TI 2012-006/1
Tinbergen Institute Discussion Paper

Transboundary Externalities and Property Rights: An International River Pollution Model

Gerard van der Laan
Nigel Moes

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

Transboundary externalities and property rights: an international river pollution model[1]

Gerard van der Laan[2] and Nigel Moes[3]

January 19, 2012

[1]This research is financially supported by Netherlands Organization for Scientific Research NWO grant 400-07-159.

[2]G. van der Laan, Department of Econometrics and Tinbergen Institute, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands. E-mail: glaan@feweb.vu.nl.
[3]N. Moes, Department of Econometrics and Tinbergen Institute, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands. E-mail: nmoes@feweb.vu.nl.

Abstract

In this paper we study international river pollution problems. We introduce a model in which the agents (countries) located along a river derive benefit while causing pollution, but also incur environmental costs of experiencing pollution from all upstream agents. We find that total pollution in the model decreases when the agents decide to cooperate. The resulting gain in social welfare can be distributed among the agents based on the property rights over the river. Using principles from international water law we suggest 'fair' ways of distributing the property rights and therefore the cooperative gain.

Keywords: international river, pollution, externality, property rights, value
JEL codes: C70, D60, Q53

1 Introduction

River water is often not only used directly for consumption (drinking water, irrigation), but also indirectly for the discharge of agricultural, biological and industrial waste products. The discharge of these products in a river can lead to pollution, which, in turn, can cause environmental damage. River pollution provides a classic example of a negative externality: when an upstream agent (e.g. country, state, city or firm) pollutes a river, this can create external costs for the agents downstream of it. Conversely, downstream agents cannot inflict external costs on upstream agents because water in a river, and therefore pollution, is not able to flow up stream. Asymmetric dependence on a water resource, like this, can cause disputes about the use of the resource, especially if property rights over it are not clearly defined. Intranational disputes about water resources are usually settled through a country's legal system, but in international disputes there typically is no third party that is able to enforce agreements. Since upstream agents obtain all the benefits but only bear part of the social costs while polluting a river, a situation of over-pollution relative to the social optimum is likely to arise in international rivers.

The well-known theorem of Coase (1960) states that when trade in an externality (pollution caused by an upstream agent to a downstream agent) is possible and there are no transaction costs, bargaining leads to an efficient outcome, regardless of the initial allocation of property rights. Because countries are able to bargain over agreements that would reduce pollution in an international river, in practice, we expect to observe similar levels of pollution in intranational and international rivers. Sigman (2002), however, finds that at water quality monitoring stations immediately upstream of international borders the pollution levels are more than 40 percent higher than the average levels at control stations. She concludes that, while rivers would seem to provide a good case for international cooperation (because they involve small numbers of countries and relatively well defined benefits and costs), cooperation on river pollution has not evolved between countries sharing rivers.[1] The reason for this lack of cooperation in international river pollution problems is the absence of clearly defined property rights over the river. All countries sharing an international river usually claim property rights over it (at least that part of the river on their territories) and none are normally willing to reduce their pollution or pay compensation to countries suffering from it.

In this paper we study how international water law doctrines can be used to solve river pollution problems through cooperation. A river is considered 'international' if it is shared by two or more sovereign states (Barrett, 1994). International rivers fall into

[1] Sigman (2002) notes that the countries in the European Union seem to be an exception. See also Barrett (1994) for an example of an agreement between Switzerland, Germany, France and the Netherlands concerning the reduction in salt pollution of the Rhine river by a French potash mine.

two categories: boundary (or contiguous) rivers and successive rivers. A boundary river flows between the territories of two or more states and hence forms the border between the states. A successive river flows from the territory of one state into the territory of another state (Garretson, Hayton and Olmstead (1967)). It is also possible that an international river is (partly) a boundary river and (partly) a successive river. In this paper we only consider successive international rivers.

Several water resource issues have been modeled using models from (cooperative) game theory, see Parrachino, Dinar and Patrone (2006) for an overview. Recently, especially the rival consumption of water from successive international rivers has received attention. Here the main problem is that water consumed by an upstream country can no longer be consumed by a downstream country. It is clear that in water stressed regions this can create tension between countries sharing a river because the population of a downstream country might (also) depend on the water inflow in the river at an upstream country. Kilgour and Dinar (1995, 2001), Ambec and Sprumont (2002), Ambec and Ehlers (2008), Khmelnitskaya (2010), Wang (2011), van den Brink, van der Laan and Moes (2011) all use game-theoretic models to investigate the distribution of water among countries sharing an international river. In Ansink and Weikard (2011) and van den Brink, Estévez-Fernández, van der Laan and Moes (2011) a closely related axiomatic approach is followed.

