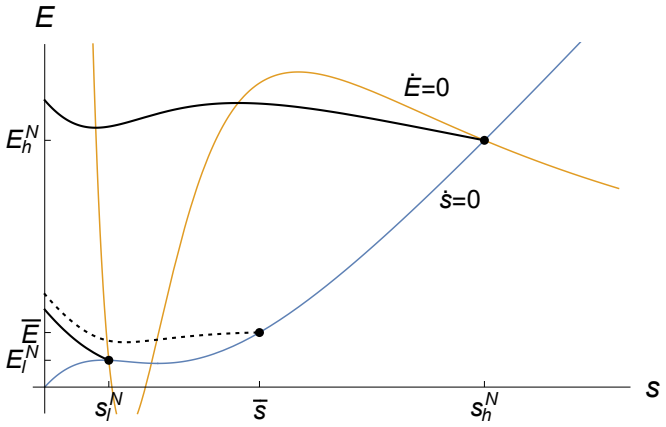


Figure 8. Trajectories: candidate Nash equilibria (solid), one-player deviation (dotted).

Analogous to the analysis in Section 2, we first check whether the candidate Nash equilibrium in the low-phosphorus area of the lake, with the steady state denoted by (E_l^N, s_l^N), is indeed a Nash equilibrium. We take as initial stock $s_0 = 0$. Figure 8 shows the trajectories of the candidate Nash equilibria to the low-phosphorus steady state (E_l^N, s_l^N) and to the high-phosphorus steady state (E_h^N, s_h^N). The necessary conditions for the best response of an individual user *i* become

$$\dot{e}_i(t) = -[r + f'(s(t))]e_i(t) + 2s(t)e_i^2(t), \tag{32}$$

together with (24), where the other users employ their Nash equilibrium strategies. The dashed curve in Figure 8 shows the resulting trajectory towards the steady state $(\bar{E},\bar{s})$. By comparing the welfare of individual user *i* on this trajectory with the welfare on the Nash equilibrium trajectory towards the steady state (E_l^N, s_l^N), it can be shown that a critical value $\bar{c}$ exists, so that for $c \geq \bar{c}$ there is no incentive to deviate, so that (E_l^N, s_l^N) is indeed the steady state of a Nash equilibrium. For *c* below this critical value, the only Nash equilibrium is the one with the trajectory leading to the high-phosphorus steady state (E_h^N, s_h^N). It should be noted that the critical value $\bar{c}$ depends on the parameters *b*, *n*, *r* and s_0, and that the calculations require advanced numerical methods, but the pattern of the results will be clear.

3.2 Coalition formation

Suppose that a coalition of size $1 < k < n$ forms in the first stage. The respective Hamiltonians for a non-cooperative Nash equilibrium between the coalition and the $n-k$ individual outsiders in the second stage become

$$\begin{aligned} H &= \sum_{j=1}^{k} \ln e_j - kcs^2 + \lambda(E - f(s)), \\ H_i &= \ln e_i - cs^2 + \lambda_i(E - f(s)), i = k+1,...,n. \end{aligned} \tag{33}$$

which yields the necessary conditions

$$\begin{aligned} E &= \sum_{i=1}^{n} e_i = -\frac{k}{\lambda} - \sum_{i=k+1}^{n} \frac{1}{\lambda_i}, \\ \dot{\lambda}(t) &= [r + f'(s(t))]\lambda(t) + 2kcs(t), \\ \dot{\lambda}_i(t) &= [r + f'(s(t))]\lambda_i(t) + 2cs(t), i = k+1,...,n. \end{aligned} \tag{34}$$

Since $\lambda = k\lambda_i$, this yields

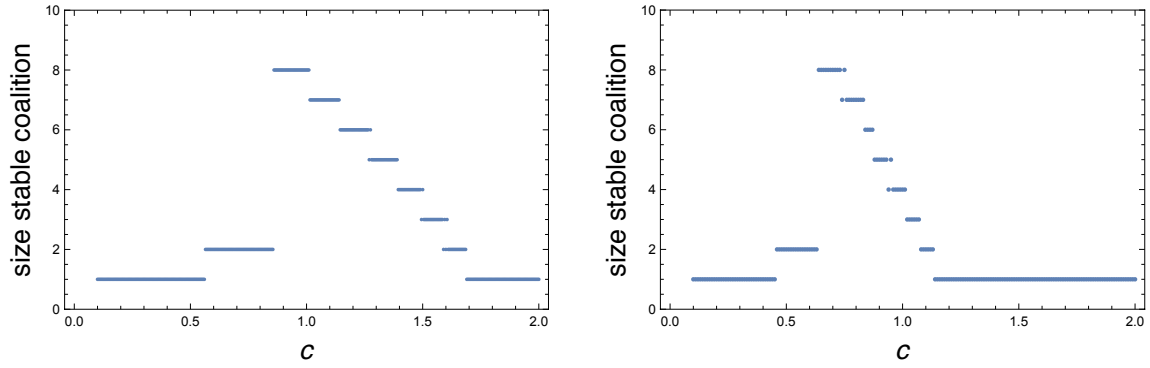
$$\dot{E}(t) = -[r + f'(s(t))]E(t) + 2\frac{c}{n-k+1}s(t)E^2(t). \tag{35}$$

Condition (35) is the same as condition (31), but for $n-k+1$ users of the lake. For these welfare indicators, the coalition effectively operates as one individual user.

