Houba, Harold; Ansink, Erik

Working Paper

Sustainable Agreements on Stochastic River Flow

Tinbergen Institute Discussion Paper, No. 13-182/II

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Houba, Harold; Ansink, Erik (2013) : Sustainable Agreements on Stochastic River Flow, Tinbergen Institute Discussion Paper, No. 13-182/II, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/87574

TI 2013-182/II
Tinbergen Institute Discussion Paper

Sustainable Agreements on Stochastic River Flow

Harold Houba
Erik Ansink

Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Sustainable agreements on stochastic river flow*

Erik Ansink† Harold Houba‡

Abstract

Many water allocation agreements in transboundary river basins are inherently unstable. Due to stochastic river flow, agreements may be broken in case of drought. The objective of this paper is to analyze whether water allocation agreements can be self-enforcing, or sustainable. We do so using an infinitely-repeated sequential game that we apply to several classes of agreements. To derive our main results we apply the Folk Theorem to the river sharing problem using the equilibrium concepts of subgame-perfect equilibrium and renegotiation-proof equilibrium. We show that, given the upstream-downstream asymmetry, sustainable agreements allow downstream agents to reap the larger share of the benefits of cooperation.

JEL Classification: C73, D74, F53, Q25

Keywords: river sharing, sustainable agreements, repeated sequential game, Folk Theorem, water allocation, renegotiation-proofness

Corresponding author:
Erik Ansink
Department of Spatial Economics
VU University Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
the Netherlands

Email: erik.ansink@vu.nl
Tel: +31 20 598 1214

*Small parts of this paper are based on a completely revised version of FEEM Working Paper 73.2009, 'Self-enforcing agreements on water allocation', by the first author. The authors thank Arjan Ruijs and Hans-Peter Weikard for helpful comments on this earlier version of the paper. The first author acknowledges financial support from FP7-IDEAS-ERC Grant No. 269788.

†Department of Spatial Economics and IVM, VU University Amsterdam, and Tinbergen Institute.

‡Department of Econometrics, VU University Amsterdam, and Tinbergen Institute.

1 Introduction

We apply the theory of repeated games to the river sharing problem. Our main contribution is the design of agreements that are sustainable to stochastic river flow in a dynamic setting. Doing so, we add to the rapidly growing literature on the analysis of solutions to the river sharing problem (cf. Béal et al., 2013; Van den Brink et al., 2012; Ambec et al., 2013), which has largely ignored dynamics.

In an international river basin, when water is scarce, countries may exchange water for side payments (Dinar, 2006; Carraro et al., 2007). This type of exchange is generally formalized in a water allocation agreement. The aim of water allocation agreements is to increase the overall efficiency of water use. This increase in efficiency can be obstructed by the stochastic nature of river flow, because countries may find it profitable to break the agreement in case of drought (Ward, 2013). A recent example is Mexico's failure to meet its required average water deliveries under the 1944 US-Mexico Water Treaty in the years 1992–1997 (Gastélum et al., 2009). Additional case study evidence on agreement breakdowns because of droughts can be found, for instance, in Barrett (1994) and Beach et al. (2000). Only a minority of current international agreements take into account the variability of river flow (De Stefano et al., 2012). Most agreements do not; they either allocate fixed or proportional shares, or they are ambiguous in their schedule for water allocation. Both the efficiency and stability (Bennett and Howe, 1998; Bennett et al., 2000; Ansink and Ruijs, 2008; Ambec et al., 2013) of such agreements may be hampered. These effects could be worsened by the impacts of climate change on river flow.

In order to accommodate for stochastic river flow, Kilgour and Dinar (2001) developed a flexible water allocation agreement that provides an efficient allocation for every possible level of river flow. This agreement maximizes the overall benefits of water use, after which side payments are made such that each country benefits from cooperation. This flexible agreement assures efficiency, but not stability because it ignores the repeated interaction of countries over time. Countries have an incentive to defect from the agreement when the benefits of defecting outweigh the benefits of compliance. Note that there is no supra-national authority that can enforce this type of international agreements. This implies that a stable agreement has to be self-enforcing or sustainable, in the sense that each agent should have an incentive to comply with the agreement. In such a setting, application of repeated-game theory to the setting of river sharing seems natural, but to the best of our knowledge, this has not been done yet.[1]

[1]This paper is therefore a contribution to the challenge raised by Carraro et al. (2007): *"Water re-*

Given the asymmetry imposed by the geography of the river, we adopt an infinitely-repeated sequential game, in which upstream agents move before downstream agents.[2] We argue that the Folk Theorem for infinitely-repeated sequential games is not very informative, since it is only a limit result on the discount factor δ. For the practical purpose of this paper, the real issue is firstly, given some δ, how to construct sustainable agreements, and secondly whether certain classes of agreements have properties that may be appealing for implementation by policy makers. For instance, we will assess the effects of non-negativity restrictions on per-period payoffs and we will look at some disadvantages of fixed-payment agreements, which are common in practice.

To derive our main results we apply the Folk Theorem to the river sharing problem using the equilibrium concepts of subgame-perfect equilibrium and renegotiation-proof equilibrium. We will see that, given the upstream-downstream asymmetry, sustainable agreements allow downstream agents to reap the larger share of the benefits of cooperation. This distribution of gains is the opposite of some papers that assess agreements on river sharing in a static setting (e.g. Ambec et al., 2013). Our results provide economic intuition for an empirical result (downstream states managing to negotiate a substantial share of upstream river water) that has, up till now, mostly been explained by political factors (Dinar, 2009; Katz and Moore, 2011).

In the next section we introduce our model and present our first result on min-max values in the river sharing problem. We derive equilibrium conditions in Section 3, which we use in an example in Section 4 and subsequent detailed analyses of four subsets of agreements in Section 5, including Nash-bargaining agreements and renegotiation-proof agreements. Finally, in Section 6, we provide some concluding remarks. Our main result, Proposition 2 in Section 5.2, presents our Folk Theorem for river sharing problems, which is further refined in Propositions 3 and 5. Proposition 4 and Corollary 2 provide a practical interpretation of these results. Two appendices contain proofs as well as detailed information on context and generalization of our analysis.

sources are intrinsically unpredictable, and the wide fluctuations in water availability are likely to become more severe over the years. Formally addressing the stochasticity of the resource, as well as the political, social, and strategic feasibility of any allocation scheme, would significantly contribute to decreasing conflicts over water."

[2]This sequence of moves according to the agents' geographical location seems most natural. One additional argument to support this sequence is that payments to compensate for water deliveries can easily be deferred while water deliveries themselves cannot.

2 Model

Consider two agents $i = 1,2$ with agent 1 upstream of agent 2 along a river. Denote river flow in period $t = (1,2,\ldots)$ by the vector $(e_{1,t}, e_{2,t})$, which includes flow contributions on the territory of agents 1 and 2. River flow is stochastic and is drawn in each period from a bivariate probability distribution with density function $f(e_1,e_2)$ on a compact subset of $[\underline{e}_1,\overline{e}_1] \times [\underline{e}_2,\overline{e}_2] \subset \boldsymbol{R}^2_+$ and marginal distributions $f_1(e_1) = \int f(e_1,e_2)de_2$ and $f_2(e_2) = \int f(e_1,e_2)de_1$. Denote water use of agent i in period t by $x_{i,t}$. Any water that was not used by agent 1 flows to the territory of agent 2. For simplicity, we suppress time when confusion cannot occur.

Benefits of water use $b_i(x_i)$ are increasing[3] and strictly concave with $b_i'(x_i) > 0$, $b_i''(x_i) < 0$, and $b_i(0) = 0$. We assume that utility is transferable through monetary payment s from agent 2 to agent 1, which is positive when agent 2 pays $|s|$ to agent 1 and vice versa. The utility of agent i depends on his water use x_i and payments s, and is given by the following quasi-linear utility function:

$$\begin{cases} u_1(x_1,s) &= b_1(x_1)+s; \\ u_2(x_2,s) &= b_2(x_2)-s. \end{cases} \tag{1}$$

There are several focal allocations of river flow that will be used extensively in the remainder of the paper: the Nash allocation[4], the efficient allocation and the minmax allocation. Definitions of the Nash allocation and efficient allocation are given below, while the minmax allocation is the subject of Proposition 1 at the end of this section.

Definition 1 (Nash allocation)**.** For every realization (e_1,e_2), the Nash allocation $x^N(e_1,e_2) = \left(x_1^N(e_1,e_2), x_2^N(e_1,e_2)\right)$ is the unique allocation where each agent uses his own endowment of river flow such that $x_1^N(e_1,e_2) = e_1$ and $x_2^N(e_1,e_2) = e_2$.

Because of increasing benefits of water use, the Nash allocation is evident in absence of water trade or other types of agreements on water use.

Definition 2 (Efficient allocation)**.** For every realization (e_1,e_2), the efficient allocation $x^*(e_1,e_2) = \left(x_1^*(e_1,e_2), x_2^*(e_1,e_2)\right)$ is the unique maximizer of the utilitarian welfare $b_1(x_1) + b_2(x_2)$ subject to the feasibility constraints $x_1 \leq e_1$ and $x_2 \leq e_1 + e_2 - x_1$.

With a strictly concave quasi-linear utility function as in (1), the efficient allocation is unique and equal to utilitarian welfare maximization as in e.g. Kilgour and Dinar

[3] We defer discussing the implications of this assumption to Appendix A.

[4] We prefer to call the unique subgame-perfect equilibrium of the non-repeated river game the Nash equilibrium in order to distinguish single-period play from repeated play.

(2001) and Houba et al. (2013).

The interesting and non-trivial case occurs when water is scarce:

Assumption 1 (Water scarcity)**.** Water is scarce such that there are incentives to cooperate for every realization (e_1, e_2).

This assumption implies $x_1^*(e_1, e_2) < e_1$ and $x_2^*(e_1, e_2) = e_1 + e_2 - x_1^*(e_1, e_2)$, and therefore $x_2^*(e_1, e_2) > e_2$. Realizations without incentives to cooperate can be ignored because any efficient agreement will specify no cooperation for each of these realizations, which is trivially sustainable.

Remark 1. Our model setup is consistent with much of the river sharing literature. The case of two agents (Ansink and Ruijs, 2008; Houba, 2008) provides no limitation for reasons similar to infinitely-repeated games with n players and full dimensionality. Full dimensionality is assured in river sharing problems because with n players the monetary transfers allow redistribution of transferable utility in all n utility dimensions. By focusing on two agents instead of the general case with more agents and more realistic river geographies (Ansink and Houba, 2012; Van den Brink et al., 2012), we are able to avoid some complexity and excessive notation. It also provides a better understanding of the issues involved in repeated interaction of the agents over time in the presence of a stochastic river flow (Kilgour and Dinar, 2001; Ambec et al., 2013). In Appendix A, we elaborate on the general case.

We proceed to describe the possibility of agreements on water allocation between the two agents. Because agent 1 is upstream of agent 2, he can take out any water, subject to feasibility. Such unilateral action causes an inefficient allocation of water whenever $x_1(e_1, e_2) \neq x_1^*(e_1, e_2)$. Instead, the agents may cooperate by signing an agreement that specifies the following three elements: (i) an allocation rule for river flow, (ii) a payment rule for monetary transfers, and (iii) punishment strategies in case one of the agents deviates from the agreement. An agreement coincides with cooperative play in the repeated game that we introduce below. In this repeated game, both the allocation of water and the payment may be contingent on realized river flow as in Kilgour and Dinar (2001). In reality, however, we also observe lump-sum payments and simpler allocation rules, including fixed and proportional allocations (Ansink and Ruijs, 2008; Drieschova et al., 2008). Punishment strategies are essential for the sustainability of the agreement because, in absence of a supra-national authority, agreements are non-binding. They determine what happens upon deviation and range from simple trigger strategies to more advanced strategies that assure *ex post* credibility of the punishment.

In actual agreements on river flow allocation, punishment strategies are often lacking, although many do contain clauses on conflict resolution (Beach et al., 2000; Ward, 2013).

To assess possible agreement specifications, we model each period as an extensive game where nature moves first by drawing a realization of (e_1, e_2). Subsequently, agent 1 moves by choosing $x_1(e_1, e_2)$ and, finally, agent 2 moves by choosing $x_2(e_1, e_2)$ and, in case of an agreement, agent 2 also decides whether to make the payment $s(e_1, e_2)$. Cooperative play consists of a stochastic sequence of water resources (e_1, e_2) that induce the water use vector $x^c(e_1, e_2) = \left(x_1^c(e_1, e_2), x_2^c(e_1, e_2)\right)$ and the payment $s^c(e_1, e_2)$ as specified in the agreement. Non-cooperative play consists of the Nash allocation and a zero payment so that the choice of $x_i(e_1, e_2)$, $i = 1, 2$, is limited to a binary strategy set $\left\{x_i^c(e_1, e_2), e_i\right\}$ and the choice of s is in essence limited to a binary strategy set $\{s^c(e_1, e_2), 0\}$. Only if the agreement maximizes utilitarian welfare we have $x^c(e_1, e_2) = x^*(e_1, e_2)$ for every realization (e_1, e_2), but this is not necessarily the case, as agents may agree otherwise.