The economic literature on the non-rival use of (international) rivers appears to be limited. Apart from the above mentioned paper of Sigman (2002), there exist three empirical papers of Gray and Shadbegian (2004), Sigman (2005) and Lipscomb and Mobarak (2007) that study transboundary river pollution between states and counties in the United States and Brazil. Mäler (1990), Barrett (1994), Fernandez (2002, 2009) and Dinar (2006) all study two-country river pollution problems. Two theoretical papers that model a multi-country setting are that of Ni and Wang (2007) and Gengenbach, Weikard and Ansink (2010). The model of Gengenbach, Weikard and Ansink (2010) is close to ours in the sense that there is a river with a unidirectional flow of pollution and the agents (countries) along the river are able to choose their own level of pollution abatement (in our model agents choose pollution levels instead of pollution abatement levels). Within this model they analyze how voluntary joint action of the agents along the river can increase pollution abatement. The main difference between the paper of Gengenbach, Weikard and Ansink (2010) and ours is that their emphasis is on the stability of coalitions of cooperating agents, while we focus on property rights and the distribution of the gain in social welfare that arises when countries along an international river switch from no cooperation on pollution levels to full cooperation. Our model also differs substantially from the river pollution model of Ni and Wang (2007). In their model pollution levels are not specified. Instead, it is assumed there is a set of agents N along an international river and each agent $i \in N$

has exogenously given (environmental) costs $C_i \in \mathbb{R}_+$ caused by the pollution of the agent i itself and all agents upstream to it. The problem then is to divide the total costs of pollution $\sum_{i \in N} C_i$ among the agents located along the river. For this problem Ni and Wang (2007) provide and characterize two solutions. The Local Responsibility Sharing method holds each agent i responsible for the costs C_i on its own territory and therefore requires that each agent i pays its own costs C_i. The Upstream Equal Sharing method recognizes that the costs C_i on the territory of agent i are caused by i and all its upstream agents and thus requires that C_i is divided equally among those agents.

In this paper we model the pollution problem by assuming that each agent (country) chooses a level of pollution. Several agents are located along the river from upstream to downstream. Each agent can perform activities that cause pollution. The higher the level of activities, the higher the corresponding level of pollution caused by the agent. An agent derives benefits from its level of activities, and thus its own level of pollution, but also incurs environmental costs if polluted river water flows through its territory. An agent therefore does not only suffer from its own level of pollution, but also from the pollution levels of all its upstream agents.[2] The agents value pollution of the river water differently in the sense that some agents have higher needs (marginal utility) for the emission of pollutants than others. The heterogeneous valuations of the agents are introduced by endowing each agent with an agent specific benefit and cost function. Together these two functions determine the utility function of the agent. The benefit function of an agent depends only on its own pollution level, its cost function depends on the pollution emissions of the agent itself and of all the agents that are located upstream of it. So, while in the rival consumption river problem the water consumption of an agent is restricted by the consumption of the agents upstream to it, in this non-rival case of pollution the use of river water by an upstream agent enters the utility functions of all agents downstream to it.

By absence of clearly defined property rights in international river situations, typically each country claims to have the right over the river on its own territory and therefore also the right to choose its own level of pollution. In our model, under non-cooperative behavior each agent chooses a pollution level that maximizes its own utility, given the pollution levels of the others. The resulting non-cooperative Nash equilibrium is usually inefficient, i.e., the sum of all utilities (social welfare) of agents along the river can be increased by coordinating the pollution levels among the agents. However, coordinating the pollution levels in order to maximize social welfare will normally result in lower utility for some of the agents, unless the agents are able to reach an agreement on both the optimal pollution levels as well as a distribution of the total social welfare by making

[2]This is, for instance, the case when river water is used in an industrial process that creates some sort of benefit for the polluter but at the same time causes environmental damage at the locations of the polluter and all agents downstream to it.