The analysis requires a substantial numerical effort but it is basically the same as in Section 2. As an example, we take $b = 0.6$, $n = 2$, $r = 0.03$ and $s_0 = 0$. For $c \geq 0.4346$, it is optimal to converge to a low-phosphorus steady state (E_l^C, s_l^C). However, if the users of the lake do not cooperate, the low-phosphorus steady state (E_l^N, s_l^N) of this candidate Nash equilibrium proves not to be a Nash equilibrium for $c < 0.6648$. It follows that for $0.4346 \leq c < 0.6648$, the lake system will tip in the absence of cooperation, but it will remain in the low-phosphorus region when the users of the lake form a coalition. The welfare implications are substantial. For example, if $c = 0.7$, the users of the lake can coordinate on a Nash equilibrium with a low-phosphorus steady state (E_l^N, s_l^N), and the individual welfare becomes -100.3824, just below the optimal welfare -100.2386. In this case, the lake system does not tip, and the gains of cooperation are very small. However, if $c = 0.6$, the users of the lake end up in the Nash equilibrium with the high-phosphorus steady state (E_h^N, s_h^N), and the individual welfare drops to -120.1273, whereas the optimal welfare is -99.8047. In this case, the lake system tips in the absence of cooperation, and the gains of cooperation are large, i.e. almost 17%.

The question is whether stable partial cooperation can prevent tipping of the lake, and a large loss of welfare, in case the candidate Nash equilibrium with the low-phosphorus steady state (E_l^N, s_l^N) proves not to be a Nash equilibrium. The analysis is the same as in Section 2.3. A larger coalition effectively means a smaller number of users of the lake, so that tipping can be prevented for smaller values of *c*, until it is not optimal to prevent tipping anymore. However, an increasing size of the

coalition also increases the free-rider benefits, so that the coalition may lose stability at some point. The result is again that it is possible to prevent tipping of the lake for a large range of values of c, with a large stable coalition, but not for all values of c for which it is optimal to prevent tipping. If a very large coalition is needed, the coalition is not stable anymore.



(a) $b = 0.60$ (b) $b = 0.52$

Figure 9. Stable dynamic non-tipping coalitions.

Because it is very complicated to present the results for the lake game in the (b,c) parameter space as in Figure 4, we present the results only for $b = 0.6$ and $b = 0.52$. Figure 9a and Figure 9b show these results for $n = 10$, $r = 0.03$, $s_0 = 0$ and $0.01 \le c \le 2$. The pattern is the same as in Figure 3. In Figure 9a, for $c \ge 1.69$, coordination on the Nash equilibrium with the low-phosphorus steady state is feasible. For $1.58 \le c < 1.69$, a coalition of size $k = 2$ prevents tipping of the lake, and it is stable. Lowering the cost parameter c further, step by step a larger coalition is needed to prevent tipping of the lake, and these coalitions are stable up to $k = 8$. Lowering c further below 0.88, at first coalitions of size $k = 9$ and $k = 10$ are needed to prevent tipping, but these coalitions are not stable, and then it is not optimal anymore to prevent tipping of the lake. It follows that for $c < 0.88$ the old grim story in the literature on international environmental agreements reappears: the stable coalitions are very small. In this specific case, for $0.58 \le c < 0.88$ a stable coalition of size $k = 2$ results, but for $c < 0.58$ stable partial cooperation is not possible. The most important result is that

for $0.88 \leq c < 1.69$ stable partial cooperation prevents tipping of the lake. Furthermore, when this is not possible anymore, because the free-rider incentive becomes too strong, the cost parameter *c* is low, so that the costs of tipping are low. Figure 9b, with $b = 0.52$, has the same pattern again. The most important difference is that coordination on a Nash equilibrium to prevent tipping of the lake is sufficient now in a larger range of values of the cost parameter *c*. The reason is that a smaller value of *b* means that the shift in the stock *s* is larger and the incentive to deviate in the candidate Nash equilibrium is smaller.

3.3 The inverse lake game

Suppose that tipping has unfortunately occurred, but that tipping back to the original domain of attraction is possible, because $0.5 < b < 3\sqrt{3}/8$. We assume that tipping back actually occurs when we reach the low tipping point. We present the results for $b = 0.6$, with $n = 10$, $r = 0.03$, $s_0 = 1.75$ and $0.1 \leq c \leq 7$. Figure 10 shows these results. The pattern is the same as in Figures 9a/b. The main difference is that the value of the cost parameter *c*, below which partial cooperation is needed to induce tipping back, is much larger, i.e. $c < 6$. This indicates that the costs of the loss of ecosystem services have to be relatively high in order to be able to induce tipping back in a Nash equilibrium. Furthermore, the value for *c*, below which partial cooperation is not stable anymore, is larger as well, i.e. $c < 1.5$. The size of the largest stable coalition is now $k = 9$. If it is still optimal to induce tipping back, but a coalition of size $k = 10$ is needed in order to achieve this, this coalition is not stable. It follows that the lake system is physically reversible but socially irreversible in this case.

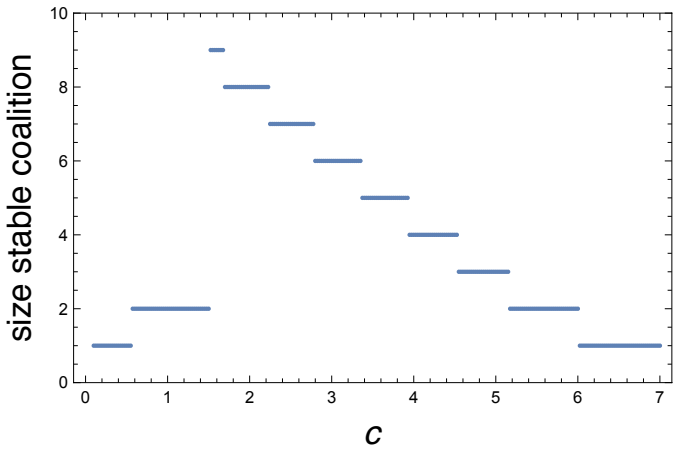


Figure 10. Stable dynamic tipping coalitions.