The Folk Theorem for infinitely-repeated sequential games states that any utility vector that yields each agent more than his minmax utility can be supported as an SPE (subgame-perfect equilibrium) utility vector for sufficiently large discount factors Wen (2002). Following this reference, we start our analysis by deriving the minmax utilities in the river sharing problem. We do this in Appendix A, where we also discuss the general case (i.e. more than two agents) as well as the implications of assuming increasing benefit functions for the minmax values. We state the following result.

Proposition 1. *For every realization* (e_1, e_2)*, agent i's minmax value is* $b_i(e_i)$*.*

The agents' minmax values coincide with the unique Nash equilibrium utilities. This result has an important implication in that the Folk Theorem can be derived within the class of trigger strategies, which simplifies the analysis (details are in Appendix A). Formally, in expectations, every $\left(V_1^c, V_2^c\right) > \left(\boldsymbol{E}\{b_1(e_1)\}, \boldsymbol{E}\{b_2(e_2)\}\right)$ can be supported for sufficiently high δ.[5] All such $\left(V_1^c, V_2^c\right)$ can be sustained by trigger strategies as the agreement's punishment strategies and, for explanatory convenience, we make this assumption. It has to be dropped only when we characterize renegotiation-proof equilibria in Section 5.4.

Assumption 2 (Trigger strategies)**.** Both agents use trigger strategies in which deviation from cooperative play, i.e. the agreement, is punished by non-cooperative play

[5]Whether boundary solutions such as $\left(\boldsymbol{E}\{b_1(e_1)\}, V_2^c\right)$ and $\left(V_1^c, \boldsymbol{E}\{b_2(e_2)\}\right)$ can also be supported by trigger strategies or require more complex punishment strategies depends upon the application. Technically speaking, in the limit as δ goes to 1, the closure of the limit set of SPE utility vectors that can be supported by trigger strategies is the set consisting of $\left(V_1^c, V_2^c\right) \geq \left(\boldsymbol{E}\{b_1(e_1)\}, \boldsymbol{E}\{b_2(e_2)\}\right)$.

forever.

Trigger strategies are based upon simple discontinuous contracts. Agent 1 only delivers his part of the agreed water allocation if he has received the agreed payment in the previous period. That is, $e_{1,t} - x_{1,t} = 0$ if $s_{t-1} < s^c_{t-1}\left(e_{1,t-1}, e_{2,t-1}\right)$ and $e_{1,t} - x_{1,t} = e_{1,t} - x^c_{1,t}\left(e_{1,t}, e_{2,t}\right)$ otherwise. Agent 2 only makes the agreed payment if he received the agreed water allocation in the same period. That is, $s_t = 0$ if $e_{1,t} - x_{1,t} < e_{1,t} - x^c_{1,t}\left(e_{1,t}, e_{2,t}\right)$ and $s_t = s^c_t\left(e_{1,t}, e_{2,t}\right)$ otherwise.

3 Equilibrium analysis

In this section, we derive equilibrium conditions for the water allocation and payment rules, using SPE as our equilibrium concept and assuming trigger strategies. Given such strategies, agent 1's optimal deviation is $x_1 = e_1$ forever, which implies $x_1 > x^c_1(e_1, e_2)$ in the current period and failing the agreed upon allocation rule in subsequent periods. Likewise, agent 2's optimal deviation is $s(e_1, e_2) = 0$ forever.

Given these punishment strategies, the *ex ante* expected value of the cooperative path to agent $i = 1, 2$ at the beginning of an arbitrary period (i.e. before nature moves) equals

$$V^c_1 = \boldsymbol{E}\left\{b_1\left(x^c_1(e_1, e_2)\right) + s^c(e_1, e_2)\right\} + \delta V^c_1 = \frac{\boldsymbol{E}\left\{b_1\left(x^c_1(e_1, e_2)\right)\right\} + \boldsymbol{E}\left\{s^c(e_1, e_2)\right\}}{1 - \delta}$$

$$V^c_2 = \boldsymbol{E}\left\{b_2\left(x^c_2(e_1, e_2)\right) - s^c(e_1, e_2)\right\} + \delta V^c_2 = \frac{\boldsymbol{E}\left\{b_2\left(x^c_2(e_1, e_2)\right)\right\} - \boldsymbol{E}\left\{s^c(e_1, e_2)\right\}}{1 - \delta}$$

where δ is the discount factor. The ex ante expected value of the non-cooperative path to agent $i = 1, 2$ at the beginning of an arbitrary period (i.e. before nature moves) equals

$$V^n_i = \boldsymbol{E}\left\{b_i(e_i)\right\} + \delta V^n_i = \frac{\boldsymbol{E}\left\{b_i(e_i)\right\}}{1 - \delta}.$$

Combining the *ex ante* expected values, we have

$$\begin{aligned} V^c_1 \;&\geq\; V^n_1 \Longleftrightarrow \boldsymbol{E}\left\{s^c(e_1, e_2)\right\} \;\geq\; \boldsymbol{E}\left\{b_1(e_1)\right\} - \boldsymbol{E}\left\{b_1\left(x^c_1(e_1, e_2)\right)\right\}, \\ V^c_2 \;&\geq\; V^n_2 \Longleftrightarrow \boldsymbol{E}\left\{s^c(e_1, e_2)\right\} \;\leq\; \boldsymbol{E}\left\{b_2\left(x^c_2(e_1, e_2)\right)\right\} - \boldsymbol{E}\left\{b_2(e_2)\right\}, \end{aligned}$$

where both right-hand sides are positive. Furthermore, the maximal per-period expected utilitarian welfare exceeds the per-period $\boldsymbol{E}\left\{b_1(e_1)\right\} + \boldsymbol{E}\left\{b_2(e_2)\right\}$, so a non-empty range of welfare-improving allocation rules $\left(\boldsymbol{E}\left\{x^c_1(e_1, e_2)\right\}, \boldsymbol{E}\left\{x^c_2(e_1, e_2)\right\}\right)$ and payments

$\boldsymbol{E}\{s(e_1,e_2)\}$ exists, compared to the non-cooperative path. Note that this range is maximal in case the allocation rule maximizes utilitarian welfare.

Given the realization of (e_1,e_2) in period t, the equilibrium conditions state that both agents prefer to continue cooperation over a single deviation,[6] knowing that non-cooperation follows forever:

$$V_1^c(e_1,e_2) \;=\; b_1\left(x_1^c(e_1,e_2)\right)+s^c(e_1,e_2)+\delta V_1^c \;\geq\; b_1(e_1)+\delta V_1^n, \tag{2}$$

$$V_2^c(e_1,e_2) \;=\; b_2\left(x_2^c(e_1,e_2)\right)-s^c(e_1,e_2)+\delta V_2^c \;\geq\; b_2\left(x_2^c(e_1,e_2)\right)+\delta V_2^n. \tag{3}$$

In repeated games it is common to derive a threshold for the discount factor δ above which cooperation can be sustained. This is not straightforward in our model setup because the lower bound on δ would become a function of the present state. In Section 5 we will analyze this in detail. Here, we simply show how a given δ imposes bounds on the payments as a function of the allocation rule for river flow:

$$\begin{aligned} s^c(e_1,e_2) \;+\; \frac{\delta}{1-\delta}\boldsymbol{E}\left\{s^c(e_1,e_2)\right\} \;\geq\; & b_1(e_1)-b_1\left(x_1^c(e_1,e_2)\right) \\ & +\; \frac{\delta}{1-\delta}\left[\boldsymbol{E}\{b_1(e_1)\}-\boldsymbol{E}\left\{b_1\left(x_1^c(e_1,e_2)\right)\right\}\right], \end{aligned} \tag{4}$$

$$s^c(e_1,e_2) \;+\; \frac{\delta}{1-\delta}\boldsymbol{E}\left\{s^c(e_1,e_2)\right\} \;\leq\; \frac{\delta}{1-\delta}\left[\boldsymbol{E}\left\{b_2\left(x_2^c(e_1,e_2)\right)\right\}-\boldsymbol{E}\{b_2(e_2)\}\right]. \tag{5}$$

Any agreement that satisfies both (4) and (5) is able to sustain cooperation. This requires a choice of allocation and payment rules that is sufficiently flexible such that the bounds are not violated for any possible realization of river flow. Note the asymmetry in these bounds with respect to the payments; the lower bound always depends upon the realization (e_1,e_2) whereas the upper bound is independent of this realization. This asymmetry is caused by the extensive form of the game, which requires agent 2 to provide a minimum compensation to agent 1 for passed water in the *current* period, which enters the lower bound. In addition, both bounds contain terms that reflect the expected benefits of cooperation in *future* periods. In the next section we use these bounds to illustrate the choice of a payment rule given the unique allocation rule that maximizes utilitarian welfare.

[6] For SPE, the one-stage deviation principle states that it is sufficient to check for profitable single deviations (Fudenberg and Tirole, 1991).

4 Example

In this section we show how bounds (4) and (5) can be used to construct a payment rule for monetary transfers given a fixed parameter value for the discount factor, using a simple example that will recur throughout the paper. In order to do so, we make two additional assumptions.

Assumption 3 (Two realizations of river flow)**.** The density function of river flow is simplified to two possible realizations of river flow (e_1, e_2), high flow $\left(e_1^H, e_2^H\right)$ with probability p and low flow $\left(e_1^L, e_2^L\right)$ with probability $1-p$.

Assumption 4 (Efficient agreement)**.** The agreement maximizes utilitarian welfare so that $x_1^c(e_1,e_2) = x_1^*(e_1,e_2)$ and $x_2^c(e_1,e_2) = x_2^*(e_1,e_2)$.

Given these assumptions, the bounds on the payments in (4) and (5) consist of only exogenous variables and we can illustrate these bounds graphically by a polygon (or a polytope of higher dimension if Assumption 3 had allowed a larger range of realizations of river flow). The line segments that bound the polygon are based on the probability of each realization of river flow. Substituting these two realizations in (4) and (5) and rearranging terms, we obtain that the bounds are given by:

$$\begin{aligned}
\text{low flow:} \qquad & (1-\delta)b_1\left(e_1^L\right) - (1-\delta)b_1\left(x_1^c\left(e_1^L, e_2^L\right)\right) + \delta[A_1] \\
& \leq (1-\delta p)\cdot s^c\left(e_1^L, e_2^L\right) + (\delta p)\cdot s^c\left(e_1^H, e_2^H\right) \leq \delta[A_2], \\
\text{high flow:} \qquad & (1-\delta)b_1\left(e_1^H\right) - (1-\delta)b_1\left(x_1^c\left(e_1^H, e_2^H\right)\right) + \delta[A_1] \\
& \leq (\delta-\delta p)\cdot s^c\left(e_1^L, e_2^L\right) + (1-\delta+\delta p)\cdot s^c\left(e_1^H, e_2^H\right) \leq \delta[A_2],
\end{aligned}$$

where A_1 and A_2 denote the terms between square brackets in the right-hand side of, respectively, (4) and (5). As discussed in Section 3, while both lower bounds depend on the realization of river flow in the current period, the upper bounds do not.

Figure 1 shows the polygon for selected parameter values, illustrating the range of possible combinations of payments that provide sustained cooperation under Assumptions 2, 3 and 4. Any point in the graph represents a payment rule, but only those in the shaded area sustain cooperation. Somewhat counter-intuitively, Figure 1 illustrates the possibility of a negative payment under one of the two possible realizations of river flow. This gives the striking possibility that agent 1 delivers water *and* a payment to agent 2. Obviously, a negative payment under one realization is accompanied by a relatively large positive payment under the alternative realization of river flow.