monetary compensations amongst each other. Therefore we assume that the agents are able to make monetary transfers to each other. Under well-specified property rights, the Coase theorem then implies that the agents would be able to reach an agreement and to determine appropriate monetary compensations. However, since typically property rights over international rivers are not specified we have to find a way to determine them. We do this by referring to doctrines from international water law. The doctrines that we consider are the principle of Absolute Territorial Sovereignty (ATS), the principle of Unlimited Territorial Integrity (UTI) and the principle of Territorial Integration of all Basin States (TIBS). We find that each of these principles allocates the property rights over the river in a different way, so that each of the principles provides a different answer to the question of what monetary transfers are appropriate and necessary to establish cooperation among agents in our international river pollution model.

In short, the main contributions of this paper are the following: we introduce a model for international river pollution problems in which the agents choose pollution levels and are able to make monetary compensations to each other. We find that the total level of pollution is always lower under cooperation (if agents coordinate their pollution levels) than under individual action. The gain in social welfare that results when the agents switch from their Nash equilibrium pollution levels to the socially optimal pollution is distributed among the agents through monetary transfers. Since in international river situations property rights are not clearly defined, we refer to three principles from international water law to provide solutions to the welfare distribution problem and the corresponding monetary transfers to implement such a distribution of the cooperative gains.

The paper is organized as follows. In Section 2 we introduce the (international) river pollution model, derive the unique Nash equilibrium and Pareto efficient pollution levels and show that the total level of pollution is always lower in the Pareto efficient outcome than in the Nash equilibrium. In Section 3 we investigate, for the two agent problem, how property rights might determine 'fair' distributions of the cooperative gain that results when changing from Nash equilibrium to Pareto efficient pollution levels. In Section 4 we extend this analysis to an arbitrary number of agents by using the solution concept of 'value' from cooperative game theory. In Section 5 we introduce a class of values that arise by applying the water rights distribution principle of Territorial Integration of all Basin States. In Section 6 we generalize the river pollution model to rivers with multiple springs and/or multiple sinks. Finally, we conclude in Section 7.

2 River pollution problems

2.1 The model

Consider a successive river flowing through a finite set of agents (countries). The set of agents is denoted by $N \subset \mathbb{N}$. Unless stated otherwise, without loss of generality we assume that $N = \{1, \ldots, n\}$ for some $n \in \mathbb{N}$ and that the agents are labeled from upstream to downstream, i.e., agent 1 is the most upstream agent, followed by agent 2 and so on until the most downstream agent n. Thus, for two agents $i, j \in N$, we have that agent i is upstream of agent j (and agent j is downstream of agent i) when $i < j$. For each agent $i \in N$, write $P^i = \{1, \ldots, i\}$ as the subset of N containing agent i and all its upstream agents, and $Q_i = \{i, \ldots, n\}$ as the subset of N containing i and all its downstream agents. In the sequel, for $K \subset \mathbb{N}$, $\mathbb{R}^K$ denotes the $|K|$-dimensional Euclidean space with elements $x \in \mathbb{R}^K$ having components $x_i,\ i \in K$. The vector $\mathbf{0} \in \mathbb{R}^K$ denotes the null-vector with all components equal to zero.

Each agent $i \in N$ can choose a level $p_i \in \mathbb{R}_+$ of pollution.[3] We collect these individual pollution levels in the $|N|$-dimensional pollution vector $p \in \mathbb{R}^N_+$. Because the river is transporting the pollution caused by some agent to all its downstream agents, the pollution experienced by agent $i \in N$ depends on the levels of pollution of the agent itself and all its upstream agents. We assume that the pollution experienced by agent i is given by the function $q_i \colon \mathbb{R}^N_+ \to \mathbb{R}_+$ defined by $q_i(p) = \sum_{j=1}^{i} p_j$, i.e., the level of pollution experienced by i is equal to the sum of all pollution levels of the agents in P^i.

We further assume that each agent along the river derives benefit while causing pollution but also incurs (environmental) costs of experiencing it. The benefit of an agent i only depends on its own pollution level and is given by a function $b_i \colon \mathbb{R}_+ \to \mathbb{R}_+$, yielding benefit $b_i(p_i)$ for every $p_i \geq 0$. The pollution costs of an agent i depends on the total pollution $q_i(p)$ of the agents in P^i and are given by a function $c_i \colon \mathbb{R}_+ \to \mathbb{R}_+$, yielding costs $c_i(q_i)$ for every $q_i \geq 0$. In the sequel b_i' and b_i'' denote the first and second order derivatives of b_i with respect to p_i, and c_i' and c_i'' denote the first and second order derivatives of c_i with respect to q_i. We make the following assumptions about the benefit and cost functions of the agents.