It is interesting to note that Figure 10 shows that for $b = 0.6$ and $c = 2$, for example, it requires a coalition of size $k = 8$ to get out of the high-phosphorus region but after tipping back to the low-phosphorus area, no coalition is needed to prevent tipping, as Figure 9 shows. This means that a substantial effort is needed to push the system from "brown" (green in case of the lake) to "green" (blue in case of the lake) environmental conditions, but then it is relatively easy to keep the system in a good state. This result is similar to the result in Acemoglu et al. (2012) where they show that only temporary policies are needed to redirect innovation toward clean inputs for production. A more sophisticated dynamical analysis to show this in our model is left for further research.

4. Conclusion

Tipping points are observed in a variety of natural systems. When a natural system tips, it shifts to another domain of attraction, which usually yields a substantial loss of ecosystem services. Tipping occurs when accumulated emissions from economic activities cross a threshold. Full cooperation of the economic agents keeps the natural system in a good condition, unless a low value is attached to the loss of ecosystem services. Non-cooperative behavior may avoid crossing the threshold as well, if the incentive to deviate is suppressed by the loss of ecosystem services. If it is not possible

to coordinate on a Nash equilibrium that avoids tipping but full cooperation would avoid tipping, the problem is that full cooperation may not be stable in the sense that the incentive to free ride is stronger than the incentive to cooperate. In such a case, the question is to what extent stable partial cooperation can solve the problem.

This paper first presents a relatively simple tipping-point model, with constant emissions and a piecewise linear response function with a shift at the threshold. In this model, it is relatively easy to derive the results. It is shown that stable partial cooperation indeed prevents tipping in a large range of parameter values and if this is not possible, the costs of tipping are low. It is also shown that in case tipping has occurred, stable partial cooperation can induce tipping back to the favorable conditions of the natural system, but again not in all cases where it is optimal to do so. This means that it may happen that tipping is physically reversible but socially irreversible, because the level of cooperation that is needed for tipping back may at some point not be stable anymore. Moreover, in general a larger level of cooperation is needed to induce tipping back than to prevent tipping.

This paper also analyses the same questions for the lake system, a well-known realistic model with a tipping point that can be seen as a metaphor for many natural systems with tipping points. The lake model has a concave-convex response function, and we allow time-dependent phosphorus loadings on the lake. The analysis uses advanced numerical methods, but the results are basically the same. An important policy conclusion is that when the users of the lake are trapped in a non-cooperative Nash equilibrium with a low level of ecosystem services, stable partial cooperation may get these users out of the pollution trap. Moreover, when the lake is in a good condition but it is not possible to coordinate on a Nash equilibrium that prevents the lake from tipping, stable partial cooperation may provide a solution. Again, a higher level of cooperation is needed to induce tipping back than to prevent tipping.

The usual grim story in the literature on international environmental agreements is that the size of the stable coalition is very small, especially when the possible gains of cooperation are large. In the presence of a tipping point, this story is reversed. This paper shows that the size of the stable coalition can be large, in order to prevent the large loss of tipping or to induce the large gain of tipping back. Moreover, when the size of the coalition cannot be increased anymore, because the incentive to free ride becomes too strong, the cost of losing ecosystem services is relatively low, so that there is not so much to gain from cooperation anyway.

In relation with environmental degradation, two wicked problems stand out: tipping points and the tragedy of the commons. In essence, the tragedy of the commons means that the incentive to free ride undermines the cooperation in managing the common good. An interesting conclusion of this paper is that the presence of a tipping point actually helps in sustaining cooperation, up to a certain point. One wicked problem is partly solved by the existence of another wicked problem.

References

Acemoglu, D., Ph. Aghion, L. Bursztyn and D. Hemous. 2012. The environment and directed technical change. *American Economic Review* 102, 1, 131-166.

d'Aspremont, C., A. Jacquemin, J. Gabszewicz and J. Weymark. 1983. On the stability of collusive price leadership. *Canadian Journal of Economics* 16, 1, 17-25.

Barrett, S. 1994. Self-enforcing international environmental agreements. *Oxford Economic Papers* 46, 878-894.

Barrett, S. 2013. Climate treaties and approaching catastrophes. *Journal of Environmental Economics and Management* 66, 2, 235-250.

Basar, T. and G.-J. Olsder. 1982. *Dynamic Noncooperative Game Theory*. Academic Press, New York.

Biggs, R., Th. Blenckner, C. Folke, L.J. Gordon, A.V. Norström, M. Nyström and G.J. Peterson. 2012. Regime shifts. In A. Hastings and L. Gross, *Encyclopedia in Theoretical Ecology*, 609-616. University of California Press, Berkeley and Los Angeles.

Brock, W.A. and D. Starrett. 2003. Managing systems with non-convex positive feedback. *Environmental & Resource Economics* 26, 4, 575-602.

Cai, Y. and T.S, Lontzek. 2019. The social cost of carbon with economic and climate risks. *Journal of Political Economy* 127, 6, 2684-2734.

Carpenter, S.R. and K.L. Cottingham. 1997. Resilience and restoration of lakes. *Conservation Ecology* 1, 2.