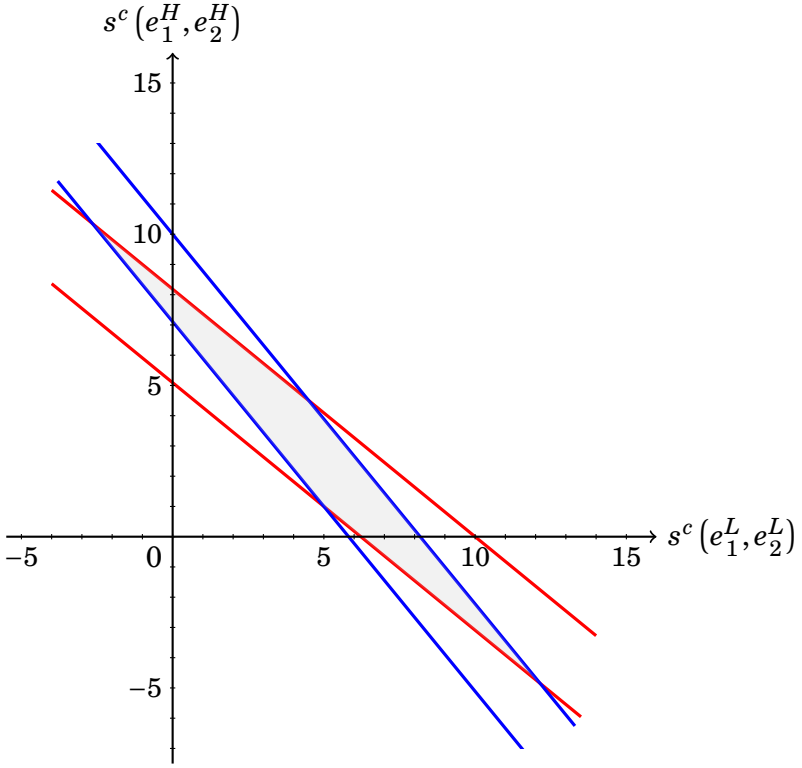
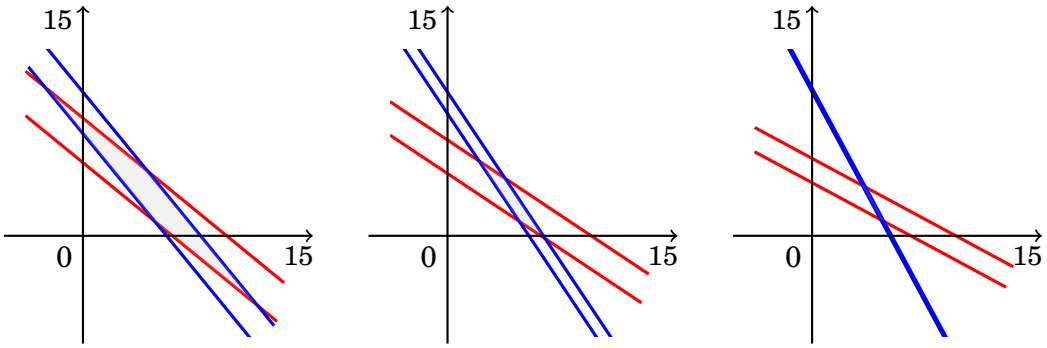


Figure 1: Combinations of payments that provide sustained cooperation under Assumptions 2, 3 and 4, for parameter values $\delta = 0.9$, $p = 0.5$, $\left(e_1^L, e_2^L\right) = (3,1)$, $\left(e_1^H, e_2^H\right) = (5,3)$, and $b_i(x_i) = -x_i^2 + 10x_i$ for $i = 1,2$.

This possibility of negative payments makes clear that, while theoretically sound, some sustainable payment rules may be inapplicable in real life situations.

Figure 2 shows a comparison of results for various values of the discount parameter δ. The figure illustrates that there is no feasible payment rule for low levels of the discount factor (for the parameter values in the figure, the threshold is $\delta = \frac{5}{7} \approx 0.7$). At the threshold, the bounds for the realization of low river flow first converge and below this threshold, they switch place and diverge such that there are no combinations of payments that provide sustained cooperation. The intuition for this result is standard in that a lower δ reduces the expected present value of the benefits of cooperation in all *future* periods, which are compared with the benefits of non-cooperation in the *current* period.

For reasons of exposition, the results in this section are constrained by two assumptions, but they can be easily generalized. Assumption 3 limits the example to two possible realizations of river flow. Having more possible realizations of river flow, say Low, Normal and High, would require three payments $s^c\left(e_1^L, e_2^L\right), s^c\left(e_1^N, e_2^N\right), s^c\left(e_1^H, e_2^H\right)$, and produce a three-dimensional figure while the analysis remains the same. Assumption 4



(a) $\delta = 0.9$ (b) $\delta = 0.8$ (c) $\delta = 0.7$

Figure 2: Panel (a) is identical to Figure 1, panels (b) and (c) differ only in the level of the discount factor δ.

limits the example to the set of efficient agreements. Its intuition, however, extends to inefficient agreements, which would yield similar diamond-shaped figures inside the shaded area but would require higher thresholds on the discount factor.

Summarizing, given a fixed parameter value for the discount factor and given trigger strategies, the illustration in Figure 1 shows how to construct sustainable agreements. Such agreements are not possible for a sufficiently low discount factor and some sustainable payment rules may be unrealistic if they include negative payments. Because it is not clear how such agreements could ever be put into practice, we assess several subsets of agreements that do not allow negative payments in the next section, while we will also drop Assumptions 3 and 4.

5 Four subsets of sustainable agreements

In this section, we assess four subsets of agreements that can be sustained in equilibrium. These subsets are fixed-payment agreements, individually-rational agreements, Nash-bargaining agreements, and renegotiation-proof agreements. Doing so, we do not need Assumptions 3 and 4. The reasoning for assessing these particular subsets is as follows. First, fixed-payment agreements reflect the lack of flexibility with respect to variability in river flow in most real-world agreements (De Stefano et al., 2012), and serve as a benchmark. Second, individually-rational agreements offer more flexibility and they exclude the unrealistic option of payoffs lower than minmax payoffs (including negative payments as discussed in Section 4). Third, Nash-bargaining agreements are assessed to illustrate how negotiations may lead to such agreements. Finally, renegotiation-proof agreements show how this additional stability requirement affects

our results.

Restricting the full set of agreements assessed in Section 4 comes at a cost. This cost is that the threshold discount factor for which agreements can be sustained will be (weakly) higher, because not all agreements are allowed. For fixed-payment agreements, Figure 3 shows that only the small subset of agreements on the 45°-line are allowed. For individually-rational agreements, Figure 3 shows that there are strict minimum and maximum bounds on the level of payments. This area contains also the subsets of Nash-bargaining agreements and renegotiation-proof agreements. A detailed explanation of these restrictions and the corresponding subsets of agreements is given in Sections 5.1–5.4.

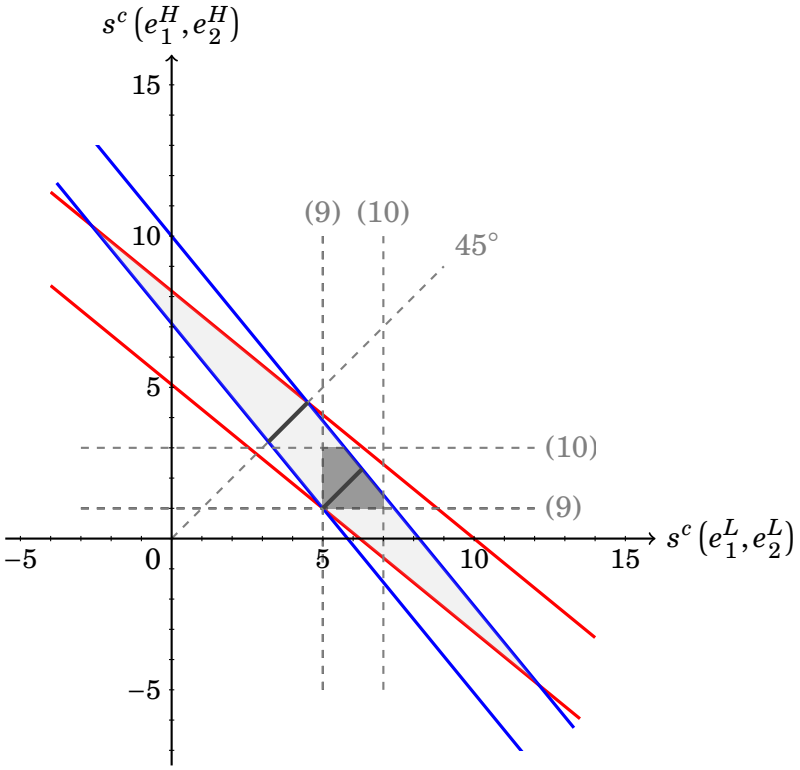


Figure 3: Identical to Figure 1, but including restrictions imposed by the subsets of agreements. The dark-shaded segment on the 45°-line through $(0,0)$ displays the subset of fixed payment agreements. The small shaded polygon displays the subset of individually rational agreements. The dark-shaded segment on the 45°-line through $(5,1)$ displays the subset of Nash-bargaining agreements that coincides with the subset of renegotiation-proof agreements.

5.1 Fixed-payment agreements

In this subsection we focus on agreements in which the payment rule is not contingent on the realization of river flow, but constant such that $s^c(e_1,e_2) = \bar{s}^c$. We call the subset of agreements that satisfies this property *fixed-payment agreements*. For such agreements, the equilibrium bounds (4) and (5) simplify to:

$$\begin{aligned}\bar{s}^c \;&\geq\; (1-\delta)\left[b_1(e_1) - b_1\left(x_1^c(e_1,e_2)\right)\right] \\ &+\; \delta\left[\boldsymbol{E}\left\{b_1(e_1)\right\} - \boldsymbol{E}\left\{b_1\left(x_1^c(e_1,e_2)\right)\right\}\right], \qquad (6)\\ \bar{s}^c \;&\leq\; \delta\left[\boldsymbol{E}\left\{b_2\left(x_2^c(e_1,e_2)\right)\right\} - \boldsymbol{E}\left\{b_2(e_2)\right\}\right]. \qquad (7)\end{aligned}$$

For sustained cooperation, these bounds have to hold for every realization (e_1,e_2). Consequently, the lower bound (6) becomes

$$\begin{aligned}\bar{s}^c \;&\geq\; (1-\delta)\max_{(e_1,e_2)}\left[b_1(e_1) - b_1\left(x_1^c(e_1,e_2)\right)\right] \\ &+\; \delta\left[\boldsymbol{E}\left\{b_1(e_1)\right\} - \boldsymbol{E}\left\{b_1\left(x_1^c(e_1,e_2)\right)\right\}\right]. \qquad (8)\end{aligned}$$

The interpretation of this lower bound is that the realization (e_1,e_2) where upstream is tempted most to deviate determines the lower bound. This observation corresponds to the analysis of stability (or robustness) of river sharing agreements used by Ansink and Ruijs (2008) and Ambec et al. (2013), the latter motivating their choice by referring to the literature on self-enforcing contracts (e.g. Gauthier et al., 1997). Given fixed payments, an agreement can be designed by first selecting the water allocation rule $x^c(e_1,e_2)$ which leaves a range of sustainable fixed payments $\bar{s}^c$ to choose from. It seems natural to select the efficient water allocation, in order to attain the maximal range of payments, but there are many examples where this is not the case (Giordano et al., 2013). The most common alternatives are fixed and proportional water allocations, and they can be assessed by simply substituting either $x_1^c(e_1,e_2) = e_1 - e_1^c$ and $x_2^c(e_1,e_2) = e_2 + e_1^c$ for fixed water allocation or $x_1^c(e_1,e_2) = \gamma e_1$ and $x_2^c(e_1,e_2) = e_2 + \gamma e_1$ for proportional water allocation. We will focus on the general case only.

The lack of flexibility of fixed-payment agreements may imply that, for a given discount factor δ and for the selected water allocation rule, there does not exist any sustainable payment rule. Existence requires a non-empty range of $\bar{s}^c$ that satisfy (7) and (8), which requires that the upper bound on $\bar{s}^c$ is larger than or equal to the lower

bound. We obtain for a given realization (e_1, e_2):

$$\delta\left[\boldsymbol{E}\left\{b_1\left(x_1^c(e_1,e_2)\right)+b_2\left(x_2^c(e_1,e_2)\right)-b_1(e_1)-b_2(e_2)\right\}\right] \geq (1-\delta)\left[b_1(e_1)-b_1\left(x_1^c(e_1,e_2)\right)\right].$$

By Assumption 1 on water scarcity, the left-hand side is positive for $x^c(e_1,e_2) = x^*(e_1,e_2)$ and these belong to a well-defined compact set of agreements $x^c(e_1,e_2)$ that admit a non-negative left-hand side. For the subset of agreements $x^c(e_1,e_2)$ that admit a positive left-hand side, as δ goes to 1, the left-hand side converges to some positive number while the right-hand side converges to 0, so that the inequality holds. Therefore, we know that for this particular subset of agreements $x^c(e_1,e_2)$ there exists a threshold discount factor above which the range of sustainable payment rules is non-empty. Obviously, agreements $x^c(e_1,e_2)$ for which the left-hand side is either zero or negative cannot be sustained for any $\delta \in [0,1)$.