Carpenter, S.R., D. Ludwig and W.A. Brock. 1999. Management of eutrophication for lakes subject to potentially irreversible change. *Ecological Applications* 9, 3, 751-771.

Carraro, C. and D. Siniscalco. 1993. Strategies for the international protection of the environment. *Journal of Public Economics* 52, 3, 309-328.

Crépin, A.-S. 2007. Using fast and slow processes to manage resources with thresholds. *Environmental & Resource Economics* 36, 2, 191-213.

Diekert, F.K. 2017. Threatening thresholds? The effect of disastrous regime shifts on the non-cooperative use of environmental goods and services. *Journal of Public Economics* 147, 30-49.

Dockner, E. and F.O.O. Wagener. 2014. Markov perfect Nash equilibria in models with a single capital stock. *Economic Theory* 56, 3, 585-625.

Finus, M. 2003. Stability and design of international environmental agreements: the case of transboundary pollution. In H. Folmer and T. Tietenberg, *The International Yearbook of Environmental and Resource Economics* 2003/2004, 82-158. Edward Elgar, Cheltenham.

Grass, D., A. Xepapadeas and A. de Zeeuw. 2017. Optimal management of ecosystem services with pollution traps: the lake model revisited. *Journal of the Association of Environmental and Resource Economists* 4, 4, 1121-1154.

Heijdra, B. and P. Heijnen. 2013. Environmental abatement and the macroeconomy in the presence of ecological thresholds. *Environmental & Resource Economics* 55, 1, 47-70.

Heijnen, P. and F.O.O. Wagener. 2013. Avoiding an ecological regime shift is sound economic policy. *Journal of Economic Dynamics and Control* 37, 7, 1322-1341.

Hoel, M. 1992. International environmental conventions: the case of uniform reduction of emissions. *Environmental & Resource Economics* 2, 2, 141-159.

Hughes, T., A. Baird, D. Bellwood, M. Card, S. Connolly, C. Folke, R. Grosberg, O. Hoegh-Guldberg, J. Jackson, J. Kleypas, J. Lough, P. Marshall, M. Nyström, S. Palumbi, J. Panolfi, B. Rosen and J. Roughgarden. 2003. Climate change, human impacts, and the resilience of coral reefs. *Science* 301, 929-933.

Kossioris, G., M. Plexousakis, A. Xepapadeas, A. de Zeeuw and K.-G. Mäler. 2008. Feedback Nash equilibria for non-linear differential games in pollution control. *Journal of Economic Dynamics and Control* 32, 4, 1312-1331.

Lemoine, D. and C.P. Traeger. 2014. Watch your step: optimal policy in a tipping climate. *American Economic Journal: Economic Policy* 6, 1, 137-166.

Lenton, T.M. and J.-C. Ciscar. 2013. Integrating tipping points into climate impact assessments. *Climate Change* 117, 3, 585-597.

Mäler, K.-G., A. Xepapadeas and A. de Zeeuw. 2003. The economics of shallow lakes. *Environmental & Resource Economics* 26, 4, 603-624.

v.d. Ploeg, F. and A. de Zeeuw. 2018. Climate tipping and economic growth: precautionary capital and the price of carbon. *Journal of the European Economic Association* 16, 5, 1577-1617.

Polasky, S., A. de Zeeuw and F.O.O. Wagener. 2011. Optimal management with potential regime shifts. *Journal of Environmental Economics and Management* 62, 2, 229-240.

Rockström, J., W. Steffen, K. Noone, A. Persson, F.S. Chapin III, E.F. Lambin, T.M. Lenton, M. Scheffer, C. Folke, H.J. Schellnhuber, B. Nykvist, C.A. de Wit, T. Hughes, S. van der Leeuw, H. Rodhe, S. Sörlin, P.K. Snyder, R. Constanza, U. Svedin, M. Falkenmark, L.

Karlberg, R.W. Corell, V.J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen and J.A. Foley. 2009. A safe operating space for humanity. *Nature* 461, 472-475.

Scheffer, M. 1997. *Ecology of Shallow Lakes*. Chapman and Hall, London.

Scheffer, M., S.R. Carpenter, J.A. Foley, C. Folke and B. Walker. 2001. Catastrophic shifts in ecosystems. *Nature* 413, 591-596.

Wagener, F.O.O. 2003. Skiba points and heteroclinic bifurcations, with applications to the shallow lake system. *Journal of Economic Dynamics and Control* 27, 9, 1533-1561.

Appendix A: the Barrett (2013) model

If the maximal amount of emissions is fixed, abatements are foregone emissions. Although in our model the maximal amount of emissions is not fixed, our model is close to the abatement model. This Appendix shows that also in his context, partial coalitions may obtain better outcomes.

Tipping, or crossing a threshold, induces a structural change in the dynamics of the climate system with a significant loss of welfare. Therefore, it can be collectively optimal to reduce total emissions more than in the absence of a tipping point, in order to prevent tipping. Moreover, it can also be the mutual best response of individual countries to prevent tipping in a Nash equilibrium, although it is to be expected that the loss of welfare must be higher in this case. In case the loss of welfare is not sufficiently high for this, stable partial cooperation solves most of the problem.

Barrett (2013) uses a simple game where *n* countries choose abatement levels $a_i, i = 1,2,...,n,$ with cost functions $C_i(a_i) = 0.5ca_i^2$ and benefit functions $B_i(a) = ba$, where a denotes total abatement. Without loss of generality, the parameters can be normalized to $b = c = 1$. It is easy to see that the full-cooperative solution is $a_i = n, i = 1,2,...,n,$ with total abatement $a = n^2$, and net benefits $0.5n^2$

for each country. The Nash equilibrium is $a_i = 1, i = 1,2,...,n,$ with total abatement $a = n$, and net benefits $n - 0.5$ for each country.

If the total abatement a is not sufficiently high and stays below the critical level denoted by a^{tp}, the climate system tips which yields a loss l for each country. It follows that the benefit functions become $B_i(a) = a - l$, if $a < a^{tp}$, and $B_i(a) = a$, if $a \geq a^{tp}$. If the countries cooperate, they choose in the absence of a tipping point total abatement equal to n^2. If $n^2 > a^{tp}$, they will continue to do so, because tipping will not occur. However, if $n^2 < a^{tp}$, tipping will occur, so that the net benefits become $0.5n^2 - l$ for each country. The countries can consider now to abate up to the critical level a^{tp}, in order to prevent tipping and the loss l to each country. Because of symmetry, each country abates a^{tp} / n. It is better for the countries collectively to abate up to the critical level a^{tp}, if the net benefits for each country are higher than in case they let tipping occur, i.e.

$$a^{tp} - 0.5\left(\frac{a^{tp}}{n}\right)^2 \geq n^2 - 0.5n^2 - l, \text{ or} \tag{A1}$$

$$l \geq 0.5\left(\frac{a^{tp}}{n} - n\right)^2. \tag{A2}$$

If the countries do not cooperate, they choose total abatement equal to n in the Nash equilibrium, in the absence of a tipping point. If $n > a^{tp}$, the Nash equilibrium does not change, because tipping will not occur. However, if $n < a^{tp}$, tipping will occur, so that the net benefits become $n - 0.5 - l$ for each country. The question now is whether $a_i = a^{tp} / n, i = 1,2,...,n,$ can be a Nash equilibrium, so that a Nash equilibrium prevents tipping. This requires that an individual country does not have an incentive to deviate, in case the other countries choose the abatement level a^{tp} / n. If a country

deviates, it chooses the abatement level $1 < a^{tp}/n$ (with costs 0.5) and accepts the loss l, because tipping will occur. It is better for a country not to deviate if

$$a^{tp} - 0.5\left(\frac{a^{tp}}{n}\right)^2 \geq 1 - 0.5 + \frac{n-1}{n}a^{tp} - l, \text{ or} \tag{A3}$$

$$l \geq 0.5\left(\frac{a^{tp}}{n} - 1\right)^2. \tag{A4}$$

The right-hand side of (A4) is larger than the right-hand side of (A2) above. This is intuitively clear, because it requires a larger loss *l* to suppress free riding than to induce the countries to avoid tipping when they cooperate.

The important conclusion is that if condition (A4) holds, it is not only collectively optimal to choose $a_i = a^{tp}/n, i = 1,2,...,n,$ and avoid tipping but this is also a Nash equilibrium. For values of l where condition (A2) holds but condition (A4) does not hold, it is collectively optimal to avoid tipping, but this cannot be achieved in a Nash equilibrium. If $a_i = a^{tp}/n, i = 1,2,...,n,$ is a Nash equilibrium, in most cases $a_i = 1, i = 1,2,...,n,$ remains a Nash equilibrium, because it is too costly for an individual country to increase abatement up to a^{tp} by itself. If the game has two Nash equilibria, it becomes a coordination game. It is clear that if it is collectively optimal to choose $a_i = a^{tp}/n, i = 1,2,...,n,$ and if both Nash equilibria exist, it is better for the countries to coordinate on the Nash equilibrium $a_i = a^{tp}/n, i = 1,2,...,n,$ than on the other Nash equilibrium. It follows that if condition (A4) holds, the cooperative outcome and the best non-cooperative Nash equilibrium coincide. It is best for an individual country to choose $a_i = a^{tp}/n$, if the other countries do the same, and the outcome is collectively optimal. For a sufficiently large loss l of tipping, a tipping point in the climate system allows the countries to coordinate on a non-cooperative Nash

equilibrium with the same outcome as when they would cooperate. The question remains if stable partial cooperation can improve the situation in the range of l where it is optimal to prevent tipping but where this cannot be realized in a Nash equilibrium, i.e.

$$0.5\left(\frac{a^{tp}}{n}-n\right)^2 \le l < 0.5\left(\frac{a^{tp}}{n}-1\right)^2. \tag{A5}$$

If a coalition of size $k<n$ forms, the first question is whether a Nash equilibrium exists between the coalition and the $n-k$ outsiders, with total abatement equal to the critical level a^{tp}. Note that the situation is asymmetric now, with coalition members and outsiders. Denote the contribution of a coalition member to the critical abatement level by a^{tpm}, and the contribution of an outsider by a^{tpo}. Given the contributions of the outsiders, the coalition has the choice to collectively abate k^2 and let tipping occur, with the costs $0.5k^2$ and the loss l to each coalition member, or to contribute each the abatement level a^{tpm}. It is better for the coalition to contribute, if the net benefits for each member are higher than in case they let tipping occur, i.e.

$$a^{tp}-0.5\left(a^{tpm}\right)^2 \ge 0.5k^2+(n-k)a^{tpo}-l, \text{ or} \tag{A6}$$

$$l \ge 0.5k^2+(n-k)a^{tpo}-a^{tp}+0.5\left(a^{tpm}\right)^2. \tag{A7}$$

Similarly, given the contribution of the coalition and the other $n-k-1$ outsiders, an outsider has the choice to abate 1 and let tipping occur, with the costs 0.5 and the loss l, or to contribute the abatement level a^{tpo}. It is better for the outsider to contribute, if the net benefits are higher than in case the outsider let tipping occur, i.e.