This threshold, which has to hold for all realizations (e_1,e_2), is given by

$$\delta \geq \max_{(e_1,e_2)} \frac{b_1(e_1)-b_1\left(x_1^c(e_1,e_2)\right)}{b_1(e_1)-b_1\left(x_1^c(e_1,e_2)\right)+\boldsymbol{E}\left\{b_1\left(x_1^c(e_1,e_2)\right)+b_2\left(x_2^c(e_1,e_2)\right)-b_1(e_1)-b_2(e_2)\right\}},$$

and it is attained for $\max_{(e_1,e_2)}\left[b_1(e_1)-b_1\left(x_1^c(e_1,e_2)\right)\right]$, i.e. where the temptation to deviate is highest to agent 1. For the example in Section 4, this threshold for existence of a sustainable fixed-payment rule occurs for $(e_1,e_2) = \left(e_1^L, e_2^L\right) = (3,1)$ and it lies at $\delta = \frac{5}{7} \approx 0.7$, similar to the threshold for the general case in Section 4. Apparently, for this particular example, the limitation to fixed-payment agreements would not constrain the possibility of sustainable agreements, but this is not a general result.

On the one hand, one may regard fixed-payment agreements as an insurance contract where downstream pays a fixed amount for a flexible scheme of water deliveries. In this interpretation, there is nothing against such agreements. On the other hand, one problem of fixed-payment rules is that they may cause payoffs lower than minmax payoffs for some realizations of river flow. This property makes such rules unattractive for application in practice. We will see in the next subsection that, for the example of Section 4, any fixed-payment rule violates this condition.

5.2 Individually-rational agreements

In this subsection we focus on agreements that offer more flexibility than the fixed-payment agreements. Instead, we introduce a condition that excludes the unrealistic option of payoffs lower than minmax payoffs. This condition thereby also excludes the

possibility of negative payments that occurred in the example of Section 4. This condition implies[7]

$$b_1\left(x_1^c(e_1,e_2)\right)+s^c(e_1,e_2) \;\geq\; b_1(e_1) \;\text{ for any } (e_1,e_2); \tag{9}$$

$$b_2\left(x_2^c(e_1,e_2)\right)-s^c(e_1,e_2) \;>\; b_2(e_2) \;\text{ for any } (e_1,e_2). \tag{10}$$

Hence, in expectations

$$\boldsymbol{E}\left\{b_1\left(x_1^c(e_1,e_2)\right)+s^c(e_1,e_2)\right\} \;\geq\; \boldsymbol{E}\left\{b_1(e_1)\right\};$$

$$\boldsymbol{E}\left\{b_2\left(x_2^c(e_1,e_2)\right)-s^c(e_1,e_2)\right\} \;>\; \boldsymbol{E}\left\{b_2(e_2)\right\}.$$

We call the subset of agreements that satisfies (9) and (10) *individually-rational agreements*. Note that the sum of expected utilities under any agreement in this subset is larger than the sum of the expected utilities under noncooperation, i.e., $V_1^c+V_2^c > \boldsymbol{E}\{b_1(e_1)\}+\boldsymbol{E}\{b_2(e_2)\}$. This covers the entire triangle of individually-rational utility vectors under the Pareto frontier, neglecting the boundary $\left(V_1^c, b_2(e_2)\right)$.

For individually-rational agreements, we show in our next result that any agreement that improves upon the utilities under the minmax allocation $x^N(e_1,e_2)=(e_1,e_2)$ is an SPE for sufficiently large $\delta<1$. To do so, we first define the threshold level

$$\underline{\delta}(x_1^c,x_2^c,s^c)=\max_{(e_1,e_2)}\frac{s^c(e_1,e_2)}{s^c(e_1,e_2)+\boldsymbol{E}\left\{b_2\left(x_2^c(e_1,e_2)\right)-b_2(e_2)\right\}-\boldsymbol{E}\left\{s^c(e_1,e_2)\right\}}. \tag{11}$$

In the proof of the following result, we show that $\underline{\delta}(x_1^c,x_2^c,s^c)<1$. The following result can be interpreted as the Folk Theorem for river sharing problems.

Proposition 2. *For any $\delta\geq\underline{\delta}(x_1^c,x_2^c,s^c)$, the individually-rational agreement with allocation rule $\left(x_1^c(e_1,e_2),x_2^c(e_1,e_2)\right)$ and payment rule $s^c(e_1,e_2)$ satisfying* (9) *and* (10) *can be sustained in equilibrium.*

Proof. Conditions (2) and (3) state the equilibrium conditions for trigger strategies, which both depend upon the realization (e_1,e_2) and have to hold for every realization. Rewriting (2) yields

$$(1-\delta)\left[b_1\left(x_1^c(e_1,e_2)\right)+s^c(e_1,e_2)\right]+\delta\boldsymbol{E}\left\{b_1\left(x_1^c(e_1,e_2)\right)+s^c(e_1,e_2)-b_1(e_1)\right\}\geq(1-\delta)b_1(e_1).$$

By (9) the first term on the left-hand side is weakly larger than the right-hand side,

[7] In Footnote 5, we mentioned that, in general, boundary payoff vectors in repeated games are hard to sustain. Here, with trigger strategies, we can sustain $\left(b_1(e_1),V_2^c\right)$ for all realizations, but not $\left(V_1^c,b_2(e_2)\right)$.

and the second term on the left-hand side is non-negative. Therefore this inequality holds for all $\delta \in [0,1]$ and independent of the realization (e_1, e_2). So, agent 1 will not deviate for any $\delta \in [0,1]$, independent of the realization of river flow. Next, (3) can be simplified to $s^c(e_1,e_2) \leq \delta\left(V_2^c - V_2^n\right)$.[8] Rewriting further yields

$$(1-\delta)s^c(e_1,e_2) \;\leq\; \delta\left(\boldsymbol{E}\left\{b_2\left(x_2^c(e_1,e_2)\right) - b_2(e_2)\right\} - \boldsymbol{E}\left\{s^c(e_1,e_2)\right\}\right).$$

By (10), the right-hand side is positive. Therefore, as δ goes to 1, the left-hand side goes to 0 and the right-hand side increases to some positive number. So, there exists a nonempty range of δ close to 1 for which this inequality holds. Solving for δ yields that agent 2 has no incentive to deviate if:

$$\delta \;\geq\; \frac{s^c(e_1,e_2)}{s^c(e_1,e_2) + \boldsymbol{E}\left\{b_2\left(x_2^c(e_1,e_2)\right) - b_2(e_2)\right\} - \boldsymbol{E}\left\{s^c(e_1,e_2)\right\}}.$$

As noted in Section 3, this threshold depends on the realization of river flow. Also, it depends upon the allocation rule $x^c(e_1,e_2)$ and payment rule $s^c(e_1,e_2)$. Because this has to hold for every realization (e_1,e_2), we must have that $\delta \geq \underline{\delta}(x_1^c, x_2^c, s^c)$. □

Note that (9) implies that $s^c(e_1,e_2) \geq b_1(e_1) - b_1\left(x_1^c(e_1,e_2)\right) \geq 0$. This class of agreements excludes the negative s^c that we observed in the example of Figure 1. Also interesting, the simplification of (3) to $s^c(e_1,e_2) \leq \delta\left(V_2^c - V_2^n\right)$ in the proof of Proposition 2, imposes, once more, a fixed upper bound on the payment, independent of realization (e_1,e_2). In the example of Section 4, however, this upper bound is dominated by the upper bounds given by (10). Specifically, for individually-rational agreements, the restrictions given by (9) and (10) on the (efficient) example in Figure 1 are:

$$\begin{aligned}
s^c\left(e_1^L, e_2^L\right) &\geq b_1(e_1=3) - b_1(x_1^c=2) \;=\; 21-16 \;=\; 5;\\
s^c\left(e_1^H, e_2^H\right) &\geq b_1(e_1=5) - b_1(x_1^c=4) \;=\; 25-24 \;=\; 1;\\
s^c\left(e_1^L, e_2^L\right) &\leq b_1(x_2^c=2) - b_1(e_2=1) \;=\; 16-9 \;=\; 7;\\
s^c\left(e_1^H, e_2^H\right) &\leq b_1(x_2^c=4) - b_1(e_2=3) \;=\; 24-21 \;=\; 3.
\end{aligned}$$

These restrictions are shown in the dark-shaded polygon in Figure 3.

Proposition 2 and its proof convey two important messages. The first is that, by construction of individually-rational agreements, agent 1 has no incentive to deviate,

[8] For $V_2^c = V_2^n$, only $s^c(e_1,e_2) = 0$ would be feasible. So, without compensating the upstream agent in the future, cooperation is impossible.

irrespective of the discount factor nor the realization of river flow. This is an important observation for the design of stable agreements. The second message is that the threshold discount factor $\underline{\delta}(x_1^c, x_2^c, s^c)$ increases in the payment that agent 2 makes to agent 1. Consequentially, the scope for agent 2's compliance with the agreement increases when the payments, contingent on the realized river flow and subject to (9) and (10), are minimized. Specifically, it is easy to check that the lowest threshold $\underline{\delta}(x_1^c, x_2^c, s^c)$ occurs for the agreement $\left(x_1^c, x_2^c, s^c\right)$ that solves

$$\min_{x_1^c, x_2^c, s^c(e_1,e_2)} \max \frac{s^c(e_1,e_2)}{s^c(e_1,e_2) + \boldsymbol{E}\left\{b_2\left(x_2^c(e_1,e_2)\right) - b_2(e_2)\right\} - \boldsymbol{E}\left\{s^c(e_1,e_2)\right\}}$$

among all possible individually-rational agreements. The solution to this problem is $x^c(e_1,e_2) = x^*(e_1,e_2)$ and $s^c(e_1,e_2) = b_1(e_1) - b_1\left(x_1^*(e_1,e_2)\right)$ for every realization (e_1,e_2). This solution selects the efficient allocation and a payment that assigns all benefits of cooperation downstream, to agent 2. We will continue discussing this solution in the context of asymmetric Nash-bargaining solutions in Section 5.3.

Note that the subset of individually-rational agreements contains various types of agreements. One example is a price-dependent agreement, in which the allocation rule is the efficient allocation and the payment implements the efficient water price, such that marginal benefits of water use are equal to both agents. This type of agreement mimics an international water market (cf. Ansink and Houba, 2012). An alternative example is the (asymmetric) Nash-bargaining solution, which we analyze in the next subsection.

As a final remark, recall the interpretation of fixed-payment agreements as insurance contracts. Individually-rational agreements can be also be seen as insurance contracts with a stochastic price $s^c(e_1,e_2)$ that depends upon the realization of river flow. This means that next to the risk over the allocation of water there is also risk with respect to this price. Because monetary payments enter the agents' quasi-linear utility functions in (1) as the linear term, agents are risk neutral with respect to this category of risk. Consequently, they are indifferent between individually-rational agreements and fixed-payment contracts with the same allocation and payment $\boldsymbol{E}\{s^c(e_1,e_2)\}$.

5.3 Nash-bargaining agreements

In this subsection we focus on agreements that are the result of a negotiation process. The obvious solution concept is then to look at asymmetric Nash-bargaining solutions (ANBS). The ANBS maximizes the product of agents' gains over a disagreement payoff,

given asymmetric bargaining strengths. Applied to the problem of river sharing in a deterministic setting, the allocation rule is the efficient allocation and the payment is based on the relative bargaining strength of the agents (Houba et al., 2013). In our stochastic setting, the ANBS will also specify the efficient allocation and the *expected* payment based on the relative bargaining strength of the agents. There are many ways to implement the expected payment over all possible realizations (e_1, e_2) and we choose a natural one: Applying the ANBS with bargaining weight $\alpha \in [0,1)$ for agent 1,[9] we take

$$s^{\alpha}(e_1,e_2) = b_1(e_1) - b_1\left(x_1^*(e_1,e_2)\right) + \alpha\left[b_1\left(x_1^*(e_1,e_2)\right) + b_2\left(x_2^*(e_1,e_2)\right) - b_1(e_1) - b_2(e_2)\right], \tag{12}$$

for each realization of river flow (e_1, e_2), which then allows us to derive $\boldsymbol{E}\{s^{\alpha}(e_1,e_2)\}$. This implementation of the ANBS chooses the efficient allocation and distributes the gains of cooperation according to the bargaining strength parameter α.

The following corollary follows from the equivalence of the expression for $s^{\alpha}(e_1,e_2)$ in (12) with conditions (9) and (10) on the subset of agreements with efficient water allocation $x^*(e_1,e_2)$.[10]

Corollary 1. *Any Nash-bargaining agreement is an efficient individually rational agreement and vice versa.*

This result implies that any efficient individually rational agreements can be implemented by a Nash-bargaining agreement for some $\alpha \in [0,1)$, and vice versa.

By Corollary 1, since the example in Section 4 assumed efficient allocations, the small shaded polygon in Figure 3 includes the set of Nash-bargaining agreements. Specifically, substituting the parameters from the example in (12), we obtain a set of Nash-bargaining agreements with $s^{\alpha}(e_1^L, e_2^L) = 5 + 2\alpha$ and $s^{\alpha}(e_1^H, e_2^H) = 1 + 2\alpha$. In Figure 3, this set lies on a segment through $(5,1)$ with slope 45° due to the equal probability of low and high river flow ($p = 0.5$).