$$a^{tp}-0.5\left(a^{tpo}\right)^2 \ge 0.5+ka^{tpm}+(n-k-1)a^{tpo}-l, \text{ or} \tag{A8}$$

$$l \ge 0.5+ka^{tpm}+(n-k-1)a^{tpo}-a^{tp}+0.5\left(a^{tpo}\right)^2. \tag{A9}$$

Finally, the total abatement of the coalition of size k and the $n-k$ outsiders must meet the critical abatement level a^{tp}, i.e.

$$ka^{tpm}+(n-k)a^{tpo}=a^{tp}. \tag{A10}$$

It is easy to show that condition (A7) and condition (A9) coincide, if

$$a^{tpm}=a^{tpo}+k-1. \tag{A11}$$

Combining conditions (A10) and (A11) yields a Nash equilibrium between the coalition of size k and the $n-k$ outsiders, given by

$$a^{tpm}=\frac{a^{tp}}{n}+\frac{(n-k)(k-1)}{n}, a^{tpo}=\frac{a^{tp}}{n}-\frac{k(k-1)}{n}, \tag{A12}$$

under the condition on the loss l, given by

$$l\geq 0.5\left(\frac{a^{tp}}{n}-\frac{n+k^2-k}{n}\right)^2. \tag{A13}$$

It is clear that for $k=n$, (A13) reduces to (A2), and for $k=0$ or $k=1$, (A13) reduces to (A4). This shows that if l decreases in (A5) from the level that is needed to prevent tipping in a Nash equilibrium to the level for which the prevention of tipping is still optimal, increasing the size k of the coalition prevents tipping. Specifically, if

$$0.5\left(\frac{a^{tp}}{n}-\frac{n+k^2-k}{n}\right)^2\leq l<0.5\left(\frac{a^{tp}}{n}-\frac{n+(k-1)^2-(k-1)}{n}\right)^2, \tag{A14}$$

a coalition of at least size k is needed to be able to prevent tipping in a Nash equilibrium between the coalition and the outsiders. Equations (A12) yield the abatement levels of coalition members and outsiders. However, the second question remains, i.e. whether this partial cooperation is stable. A coalition of size $k>1$ is internally stable if a member of the coalition does not have an incentive to become an outsider to the coalition of size $k-1$. A coalition of size $k<n$ is externally stable

if an outsider does not have an incentive to join the coalition and increase the size of the coalition to $k+1$. In the absence of a tipping point, a coalition of size k is internally stable if

$$0.5k^2 + n - k \geq 0.5 + (k-1)^2 + n - k \Rightarrow 0.5(k-1)(k-3) \leq 0, \quad \text{(A15)}$$

and externally stable if

$$0.5 + k^2 + n - k - 1 \geq 0.5(k+1)^2 + n - k - 1 \Rightarrow 0.5k(k-2) \geq 0. \quad \text{(A16)}$$

This implies that only the coalitions of size 2 and size 3 are stable in this case.

For l satisfying condition (A14), leaving the coalition means that tipping cannot be avoided. The remaining coalition will then choose $(k-1)^2$ and the outsiders will choose 1, inducing the loss l for each country because tipping will occur. It follows that a coalition member does not have an incentive to leave the coalition, if the net benefits as a member of the coalition of size k are larger than the net benefits as an outsider to the coalition of size $k-1$, i.e.

$$a^{tp} - 0.5\left(a^{tpm}(k)\right)^2 \geq 0.5 + (k-1)^2 + n - k - l, \quad \text{(A17)}$$

where $a^{tpm}(k)$ is the contribution of a member of the coalition of size k to the critical level a^{tp}.

Proposition A1: if condition (A14) on the tipping loss l holds, so that a Nash equilibrium between the coalition of size k and the $n-k$ outsiders prevents tipping, this coalition is internally stable if

$$0.5(k-1)(k-3) \leq (n-k)\sqrt{2p(k, a^{tp}, n)}, \quad \text{(A18)}$$

where $p(k, a^{tp}, n)$ denotes the lower bound of the tipping loss l in (A14).

Proof:

Using (A12) and (A14), the condition for internal stability (A17) holds if

$$a^{tp} - 0.5\left(\frac{a^{tp}}{n} + \frac{(n-k)(k-1)}{n}\right)^2 \geq$$
$$0.5 + (k-1)^2 + n - k - 0.5\left(\frac{a^{tp}}{n} - \frac{n+k^2-k}{n}\right)^2. \quad \text{(A19)}$$

Using $(n-k)(k-1) + n + k^2 - k = nk$, condition (A19) can be rewritten as

$$\frac{n-k}{n} a^{tp} \geq -0.5k^2 + \frac{k(n-k)(k-1)}{n} + (k-1)^2 + (n-k+1) - 0.5, \quad \text{(A20)}$$

which yields

$$0.5(k-1)(k-3) \leq (n-k)\left(\frac{a^{tp}}{n} - \frac{n+k^2-k}{n}\right). \quad \text{(A21)}$$

The second term between brackets on the right-hand side of condition (A21) is the square root of twice the lower bound of the tipping loss l in (A14). Q.E.D.

Comparing conditions (A18) and (A15), it follows that the size of coalition that is internally stable can be much larger if tipping is possible. It is immediately clear that the coalitions of size 2 and size 3 are internally stable, and that the coalition of size *n* is not internally stable if $n \geq 4$. Note that if k increases, the left-hand side of (A18) increases and the right-hand side decreases, so that at some point internal stability is lost. For example, if $a^{tp} = 150$ and $n = 10$, condition (A18) still holds for $k = 7$, but it does not hold anymore for $k = 8$. This implies that for $48.02 \leq l < 60.5$ (i.e., inequalities (A14), $k = 7$), the coalition of size $k = 7$ prevents tipping and is internally stable. However, for $35.28 \leq l < 48.02$ (i.e., inequalities (A14), $k = 8$), increasing the size of the coalition to $k = 8$ prevents tipping, but it is not internally stable anymore. At some point, the incentive to free ride dominates the incentive to prevent tipping, because of the high level of cooperation and the low level of tipping costs l. The coalition of size $k = 7$ is also externally stable, because it is

not beneficial for an outsider to join the coalition and contribute a higher level of abatement to the critical level a^{tp} as a member of the coalition of size $k=8$ than as an outsider to the coalition of size $k=7$ (see (A12)). It follows that the size of the stable coalition is $k=7$ in this case.

Inequalities (A5) show values of the tipping loss l for which it is collectively optimal to prevent tipping but for which this cannot be sustained in a Nash equilibrium. In our example, this range is $12.5 \le l < 98$. For $l \ge 98$, a Nash equilibrium prevents tipping, and for $l < 12.5$, it is better to let tipping occur. Stable partial cooperation covers a large part of this range, i.e. $48.02 \le l < 98$, with coalitions up to size $k=7$. In the range $12.5 \le l < 48.02$, partial cooperation with larger coalitions can prevent tipping, but these coalitions are not internally stable. However, the good news is that in this case, the tipping loss l is small.

Appendix B: Proofs

Proof Proposition 2. From (13) we have

$$\bar{c}(b,n) = \frac{1}{2sx}, \tag{B1}$$

with

$$x = s - b - \frac{n-1}{n}, F(s,b,n) = 2sx\ln(nx) - s^2 + 1 = 0. \tag{B2}$$

For $b=0$, $F(s,0,n)=0$ has a unique solution $\bar{s}(0,n)=1$, so that $\bar{c}(0,n)=n/2$.

For $b>0$, $F(s,b,n)<0$, if $1 \le s \le b+1$, and $F(s,b,n) \to \infty$, if $s \to \infty$. For $s > b+1$ ($x > 1/n$),

$$\frac{\partial F(s,b,n)}{\partial s} = 2(s+x)\ln(nx) > 0. \tag{B3}$$

Hence, for $b>0$, $F(s,b,n)=0$ has a unique solution $\bar{s}(b,n) > b+1$.

It follows that $\bar{c}(b,n) < n/(2(b+1))$. The implicit function theorem yields

$$\frac{\partial \bar{s}(b,n)}{\partial b} = -\frac{\partial F}{\partial b} / \frac{\partial F}{\partial s} = \frac{\bar{s}}{\bar{s}+\bar{x}} \frac{\ln(n\bar{x})+1}{\ln(n\bar{x})}, \tag{B4}$$

so that

$$\frac{\partial \bar{c}(b,n)}{\partial b} = -\frac{(\bar{s}+\bar{x})\partial \bar{s}(b,n)/\partial b - \bar{s}}{2\bar{s}^2\bar{x}^2} = -\frac{1}{2\bar{s}\bar{x}^2 \ln(n\bar{x})} < 0. \tag{B5}$$

Similarly, the implicit function theorem yields

$$\frac{\partial \bar{s}(b,n)}{\partial n} = -\frac{\partial F}{\partial n} / \frac{\partial F}{\partial s} = \frac{\bar{s}}{\bar{s}+\bar{x}} \frac{\ln(n\bar{x})+1-n\bar{x}}{n^2 \ln(n\bar{x})}, \tag{B6}$$

so that

$$\frac{\partial \bar{c}(b,n)}{\partial n} = -\frac{(\bar{s}+\bar{x})\partial \bar{s}(b,n)/\partial n - \bar{s}/n^2}{2\bar{s}^2\bar{x}^2} = \frac{n\bar{x}-1}{2\bar{s}\bar{x}^2 n^2 \ln(n\bar{x})} > 0. \tag{B7}$$

The properties of $\bar{c}(b,n)$ in Proposition 2 follow immediately.

Proof Proposition 4. The critical level $\bar{c}(b, n-k+1)$ is determined by the system (17) of equations in $(\bar{c},\bar{x},\bar{s})$. Omitting upper bars for convenience, and introducing $y=(n-k+1)x$ as parameter, the system (17) can be rewritten as the system of equations

$$s^2 - 2\frac{y\ln y}{n-k+1}s - 1 = 0, c = \frac{n-k+1}{2sy}, b = s - \frac{y+n-k}{n-k+1} \tag{B8}$$

in (s,c,b).

Similarly, the critical value of c for internal stability, given by equations (19) can be written as the system of equations

$$c = \frac{1}{2sx}, (n-k+2)x = s-b, \ln[k(n-k+1)x] = cs^2 - c, \tag{B9}$$

and by introducing $z=k(n-k+1)x$ as parameter, also as the system of equations

$$s^2 - 2\frac{z\ln z}{k(n-k+1)}s - 1 = 0, c = \frac{k(n-k+1)}{2sz}, b = s - \frac{(n-k+2)z}{k(n-k+1)} \tag{B10}$$

in (s,c,b).