For Nash-bargaining agreements, given weight α, we show in our next result that any efficient agreement that improves upon the utilities under the minmax allocation $x^N(e_1,e_2) = (e_1,e_2)$ is an SPE for sufficiently large $\delta < 1$. The difference with Proposition 2 is subtle. By construction, the ANBS as applied in (12) satisfies conditions

[9] For similar reasons as in (9) and (10), we allow for $\alpha = 0$ i.e. agent 2 is a dictator, but not for $\alpha = 1$.

[10] To see this, note the following: First, for $\alpha = 0$, (12) is equivalent to (9) with equality, while trivially satisfying (10). Second, for the full range of $\alpha \in (0,1)$, (12) is equivalent to the combination of (9) and (10) with strict inequalities.

(9) and (10) that describe the set of individually-rational agreements and, therefore, Nash-bargaining agreements are a subset of this set. The Pareto efficiency of Nash-bargaining agreements implies that the threshold discount factor for which an agreement can be sustained will coincide with the one for the associated Pareto efficient individually-rational agreement.

We show in our next result that any Nash-bargaining agreement is an SPE for sufficiently large $\delta < 1$. To do so, we first define the threshold level

$$\underline{\delta}^{\alpha}(x_1^*, x_2^*) = \max_{(e_1,e_2)} \frac{s^{\alpha}(e_1,e_2)}{s^{\alpha}(e_1,e_2) + \boldsymbol{E}\left\{b_2\left(x_2^*(e_1,e_2)\right) - b_2(e_2)\right\} - \boldsymbol{E}\left\{s^{\alpha}(e_1,e_2)\right\}}. \tag{13}$$

Proposition 3. *For any $\alpha \in [0,1)$ and $\delta \geq \underline{\delta}^{\alpha}(x_1^*, x_2^*)$, the Nash-bargaining agreement with allocation rule $x^*(e_1,e_2)$ and payment rule $s^{\alpha}(e_1,e_2)$ satisfying* (12) *can be sustained in equilibrium.*

Proof. The proof is similar to the proof of Proposition 2, with two small differences. One is that, without implications, (12) is used rather than (9) and (10). The second difference is that $x^c(e_1,e_2)$ and $s^c(e_1,e_2)$ are replaced by $x^*(e_1,e_2)$ and $s^{\alpha}(e_1,e_2)$, following the definition of ANBS and (12).

Hence, we know that agent 1 will not deviate for any $\delta \in [0,1]$, and we know that there exists a nonempty range of δ for which agent 2 will also not deviate, which is given by

$$\delta \geq \frac{s^{\alpha}(e_1,e_2)}{s^{\alpha}(e_1,e_2) + \boldsymbol{E}\left\{b_2\left(x_2^*(e_1,e_2)\right) - b_2(e_2)\right\} - \boldsymbol{E}\left\{s^{\alpha}(e_1,e_2)\right\}}.$$

Finally, because this has to hold for every realization (e_1,e_2), we must have that $\delta \geq \underline{\delta}^{\alpha}(x_1^*, x_2^*)$. □

From (11), we already know that $\underline{\delta}(x_1^c, x_2^c, s^c)$ is increasing in the payment scheme $s^c(e_1,e_2)$. Because the payment $s^{\alpha}(e_1,e_2)$ in (12) is increasing in α, we immediately obtain that $\underline{\delta}^{\alpha}(x_1^*, x_2^*)$ in (13) is also increasing in α. The interpretation of this relation is that the threshold discount factor for which Nash-bargaining agreements can be sustained is increasing in the bargaining strength of agent 1. In other words, the scope for sustainable Nash-bargaining agreements decreases when the upstream agent gets a larger share of the pie. Likewise, this scope increases when the downstream agent gains most.

As is standard in the literature on repeated games, we have presented Proposition 3 in terms of a threshold discount factor, above which subsets of agreements can be sus-

tained in equilibrium. For the practical purpose of this paper, the real issue is, given some δ, how to construct sustainable agreements, which in this case means how to construct the range of sustainable Nash bargaining agreements. To do so, we first define the upper bound

$$\alpha^{\delta}(x_1^*, x_2^*) \equiv \min_{(e_1,e_2)} \frac{\delta \boldsymbol{E}\{h(e_1,e_2)\} - (1-\delta)\left[b_1(e_1) - b_1\left(x_1^*(e_1,e_2)\right)\right]}{\delta \boldsymbol{E}\{h(e_1,e_2)\} + (1-\delta)[h(e_1,e_2)]} < 1.$$

We now present this alternative, more practical, interpretation to Proposition 3.

Proposition 4. *For any $\delta \in [0,1)$ and $\alpha \leq \alpha^{\delta}(x_1^*, x_2^*)$, the Nash-bargaining agreement with allocation rule $\left(x_1^*(e_1,e_2), x_2^*(e_1,e_2)\right)$ and payment rule $s^{\alpha}(e_1,e_2)$ satisfying* (12) *can be sustained in equilibrium.*

Proof. The proof consists of rewriting the equilibrium condition from Proposition 3, using (12) and (13). Denote the cooperative surplus by $h(e_1,e_2) = b_1\left(x_1^*(e_1,e_2)\right) + b_2\left(x_2^*(e_1,e_2)\right) - b_1(e_1) - b_2(e_2)$, then rewriting and substituting for $s^{\alpha}(e_1,e_2)$ we obtain

$$\alpha \leq \frac{\delta \boldsymbol{E}\{h(e_1,e_2)\} - (1-\delta)\left[b_1(e_1) - b_1\left(x_1^*(e_1,e_2)\right)\right]}{\delta \boldsymbol{E}\{h(e_1,e_2)\} + (1-\delta)[h(e_1,e_2)]} < 1.$$

Because this inequality has to hold for every realization (e_1,e_2), we must have that $\alpha \leq \alpha^{\delta}(x_1^*, x_2^*)$. □

This result, which is the converse of Proposition 3, shows how a given discount factor restricts the subset of sustainable Nash bargaining agreements. Note that $\alpha^{\delta}(x_1^*, x_2^*)$ is increasing in δ. So, as agents become more patient a larger subset of Nash bargaining agreements can be sustained. Moreover, $\alpha^{\delta}(x_1^*, x_2^*)$ converges to 1.[11] Therefore, when both agents become perfectly patient this allows the complete range of $\alpha \in [0,1)$.

We further illustrate Nash-bargaining agreements in Figure 4, using the example of Section 4 for two values of α. Given transferable utility, the Pareto frontier is linear and given by the set of efficient agreements where total utility equals

$$2\left[p \cdot b_i\left(x_i^c\left(e_1^H, e_2^H\right)\right) + (1-p) \cdot b_i\left(x_i^c\left(e_1^L, e_2^L\right)\right)\right].$$

Using the parameter values of Figure 1, but without specifying p and δ, total utility equals $2[p \cdot 24 + (1-p) \cdot 16] = 32 + 16p$ per period. The disagreement point for agent 1 equals $21 + 4p$ per period and the disagreement point for agent 2 equals $9 + 12p$ per period.

[11] Formally, from substituting $\delta = 1$ in $\alpha^{\delta}(x_1^*, x_2^*)$ we obtain $\lim_{\delta \to 1} \alpha^{\delta}(x_1^*, x_2^*) = \frac{\boldsymbol{E}\{h(e_1,e_2)\}}{\boldsymbol{E}\{h(e_1,e_2)\}} = 1$.

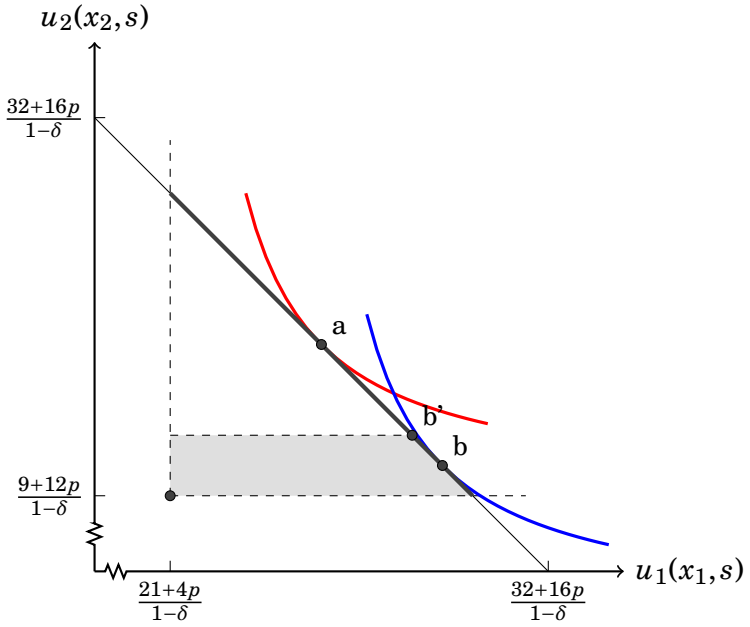


Figure 4: The ANBS for $\alpha = 0.5$ (point a where the red level curve touches the Pareto frontier) and $\alpha = 0.9$ (point b where the blue level curve touches the Pareto frontier), given $p = 0.5$. For sufficiently low δ part of the individually rational utility vectors under the Pareto frontier cannot be sustained in equilibrium because agent 2 will deviate (shaded area). This may shift the ANBS for agreements with high α, as indicated by point b' for $\alpha = 0.9$.

For $\alpha = 0.5$ the (symmetric) ANBS is the unconstrained optimum in point a of Figure 4, located at the intersection of the Pareto frontier and a 45° line through the disagreement point. For $\alpha = 0.9$ the ANBS may be constrained if δ is sufficiently low as indicated by the shaded area in the figure. This shaded area relates to the result in Proposition 3 that agent 2 deviates for $\delta < \underline{\delta}^{\alpha}(x_1^*, x_2^*)$, indicating that the expected utility of the cooperative path to agent 2 is not sufficiently high to prevent deviation for low δ. Hence, when agent 2 is sufficiently impatient, agreements in the shaded area cannot be sustained and only the largest α' such that $\underline{\delta}^{\alpha'}(x_1^*, x_2^*) \leq \delta$ can be implemented. The interpretation of this situation is that the bargaining strength of upstream agent 1 may be limited by the absence of a supra-national authority that can enforce agreements. For $\alpha = 0.9$, this implies that the ANBS at point b shifts to the corner solution at point b', where the agreement can be sustained. Agent 2's impatience yields him a higher payoff at the cost of agent 1.

This impact of impatience illustrates the (lack of) *robustness* of Nash-bargaining agreements. Because agent 1 has no incentive to deviate, the most robust ANBS occurs for $\alpha = 0$ in the sense that this solution brings about the lowest threshold level

$\underline{\delta}^{\alpha}(x_1^c, x_2^c)$ and is always contained in the sustainable range of Nash bargaining agreements. By (12), when $\alpha = 0$, agent 1 is only compensated for his forgone benefits but does not share in the surplus generated by the agreement. This extreme solution in terms of the distribution of the surplus of cooperation coincides with the downstream incremental distribution, proposed by Ambec and Sprumont (2002) as a compromise between two legal doctrines for river sharing. This solution is assessed for robustness in a static setting by Ambec et al. (2013), and assessed for time-consistency in a dynamic cooperative game by Beard and McDonald (2007).

In the next subsection we will see that the requirement of renegotiation-proofness adds even more credibility to this extreme solution.

5.4 Renegotiation-proof equilibria

In this subsection we assess the implications of requiring agreements to satisfy renegotiation-proofness. Up till here, we assumed trigger strategies, but these strategies have an important disadvantage: The agent who carries out the punishment by switching to non-cooperative play is also punishing himself. This gives the punisher an incentive to abolish his punishment, and re-negotiate with the defector in order to revert to cooperative play, which Pareto-dominates the non-cooperative path. As a result, punishments by trigger strategies lack credibility.

In response to this lack of credibility, the key idea of renegotiation-proof equilibria (RPE) is to construct strategy profiles with punishments that include a non-negative reward to the punisher. Based on the concept of *weakly renegotiation-proof equilibrium* proposed by Farrell and Maskin (1989), our following result provides additional support for solutions in which the downstream agent gains most. Appendix B contains additional background information on RPE as well as an extensive proof of Proposition 5.

We will first summarize the results of Appendix B before stating our main results. Since we are interested in Pareto efficient solutions, we focus, without loss of generality, on sustaining Nash bargaining agreements in RPE. In order to sustain the Nash bargaining agreement $\alpha^0 \in [0,1)$, we introduce two punishment paths, one for each agent. The punishment path for agent 1 corresponds to the Nash bargaining agreement $\alpha^1 = 0$ in every period, a path on which agent 1 has no incentive to deviate. The punishment path for agent 2 is somewhat more involved. In the first period of this path, agent 1 does not deliver any water and agent 2 pays a penalty that is equal to the present value of the entire expected net surplus of all future periods from the second period onwards,

which is independent of the current realization. From the second period onward, we switch to the Nash bargaining agreement $\alpha^2 = 0$ in every period, which equals agent 1's punishment path (we postpone an explanation). For the moment, agent 2's punishment path coincides with agent 1's best RPE path. Therefore, agent 2's punishment path is Pareto efficient from the second period onwards with unavoidable efficiency losses only in the first period.

In our next result, we state the theoretically largest set of Pareto efficient RPE Nash bargaining agreements that are derived in Appendix B.

Proposition 5. *For any $\alpha^0 \in [0,1)$ and $\delta \geq \underline{\delta}^{\alpha^0}(x_1^*, x_2^*)$, the Nash bargaining agreement with allocation rule $\left(x_1^*(e_1,e_2), x_2^*(e_1,e_2)\right)$ and payment rule $s^{\alpha^0}(e_1,e_2)$ satisfying* (12) *can be sustained in weakly renegotiation-proof equilibrium.*

Corollary 2. *For any $\delta \in [0,1)$ and $\alpha^0 \leq \alpha^\delta(x_1^*, x_2^*)$, the Nash bargaining agreement with allocation rule $\left(x_1^*(e_1,e_2), x_2^*(e_1,e_2)\right)$ and payment rule $s^{\alpha^0}(e_1,e_2)$ satisfying* (12) *can be sustained in weakly renegotiation-proof equilibrium.*

The positive news is that weakly renegotiation-proof equilibria exist and that the underlying strategies can replace trigger strategies. The most striking part of this result is that imposing a more restrictive equilibrium concept does not result in a higher threshold level for the discount factor or a reduced upper bound on agents 1's bargaining weight. In Appendix B we show that agent 2 obtains exactly his minmax payoff in his punishment path. Since this is the same payoff as under the trigger strategies, both strategies employed as punishment strategies are payoff-equivalent to agent 2. As before, only agent 2 has an incentive to deviate. Given the payoff-equivalent punishments to agent 2, this must result in the same thresholds as derived under trigger strategies.

In Appendix B we also derive that $\underline{\delta}^\alpha(x_1^*, x_2^*)$ is increasing in α so that the threshold discount factor for which RPE can be sustained is increasing in the payment $s^\alpha(e_1,e_2)$. Analogous to the interpretation of Proposition 3 on Nash-bargaining agreements, the most robust agreement occurs for $\alpha^0 = 0$ in the sense that this solution brings about the lowest threshold level $\underline{\delta}^\alpha(x_1^*, x_2^*)$ and thereby maximizes the scope for sustainable agreements. As mentioned in the previous section, for $\alpha = 0$ agent 1 is only compensated for his forgone benefits, while agent 2 receives the complete surplus of cooperation.

Remark 2. We close this section with several remarks on the implications of the penalty in player 2's punishment path. Note that agent 2's worst RPE payoff consists of his expected payoff $\mathbf{E}\{b_2(e_2)\}$ minus a penalty in the first period, which is below his minmax value, followed by his expected maximum RPE payoff in all future periods, which

is above his minmax value. Although we objected against payoffs below the minmax value in the last paragraph of Section 5.1, this argument does not apply here because, by paying the penalty, agent 2 invests in restoring the cooperation, which might be interpreted as either to repent and show remorse, or to regain agent 1's trust.

Furthermore, the penalty equals the present value of the entire expected net surplus of all future periods from the second period onwards, which can be quite substantial. Even though agent 1 receives his minmax payoff in all future periods from the second period onwards, this agent receives almost the entire present value of the overall net surplus as the penalty in the first period (he only misses out on the net surplus of the first period, in which he only receives $\mathbf{E}\{b_1(e_1)\}$). Theoretically, such a substantial penalty is fine, but in practice it might not be realistic.

For practical purposes, one might resort to weights $\alpha^1 < \alpha^0 < \alpha^2$ that are closer together and construct paths similar as described above with one exception: the punishment path for agent 2 specifies (e_1, e_2) and some small penalty in the first period, followed by the Nash bargaining agreement α^2 for several periods before it continues forever with Nash bargaining agreement α^0 (instead of α^1).[12] Then, as before, agent 1 has no incentive to deviate and agent 2 should be given enough incentives to undergo any of these three paths. We leave this option for future research.

The main message of this section is that Pareto efficient weakly renegotiation-proof equilibria can be derived and that we characterized its theoretically largest set, which happens to coincide with the set of sustainable Nash bargaining agreements in SPE.

6 Conclusion

This paper is the first to systematically assess the implications of repeated interaction for the stability of river sharing agreements between riparian neighbors. Our Folk Theorem for river sharing problems in Proposition 2, further refined in Propositions 3 and 5, provides clear conditions for sustainable agreements in terms of the distribution of the gains from cooperation. These conditions are stated in terms of minimal thresholds on the discount factor. For the practical purpose of this paper, the real issue is, given some δ, how to construct sustainable agreements. The relevance of this issue is illustrated in the example of Section 4 and further demonstrated in Proposition 4 and Corollary 2 where ranges of sustainable Nash bargaining agreements were

[12] This construction mimics the Pareto efficient RPE in Van Damme (1989) for the infinitely repeated Prisoners' Dilemma, in which both agents return to cooperate forever after undergoing their punishment for several periods, in each agent's punishment.

derived. Remarkable is that the set of sustainable Nash bargaining agreements under subgame perfect equilibrium can also be sustained by the more restrictive weakly renegotiation-proof equilibrium.

Repeated interaction tends to favor the downstream agent, which may seem counter-intuitive at first, but may explain empirical observations on downstream states managing to negotiate a substantial share of upstream river water. Our results provide non-cooperative support for solutions that assign larger shares of the pie to downstream agents. At the lowest possible threshold on the discount factor, only the downstream incremental distribution, proposed by Ambec and Sprumont (2002), that assigns *all* gains from cooperation to downstream agents can be sustained and this distribution remains sustainable for higher discount factors.

Finally, the model developed in this paper offers ample scope for extensions and applications. One obvious extension is to allow for more general river geographies as in Khmelnitskaya (2010), Ansink and Houba (2012), or Van den Brink et al. (2012). One obvious application is to repeat the analysis of the Bishkek Treaty in the Aral Sea basin by Ambec et al. (2013) in the dynamic setting of this paper. Their static analysis showed that actual payments under this agreement approximate the payment rule induced by the downstream incremental distribution. In a static setting, this result implies instability. When considering repeated interaction, however, the results in our paper suggest that this payment rule may actually be well-chosen.

A On minmax values

In this appendix, we derive the minmax values for each agent which proves Proposition 1, discuss the implications of assuming increasing benefit functions and describe how to derive minmax values in the general case with more than two agents.

A.1 Proof of Proposition 1

Since the infinitely-repeated river sharing problem is an infinitely-repeated sequential game, the characterization of minmax values in Wen (2002) is appropriate. The order of moves is trivial in the river sharing problem: agent 1 moves before agent 2. Each minmax value is solved recursively and backward.

First, for any realization (e_1, e_2), agent 2's best response to any strategy $x_1(e_1, e_2)$ by agent 1 is $x_2 = e_1 + e_2 - x_1(e_1, e_2)$ due to the increasing benefit function. The strategy of agent 1 that minmaxes agent 2 is therefore given by $\min_{x_1 \in [0, e_1]} b_2(e_1 + e_2 - x_1)$. This implies $x_1(e_1, e_2) = e_1$. So, $b_2(e_2)$ is agent 2's minmax value and it is equal to

$$\min_{x_1 \in [0, e_1]} \max_{x_2 \in [0, e_1 + e_2 - x_1]} b_2(x_2). \tag{14}$$

Second, for any realization (e_1, e_2) and agent 1's strategy x_1, any strategy $x_2 \in [0, e_1 + e_2 - x_1]$ does not affect agent 1's benefit function $b_1(x_1)$, which implies that agent 2 can not punish agent 1. So, in deriving agent 1's minmax value agent 1 maximizes first over x_1 taking into account the minmax response of agent 2. Formally,

$$\max_{x_1 \in [0, e_1]} \min_{x_2 \in [0, e_1 + e_2 - x_1]} b_1(x_1), \tag{15}$$

which yields minmax value $b_1(e_1)$ because b_1 is increasing. This completes the derivation of minmax values.

The implication for our analysis of the infinitely-repeated river sharing game is that forever playing the Nash equilibrium is not only the worst punishment for both agents but it is also a credible punishment in terms of the SPE.

A.2 Remark on increasing benefit functions

In the above derivation of minmax values, we invoked that the benefit functions are increasing. We will assess the importance of this assumption for our main results by determining the minmax values when we allow for satiation. Recall that $[\underline{e}_1, \overline{e}_1] \times [\underline{e}_2, \overline{e}_2]$ is the domain of the probability distribution. Denote x_i^S as agent i's satiation point. By

strict concavity, benefit function b_i increases on the interval $\left[0, x_i^S\right)$ and decreases for $x_i > x_i^S$. We will not investigate all possible cases, but rather concentrate on the case $x_1^S < \underline{e}_1$ and $x_2^S > \overline{e}_2$ for explanatory simplicity.[13] Assumption 1 on water scarcity further imposes $x_1^*(e_1, e_2) < x_1^S$ for all realizations (e_1, e_2). Then, for any realization (e_1, e_2), the unique Nash equilibrium features $\hat{x}_1^N(e_1, e_2) = x_1^S$ and

$$\hat{x}_2^N(e_1, e_2) = \min\left\{x_2^S, e_1 + e_2 - x_1^S\right\} = e_1 + e_2 - x_1^S > e_2.$$

Obviously, $x_2^*(e_1, e_2) > \hat{x}_2^N(e_1, e_2)$.

First, in order to derive agent 1's minmax value we again apply (15) and obtain that this agent's minmax value is $b_1\left(x_1^S\right) > b_1(e_1)$. Similar to the case without satiation, agent 1's minmax value is supported by the Nash equilibrium and only quantitatively we have to deal with the difference $b_1\left(x_1^S\right) > b_1(e_1)$. The implication for our analysis of the infinitely-repeated river sharing game is that forever playing the Nash equilibrium remains available as the worst punishment for agent 1 that is also credible.

Second, agent 2's minmax value is again derived from (14) and we obtain that

$$x_1 = e_1 > x_1^S \quad \text{and} \quad x_2 = \min\left\{x_2^S, e_2\right\} = e_2 < \hat{x}_2^N(e_1, e_2)$$

support this agent's minmax value $b_2(e_2) < b_2\left(\hat{x}_2^N(e_1, e_2)\right)$. So, agent 2's minmax value remains similar to the case without satiation, but it is no longer supported by the Nash equilibrium. In punishing agent 2, agent 1 incurs opportunity costs $b_1(e_1) - b_1\left(x_1^S\right) > 0$ due to overconsumption. Care should be taken in designing punishments if agent 1 would choose not to to minmax agent 2. This can be seen as a standard exercise in repeated games that we forgo.

Summarizing, the restriction to increasing benefit functions is technically convenient because it avoids some technicalities in sustaining credible punishments and it allows to consider the simple class of trigger strategies in sustaining SPE. Increasing benefit functions are also notationally convenient, because all minmax values become $b_i(e_i)$.

[13] Note that for $x_1^S \in [\underline{e}_1, \overline{e}_1]$, all realizations $e_1 \leq x_1^S$ are equivalent to non-satiation and, similar, all $e_1 > x_1^S$ correspond to satiation. For $x_2^S \in [\underline{e}_2, \overline{e}_2]$, all realizations e_2 such that $e_1 + e_2 - x_1^S \leq x_2^S$ are equivalent to non-satiation and otherwise we have satiation and no incentives for agent 2 to cooperate, which is trivial. For completeness, $x_1^S > \overline{e}_1$ corresponds to non-satiation and $x_2^S < \underline{e}_2$ implies agent 2 also has no incentives to cooperate, violating Assumption 1 on water scarcity.

A.3 Remark on the general case

Finally, we discuss the minmax values for the general case with n agents and more realistic river geographies (Ansink and Houba, 2012; Van den Brink et al., 2012). Then, applying the characterization of minmax values in Wen (2002) to river sharing problems implies that only upstream agents can minimize some agent's payoff while all downstream agents (and agents on disjoint tributaries) cannot affect this agent. Consider some agent $i = 1,\ldots,n$, denote the set of *all* of agent i's upstream agents as P^i, and denote the vector of these agents' water uses as $\left(x_j\right)_{j\in P^i}$. Then, (14) becomes

$$\min_{(x_j)_{j\in P^i}:\Sigma_{j\in P^i}x_j\leq\Sigma_{j\in P^i}e_j} \quad \max_{x_i\in[0,e_i+\Sigma_{j\in P^i}(e_j-x_j)]} b_i(x_i).$$

Note that for any most-upstream agent i that cannot receive inflow from any other agent, the set $P^i = \emptyset$ and, consequently, this agent's minmax value is his Nash equilibrium benefit $b_i\left(x_i^N(e_1,\ldots,e_n)\right)$, where $x_i^N(e_1,\ldots,e_n) = e_i$ denotes agent i's Nash equilibrium water use. Again, the worst punishment for the most-upstream agents is the Nash equilibrium, which is also credible. For all other agents, under mild assumptions, their minmax values are $b_i(e_i) \leq b_i\left(x_i^N(e_1,\ldots,e_n)\right)$. To be specific, in case agent i's satiation water use x_i^S is larger than the upper bound of his own stochastic resources, denoted $\overline{e}_i$, then agent i's minmax value is $b_i(e_i)$, for similar reasons as before.