Starting with the system of equations (B8), the positive root of the quadratic equation in s is

$$s_1 = \frac{y\ln y}{n-k+1} + \frac{y\ln y}{n-k+1}\sqrt{1+\left(\frac{n-k+1}{y\ln y}\right)^2} = \frac{2y\ln y}{n-k+1} + O((y\ln y)^{-1}). \tag{B11}$$

Furthermore,

$$\frac{1}{s_1} = -\frac{y\ln y}{n-k+1} + \frac{y\ln y}{n-k+1}\sqrt{1+\left(\frac{n-k+1}{y\ln y}\right)^2} = \frac{n-k+1}{2y\ln y} + O((y\ln y)^{-3}). \tag{B12}$$

Using (B8) and (B11), it follows that

$$b_1 = s_1 - \frac{y+n-k}{n-k+1} = \frac{2y\ln y - y - (n-k)}{n-k+1} + O((y\ln y)^{-1}), \tag{B13}$$

and using (B8) and (B12), it follows that,

$$c_1 = \frac{n-k+1}{2s_1 y} = \frac{(n-k+1)^2}{4y^2\ln y} + O(y^{-4}(\ln y)^{-3}). \tag{B14}$$

Similarly, starting with the system of equations (B10), the positive root of the quadratic equation in s is

$$s_2 = \frac{2z\ln z}{k(n-k+1)} + O((z\ln z)^{-1}), \tag{B15}$$

so that

$$b_2 = s_2 - \frac{(n-k+2)z}{k(n-k+1)} = \frac{2z\ln z - (n-k+2)z}{k(n-k+1)} + O((z\ln z)^{-1}), \tag{B16}$$

and

$$c_2 = \frac{k(n-k+1)}{2s_2 z} = \frac{k^2(n-k+1)^2}{4z^2\ln z} + O(z^{-4}(\ln z)^{-3}). \tag{B17}$$

In order to combine the requirements that the coalition of size k prevents tipping and is internally stable, the solution sets (B13, B14) and (B16, B17) for b and c have to be consistent. For small values of c or equivalently, for large values of b, which correspond to large values of y and z, we can ignore the lower-order terms. First, we claim the relationship $z = ky/\sqrt{1+p}$ between z and y, and consider the difference $\Delta_c(p) := c_2 - c_1$, using (B14) and (B17):

$$\Delta_c(p) = \frac{(n-k+1)^2}{4y^2 \ln y \ln\left(ky/\sqrt{1+p}\right)}(p \ln y - \ln k + 0.5\ln(1+p)). \tag{B18}$$

We search for p such that $\Delta_c(p) = 0$. We can focus on p such that the second term of the right-hand side of (B18) is 0. For small p, we can approximate $0.5\ln(1+p)$ by $0.5p$. It follows that

$$\frac{\ln k}{\ln y + 0.5} < p < \frac{\ln k}{\ln y} \Rightarrow p = \frac{\ln k}{\ln y} + O((\ln y)^{-2}), \tag{B19}$$

and thus

$$\begin{aligned} z &= ky - 0.5ky\frac{\ln k}{\ln y} + O(y(\ln y)^{-2}), \\ \ln z &= \ln k + \ln y + O((\ln y)^{-1}), \\ z\ln z &= ky\ln y + 0.5ky\ln k + O(y(\ln y)^{-1}). \end{aligned} \tag{B20}$$

Finally, we consider the difference $\Delta_b := b_2 - b_1$, using (B13), (B16) and (B20):

$$\Delta_b = \frac{y}{n-k+1}(\ln k - (n-k+1)) + O(y(\ln y)^{-1}). \tag{B21}$$

It follows that for small values of c, $\Delta_b < 0$, if $k - 1 + \ln k < n$, and $\Delta_b > 0$, if $k - 1 + \ln k > n$. Therefore, if k^* is the largest integer that satisfies $k - 1 + \ln k < n$, a coalition of size k^* is the largest coalition that prevents tipping and is internally stable.

Proof Proposition 5. The Lagrangian for the best response of an economic agent in the inverse tipping game is

$$L_i = \ln e_i - cs^2 + \lambda_i \left(e_i + \frac{n-1}{n}(1-b) - f(s) \right). \tag{B22}$$

With the transformation $\hat{e}_i = e_i / (1-b)$, $\hat{s} = s/(1-b)$, $\hat{b} = b/(1-b)$, $\hat{c} = c(1-b)^2$, $\hat{\lambda}_i = \lambda_i (1-b)$, this Lagrangian changes into

$$\hat{L}_i = \ln \hat{e}_i - \hat{c}\hat{s}^2 + \ln(1-b) + \hat{\lambda}_i \left(\hat{e}_i + \frac{n-1}{n} - \hat{f}(\hat{s}) \right), \tag{B23}$$

where the response function $\hat{f}$ tips at $\hat{s}^{tp} = 1/(1-b)$. This is the same Lagrangian as the one for the best response of an economic agent in the tipping game, except for the constant $\ln(1-b)$. It follows that the two problems are equivalent.

The internal stability conditions are equivalent as well. The inverse tipping game has the condition

$$\ln \frac{1-b}{k(n-k+1)} - c(1-b)^2 \geq \ln e_i - cs^2. \tag{B24}$$

With the transformation above, this condition becomes

$$\ln \frac{1}{k(n-k+1)} - \hat{c} \geq \ln \hat{e}_i - \hat{c}\hat{s}^2, \tag{B25}$$

which is the condition for the tipping game.

It follows that proposition 4 proves proposition 5. The parameter transformation $\hat{b} = b/(1-b)$ and $\hat{c} = c(1-b)^2$ maps the regions of the inverse tipping game in Figure 5 into the regions of the tipping game in Figure 3.