Summarizing, the restriction to two agents is, from a conceptual point of view, qualitatively similar to the case of more than two agents while it requires less notation and is more insightful.

B On renegotiation-proof equilibria

In this appendix, we derive the subset of weakly renegotiation-proof equilibria (RPE), as defined in Farrell and Maskin (1989), that are also Pareto efficient. Before doing so, we first summarize some key ideas from the literature on repeated normal-form games that we need to modify in order to derive our results.

B.1 Repeated normal-form games and RPE

For standard infinitely-repeated games, Abreu (1988) showed that in n-player games it suffices to consider $n+1$ infinite paths of actions in the stage game. These paths are often denoted as $\pi^0, \pi^1, \ldots, \pi^n$, a notation that we will follow. Path π^0 represents the intended (subgame-perfect) equilibrium path and path π^i, $i = 1, \ldots, n$, represents player i's worst (subgame-perfect) equilibrium path. Sustaining these $n+1$ paths in equilibrium is based upon the following key ideas: First, the worst equilibrium path can be used as a credible punishment to sustain any equilibrium path π^0. If the worst equilibrium path cannot sustain π^0 as an equilibrium, then there does not exist any equilibrium punishment strategy that can sustain π^0. Second, if player i deviates from any of these paths, including π^i, then all players immediately switch to playing path π^i, an idea called equilibrium switching. So, if player i does not comply to π^i, π^i will be started over and over again. Of course, in equilibrium player i is given sufficient incentives to follow π^i and such non-compliance will not occur.

Trigger strategies, which we have used so far in the main text have a very simple structure, namely $\pi^1 = \ldots = \pi^n$ describe to always play the same Nash equilibrium. If this Nash equilibrium supports each player's minmax value then it becomes the credible worst punishment. Trigger strategies are criticized, however, because the player who carries out the punishment of a deviating player also hurts himself. For two-player games, this gives the punisher an incentive to forgive the defector and continue playing π^0 instead, but then π^0 becomes unsustainable because each player knows punishments, which are supposed to make π^0 stable, are never carried out.

In response to this criticism, the key idea of RPE is to construct punishments that include a non-negative reward to the punisher. A weakly renegotiation-proof equilibrium is an SPE that additionally requires that for all player j's, $j \neq i$, the equilibrium payoffs associated with π^i are larger or equal to player j's payoffs of following π^0. Also, along each path the equilibrium payoffs should be non-decreasing for reasons that we forgo. It gives each path a *stick and carrot* flavor.

B.2 Proof of Proposition 5

We first focus on characterizing the π^1 and π^2 (initially imposing $\pi^0 = \pi^1$) to sustain RPEs and then extend the analysis to characterize all Pareto efficient π^0 that can be supported by π^1 and π^2 as RPE. We denote agent i's minimum expected RPE payoff (before the realization of river flow) as m_i, and his maximum expected RPE payoff as M_i. We must derive these RPE payoffs in characterizing π^1 and π^2.

The results for sustainable individually-rational agreements in Section 5.2 suggest the following very convenient punishment path for agent 1 that gives agent 2 the highest expected payoff attainable in the set of individually-rational expected payoff vectors (V_1, V_2):

- π^1 : In every period, $x^1(e_1,e_2) = x^*(e_1,e_2)$ and $s^1(e_1,e_2) = b_1(e_1) - b_1\left(x_1^*(e_1,e_2)\right)$.

For similar reasons as before, agent 1 is kept to his minmax value $\boldsymbol{E}\{b_1(e_1)\}$ and has no incentive to deviate for all $\delta \in [0,1]$. Under the hypothesis that δ is sufficiently large to sustain π^1 as agent 1's worst punishment in the RPE we try to construct, it must be that

$$\begin{aligned} m_1 &= \frac{\boldsymbol{E}\{b_1(e_1)\}}{1-\delta}, \\ M_2 &= \frac{\boldsymbol{E}\left\{b_1\left(x_1^*(e_1,e_2)\right) + b_2\left(x_2^*(e_1,e_2)\right)\right\} - \boldsymbol{E}\{b_1(e_1)\}}{1-\delta}. \end{aligned}$$

Note that M_2 can be rewritten as $\left(\frac{1}{1-\delta}\right)\cdot \boldsymbol{E}\{b_2(e_2)\}$ plus the maximal net expected surplus from cooperation.

If agent 2 deviates by not paying $s^1(e_1,e_2)$, then he will be punished by an immediate switch to the yet unknown path π^2 with unknown m_2 as his worst RPE payoff. Agent 2 will comply to every period of π^2 if the following equilibrium condition holds for any realization (e_1,e_2):

$$-s^1(e_1,e_2) + \delta M_2 \geq 0 + \delta m_2. \tag{16}$$

This condition reveals the minimal difference between M_2 and m_2, that we will use below. Under the hypothesis that π^1, m_1 and M_2 are part of the RPE we are after, we will now characterize π^2, m_2, M_1 and the threshold on δ that sustain π^1 and π^2 as an RPE.

As discussed above, the path π^2 needs a stick and carrot flavor. The stick is a non-negative monetary payment, denoted $p(e_1,e_2) \geq 0$, that agent 2 has to pay to agent 1 in the first period of the infinite path π^2 in case of realization (e_1,e_2). Agent i's expected

continuation RPE payoff from the second period of π^2 onwards is denoted by v_i, where $v_i \in [m_i, M_i]$. Given realization (e_1, e_2) and that π^2 will be restarted next period if agent 2 does not pay, which yields him his worst continuation RPE payoff of m_2, we obtain the following equilibrium condition for agent 2 to comply to the first period of π^2:

$$-p(e_1, e_2) + \delta v_2 \geq 0 + \delta m_2.$$

This condition reveals a trade-off between the stick $p(e_1, e_2)$ and the carrot v_2, larger sticks requiring larger carrots. From rewriting and applying $v_2 \leq M_2$, we obtain

$$p(e_1, e_2) \leq \delta(v_2 - m_2) \leq \delta(M_2 - m_2),$$

which resembles (16). The continuation payoff δm_2 can be attained by equating the first inequality, and the maximal RPE payment that implements δm_2 is

$$p(e_1, e_2) = \delta(M_2 - m_2), \tag{17}$$

which makes $p(e_1, e_2)$ independent of realization (e_1, e_2). Setting $p(e_1, e_2) = \delta(M_2 - m_2)$ implies that, from the second period of π^2, we must follow π^1, otherwise agent 2 cannot attain M_2 and the equilibrium condition (16) would fail. So, the harshest stick available is followed by the sweetest carrot available. To avoid any misunderstanding, $s^1(e_1, e_2) = p(e_1, e_2)$ also satisfies agent 2's equilibrium condition (16) and this agent will comply to paying $p(e_1, e_2)$ to agent 1.

In order to complete the characterization of π^2, we also have to characterize the allocation $x^2(e_1, e_2)$ in the first period of π^2. The equilibrium condition for agent 1 to comply to $x_1^2(e_1, e_2)$, for every realization (e_1, e_2), is given by

$$b_1\left(x_1^2(e_1, e_2)\right) + p(e_1, e_2) + \delta m_1 \geq b_1(e_1) + \delta m_1, \tag{18}$$

where $p(e_1, e_2) + \delta m_1$ is consistent with π^2. Given the non-negativity of $p(e_1, e_2)$, we have that $x_1^2(e_1, e_2) = e_1$ trivially satisfies this equilibrium condition for all realizations of (e_1, e_2) for all $\delta \in [0, 1)$. By definition of m_2, $p(e_1, e_2)$ and π^2, we have that m_2 is the

minimal RPE payoff that satisfies the equilibrium conditions:

$$\begin{aligned} m_2 &= \min_{x^2(e_1,e_2)} \boldsymbol{E}\left\{b_2\left(x_2^2(e_1,e_2)\right) - p(e_1,e_2) + \delta M_2\right\}, \text{ s.t. (16), (17) and (18),} \\ &= \min_{x^2(e_1,e_2)} \boldsymbol{E}\left\{b_2\left(x_2^2(e_1,e_2)\right)\right\} + \delta m_2, \text{ s.t. (16) and (18),} \\ &= \boldsymbol{E}\{b_2(e_2)\} + \delta m_2, \text{ s.t. (16)} \\ &= \frac{\boldsymbol{E}\{b_2(e_2)\}}{1-\delta}, \text{ s.t. (16).} \end{aligned}$$

Agent 2 obtains exactly his minmax payoff in his worst RPE. Note, however, that here it consists of his expected payoff $E\{b_2(e_2)\} - p(e_1,e_2)$ in the first period, which is below his minmax value, followed by his expected maximum RPE payoff in all future periods, which is above his minmax value. Since this is the same payoff as under the trigger strategies, both strategies punishments are payoff-equivalent to agent 2.

Agent 1 does not have an incentive to deviate from $x^2(e_1,e_2) = (e_1,e_2)$ as his current period utility will be lower (less benefit from water use and a foregone payment) followed by his worst RPE from the next period onward. Moreover, note that the difference $M_2 - m_2$ is equal to the present value of the expected net surplus of efficient cooperation and, therefore,

$$p(e_1,e_2) = \delta \frac{\boldsymbol{E}\left\{b_1\left(x_1^*(e_1,e_2)\right) + b_2\left(x_2^*(e_1,e_2)\right) - b_1(e_1) - b_2(e_2)\right\}}{1-\delta}. \tag{19}$$

Note that this transfer is equal to the present value of the entire expected net surplus of all future periods from the second period onwards. Theoretically, this is fine, but in practice it might not be applicable because it can be quite a substantial payment.

So, under the hypothesis that π^1, m_1 and M_2 are part of the RPE, we have characterized the following punishment path for agent 2:

- π^2 : In the first period of π^2, $x^2(e_1,e_2) = (e_1,e_2)$ and $p(e_1,e_2)$ is given by (19). In the second period of π^2 : Switch to π^1.

Finally, we have to check whether our hypothesis holds. For realization (e_1,e_2) and $s^1(e_1,e_2) = p(e_1,e_2)$, (16) can be rewritten as

$$(1-\delta)\left[b_1(e_1) - b_1\left(x_1^*(e_1,e_2)\right)\right] \leq \delta\left[\boldsymbol{E}\left\{b_1\left(x_1^*(e_1,e_2)\right) + b_2\left(x_2^*(e_1,e_2)\right) - b_1(e_1) - b_2(e_2)\right\}\right].$$

Because both terms between square brackets are positive, this conditions holds for sufficiently large $\delta < 1$. Since this condition has to hold for all realizations (e_1,e_2), we

obtain the threshold

$$\delta \geq \underline{\delta}^* \equiv \max_{e_1,e_2} \frac{b_1(e_1) - b_1\left(x_1^*(e_1,e_2)\right)}{b_1(e_1) - b_1\left(x_1^*(e_1,e_2)\right) + \boldsymbol{E}\left\{b_1\left(x_1^*(e_1,e_2)\right) + b_2\left(x_2^*(e_1,e_2)\right) - b_1(e_1) - b_2(e_2)\right\}}.$$

This thresholds corresponds to the lowest threshold $\underline{\delta}\left(x_1^c, x_2^c, s^c\right)$ to sustain individually-rational agreements in SPE that we derived in Section 5.2. Because the denominator is larger than the numerator, we have $\underline{\delta}^* < 1$. Its maximum level is attained for $\max_{e_1,e_2}\left[b_1(e_1) - b_1\left(x_1^*(e_1,e_2)\right)\right]$. This is the realization (e_1,e_2) where the payment to agent 1 is maximal and, therefore, the temptation for agent 1 to defect is highest. This establishes the threshold for which δ is sufficiently large to sustain the pair of paths π^1 and π^2 as the agents' worst possible punishments in any RPE.

For completeness,

$$\begin{aligned} M_1 &= \boldsymbol{E}\{b_1(e_1) + p(e_1,e_2)\} + \delta m_1, \\ &= \frac{(1-\delta)\boldsymbol{E}\{b_1(e_1)\} + \delta\left[\boldsymbol{E}\left\{b_1\left(x_1^*(e_1,e_2)\right) + b_2\left(x_2^*(e_1,e_2)\right)\right\} - \boldsymbol{E}\{b_2(e_2)\}\right]}{1-\delta} \\ &< \frac{\boldsymbol{E}\left\{b_1\left(x_1^*(e_1,e_2)\right) + b_2\left(x_2^*(e_1,e_2)\right)\right\} - \boldsymbol{E}\{b_2(e_2)\}}{1-\delta}, \end{aligned}$$

where the last expression is agent 1's *utopia* payoff in the set of individually-rational payoff vectors. As δ goes to 1, the entire set can be sustained as RPE payoffs. Note that, due to the sequential setting, agent 1's best RPE payoff M_1 is always less than his utopia payoff, but agent 2 has an RPE in which he can attain his utopia payoff.

We now extend the analysis to characterize the largest set of Pareto efficient paths π^0 that can be supported by π^1 and π^2 as an RPE. For parameter $\alpha \in [0,1)$, we consider the following intended equilibrium path in order to support ANBS with payment rule (12):

- $\pi^0(\alpha)$: In every period, $x^0(e_1,e_2) = x^*(e_1,e_2)$ and $s^{\alpha^0}(e_1,e_2) = b_1(e_1) - b_1\left(x_1^*(e_1,e_2)\right) + \alpha\left[b_1\left(x_1^*(e_1,e_2)\right) + b_2\left(x_2^*(e_1,e_2)\right) - b_1(e_1) - b_2(e_2)\right]$.

This path selects the efficient allocation combined with a payment that depends on parameter $\alpha \in [0,1)$, which distributes the cooperative surplus. Note that the boundary $\alpha = 1$ is not included, because then π^0 Pareto-dominates π^2, which is not allowed in RPE.

Figure 5 illustrates the range of ex ante expected values associated with $\pi^0(\alpha)$, π^1 and π^2 for parameter values as introduced in Figure 1. Note that π^2 is Pareto inefficient because $(b_1(e_1), b_2(e))$ determines the first period's payoff. Because of this inefficiency,

the range of Pareto efficient RPEs does not satisfy the stronger concept of Strong Perfect Equilibrium in e.g. Rubinstein (1980), in which all three paths of $\left(\pi^0(\alpha),\pi^1,\pi^2\right)$ have to be Pareto efficient.

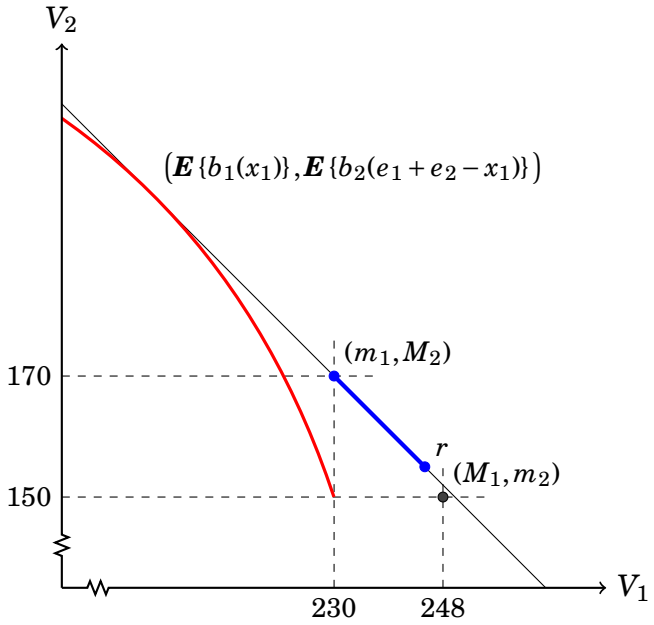


Figure 5: The range of *ex ante* expected values associated with $\pi^0(\alpha)$, π^1 and π^2 is shown for parameter values as used in Figure 1. The set of $\pi^0(\alpha)$ is the thick blue segment on the Pareto frontier from payoff pair (m_1,M_2) up to some payoff pair r in south-east direction, depending on δ; The red curve is the parametric equation $\left(E\{b_1(x_1)\},E\{b_2(e_1+e_2-x_1)\}\right)$, assuming without loss of generality $x_1\left(e_1^H,e_2^H\right) = x_1\left(e_1^L,e_2^L\right)+2$. This curve determines both the location of the Pareto frontier as well as the location of payoff pair (m_1,M_2). We obtain $(m_1,M_2) = (230,170)$, $(M_1,m_2) = (248,150)$, and $r = (245,155)$.

The final step of this appendix is to check for which α and δ do $\left(\pi^0(\alpha),\pi^1,\pi^2\right)$ form an RPE. By construction, $\pi^0(\alpha)$ constitutes a Pareto efficient individually-rational agreement. For similar reasons as before, agent 1 has no incentive to deviate for all $\delta\in[0,1]$. Therefore, we focus on deterring deviations by agent 2, which would trigger his worst RPE path π^2. Denote agent i's *ex ante* expected values of the cooperative path $\pi^0(\alpha)$ as $V_i^c(\alpha)$, $i=1,2$. For realization (e_1,e_2), agent 2 has no incentive to deviate if

$$-s^{\alpha^0}(e_1,e_2)+\delta V_2^c(\alpha)\geq 0+\delta m_2 \qquad\Longleftrightarrow\qquad s^{\alpha^0}(e_1,e_2)\leq\delta\left(V_2^c(\alpha)-m_2\right).$$

By substituting the expressions for $\delta V_2^c(\alpha)$ and m_2, we obtain

$$s^{\alpha^0}(e_1,e_2)+\frac{\delta}{1-\delta}\boldsymbol{E}\left\{s^{\alpha^0}(e_1,e_2)\right\}\leq\frac{\delta}{1-\delta}\left[\boldsymbol{E}\left\{b_2\left(x_2^*(e_1,e_2)\right)\right\}-\boldsymbol{E}\left\{b_2(e_2)\right\}\right],$$

which resembles (5).

We do not directly substitute for $s^{\alpha^0}(e_1,e_2)$ in order to avoid a messy condition. Instead, we solve for δ to obtain that agent 2 has no incentive to deviate, for all realizations (e_1,e_2), if:

$$\delta \geq \underline{\delta}^{\alpha}(x_1^*,x_2^*) \equiv \max_{e_1,e_2} \frac{s^{\alpha^0}(e_1,e_2)}{s^{\alpha^0}(e_1,e_2)+\boldsymbol{E}\left\{b_2\left(x_2^*(e_1,e_2)\right)-b_2(e_2)\right\}-\boldsymbol{E}\left\{s^{\alpha^0}(e_1,e_2)\right\}},$$

with $s^{\alpha^0}(e_1,e_2)$ according to π^0. This establishes the threshold $\underline{\delta}^{\alpha}(x_1^*,x_2^*)$ for which δ is sufficiently large that $\left(\pi^0(\alpha),\pi^1,\pi^2\right)$ forms an RPE. Because the payment rule is the ANBS payment rule in (12), this threshold coincides with the ANBS threshold (13).

Because $s^{\alpha^0}(e_1,e_2)$ is increasing in α, also its expectation $E\left\{s^{\alpha^0}(e_1,e_2)\right\}$ is increasing in α. As a result, it is straightforward to verify that $\underline{\delta}^{\alpha}(x_1^*,x_2^*)$ is increasing in α. The interpretation of this relation is that the threshold discount factor for which RPE can be sustained is increasing in the payment within the bounds set by $\pi^0(\alpha)$. This interpretation corresponds to the observation made in Section 5.3 with respect to the bargaining strength of agent 1.

Finally, in the limit when δ goes to 1, division by $1-\delta$ causes a mathematical problem in deriving limits of the m_i's and M_i's. To overcome this problem, the literature on repeated games works with normalized discounted payoffs, i.e. $(1-\delta)\sum_{t=0}^{\infty}\delta^t u_t$, where u_t is the payoff in period t. As δ goes to 1, the limit of these normalized discounted payoffs is well defined. In addition, optimal strategies are unaffected by this change in payoffs and normalized discounted payoffs converge to the limiting average of the stream of undiscounted payoffs, i.e. $\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T}u_t$. By taking the limit δ goes to 1 of the normalized discounted payoffs of the renegotiation-proof equilibria, we obtain that (M_1,m_2) moves straight east to the Pareto frontier in Figure 5, where it coincides with payoff-pair r in the limit at agent 1's utopia payoff. The limit point cannot be sustained for $\delta \in [0,1)$.

References

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica 56*(2), 383–396.

Ambec, S., A. Dinar, and D. McKinney (2013). Water sharing agreements sustainable to reduced flows. Forthcoming in *Journal of Environmental Economics and Management*.

Ambec, S. and Y. Sprumont (2002). Sharing a river. *Journal of Economic Theory 107*(2), 453–462.

Ansink, E. and H. Houba (2012). Market power in water markets. *Journal of Environmental Economics and Management 64*(2), 237–252.

Ansink, E. and A. Ruijs (2008). Climate change and the stability of water allocation agreements. *Environmental and Resource Economics 41*(2), 249–266.

Barrett, S. (1994). Conflict and cooperation in managing international water resources. World Bank Policy Research Working Paper 1303.

Beach, H., J. Hammer, J. Hewitt, E. Kaufman, A. Kurki, J. Oppenheimer, and A. Wolf (2000). *Transboundary Freshwater Dispute Resolution: Theory, Practice, and Annotated References*. Tokyo: United Nations University Press.

Béal, S., A. Ghintran, E. Rémila, and P. Solal (2013). The river sharing problem: A survey. *International Game Theory Review 15*(3), 1340016.

Beard, R. and S. McDonald (2007). Time-consistent fair water sharing agreements. In S. Jørgensen, M. Quincampoix, and T. L. Vincent (Eds.), *Advances in Dynamic Game Theory*. Boston: Birkhäuser.

Bennett, L. and C. Howe (1998). The interstate river compact: Incentives for noncompliance. *Water Resources Research 34*(3), 485–495.

Bennett, L., C. Howe, and J. Shope (2000). The interstate river compact as a water allocation mechanism: Efficiency aspects. *American Journal of Agricultural Economics 82*(4), 1006–1015.

Carraro, C., C. Marchiori, and A. Sgobbi (2007). Negotiating on water: Insights from non-cooperative bargaining theory. *Environment and Development Economics 12*(2), 329–349.

De Stefano, L., J. Duncan, S. Dinar, K. Stahl, K. Strzepek, and A. Wolf (2012). Climate change and the institutional resilience of international river basins. *Journal of Peace Research 49*(1), 193–209.

Dinar, S. (2006). Assessing side-payment and cost-sharing patterns in international water agreements: The geographic and economic connection. *Political Geography 25*(4), 412–437.

Dinar, S. (2009). Power asymmetry and negotiations in international river basins. *International Negotiation 14*(2), 329–360.

Drieschova, A., M. Giordano, and I. Fischhendler (2008). Governance mechanisms to address flow variability in water treaties. *Global Environmental Change 18*(2), 285–295.

Farrell, J. and E. Maskin (1989). Renegotiation in repeated games. *Games and Economic Behavior 1*(4), 327–360.

Fudenberg, D. and J. Tirole (1991). *Game Theory*. Cambridge, MA: MIT Press.

Gastélum, J., J. Valdés, and S. Stewart (2009). A decision support system to improve water resources management in the Conchos basin. *Water Resources Management 23*(8), 1519–1548.

Gauthier, C., M. Poitevin, and P. González (1997). Using ex ante payments in self-enforcing risk-sharing contracts. *Journal of Economic Theory 76*(1), 106–144.

Giordano, M., A. Drieschova, J. A. Duncan, Y. Sayama, L. De Stefano, and A. T. Wolf (2013). A review of the evolution and state of transboundary freshwater treaties. Forthcoming in *International Environmental Agreements*.

Houba, H. (2008). Computing alternating offers and water prices in bilateral river basin management. *International Game Theory Review 10*(3), 257–278.

Houba, H., G. Van der Laan, and Y. Zeng (2013). Asymmetric Nash solutions in the river sharing problem. Tinbergen Institute Discussion Paper 2013-051/II.

Katz, D. and M. Moore (2011). Dividing the waters: An empirical analysis of interstate compact allocation of transboundary rivers. *Water Resources Research 47*(6), W06513.

Khmelnitskaya, A. (2010). Values for rooted-tree and sink-tree digraph games and sharing a river. *Theory and Decision 69*(4), 657–669.

Kilgour, D. and A. Dinar (2001). Flexible water sharing within an international river basin. *Environmental and Resource Economics 18*(1), 43–60.

Rubinstein, A. (1980). Strong perfect equilibrium in supergames. *International Journal of Game Theory 9*(1), 1–12.

Van Damme, E. (1989). Renegotiation-proof equilibria in repeated prisoners' dilemma. *Journal of Economic Theory 47*(1), 206–217.

Van den Brink, R., G. van der Laan, and N. Moes (2012). Fair agreements for sharing international rivers with multiple springs and externalities. *Journal of Environmental Economics and Management 63*(3), 388–403.

Ward, F. (2013). Forging sustainable transboundary water-sharing agreements: Barriers and opportunities. *Water Policy 15*(3), 386–417.

Wen, Q. (2002). A Folk theorem for repeated sequential games. *Review of Economic Studies 69*(2), 493–512.