Assumption 2.1
1. *For every $i \in N$: $b_i(0) = 0$ and, for all $p_i > 0$, b_i is twice differentiable with $b_i'(p_i) > 0$ and $b_i''(p_i) < 0$. In addition, $b_i'(p_i) \to \infty$ as $p_i \to 0$ and $b_i'(p_i) \to 0$ as $p_i \to \infty$.*

[3] We could also let the agents choose the level of production in some industrial process that causes river pollution. If we would assume that pollution is strictly increasing in the production level and modify the subsequent assumptions appropriately, this model would lead to similar conclusions.

2. *For every* $i \in N$: $c_i(0) = 0$ *and, for all* $q_i > 0$, c_i *is twice differentiable with* $c_i'(0) > 0$ *and* $c_i''(q_i) > 0$.

The first assumption states that agents obtain no benefit when there is no pollution and that the marginal benefits of pollution are positive and (strictly) decreasing. Further, the marginal benefits tend to infinity when pollution tends to zero and tend to zero when pollution tends to infinity. The second assumption states that agents incur no costs when there is no pollution and implies that the marginal costs of pollution are positive and (strictly) increasing. Notice that under Assumption 2.1, for every $i \in N$ there exists a unique positive real number, say r_i, such that $b_i'(r_i) = c_i'(r_i)$. Let $r \in \mathbb{R}_+^N$ be the vector of these positive real numbers.

Pollution levels $p \in \mathbb{R}_+^N$ result in utilities

$$u_i(p) = b_i(p_i) - c_i(q_i(p)), \quad i \in N.$$

That is, the utility of agent i is the difference between its pollution benefit $b_i(p_i)$ and the pollution costs $c_i(q_i(p)) = c_i(\sum_{j=1}^{i} p_j)$. We assume that utility is transferable. This means that agents are able to transfer utility to each other by making monetary transfers. The monetary transfer to agent $i \in N$ is equal to $t_i \in \mathbb{R}$. When $t_i > 0$ agent i receives a monetary transfer and when $t_i < 0$ agent i pays a monetary transfer. A (monetary) compensation scheme is a vector $t \in \mathbb{R}^N$ that satisfies the restriction

$$\sum_{i=1}^{n} t_i \leq 0, \tag{2.1}$$

i.e., the sum of all monetary transfers is at most equal to zero. A compensation scheme is said to be budget balanced if $\sum_{i=1}^{n} t_i = 0$. Pollution levels p and a compensation scheme t result in payoffs

$$z_i(p, t) = u_i(p) + t_i, \quad i \in N.$$

In the sequel we assume that the agents in the model are rational utility maximizers and that all benefit and cost functions are common knowledge. The tuple (N, b, c), with $b = \{b_i | i \in N\}$ the collection of benefit functions and $c = \{c_i | i \in N\}$ the collection of costs functions, constitutes the input of a river pollution model. The output of the model is a pair (p, t) of pollution levels and monetary transfers, yielding payoffs $z_i(p, t)$, $i \in N$. Given the input of a river pollution model (N, b, c), the aim of this paper is to make both positive and normative statements about the output (p, t) under the restriction that t satisfies (2.1).

2.2 The Nash equilibrium output

We start the analysis of the river pollution model (N, b, c) by considering the situation in which each agent acts individually. In this situation there (clearly) are no monetary transfers and each agent $i \in N$ maximizes its utility $u_i(p)$ with respect to the variable p_i under its control. So, each agent i chooses its pollution level p_i so as to maximize its own utility, given the pollution levels of the other agents. This behavior results in Nash equilibrium pollution levels. The next proposition shows that in the Nash equilibrium each agent $i \in N$ sets the unique (strictly positive) optimal pollution level $\widehat{p}_i$ at which its marginal benefit of pollution is equal to its marginal cost.

Proposition 2.2 *For a river pollution model (N, b, c) that satisfies Assumption 2.1, there exists a unique Nash equilibrium pollution vector $\widehat{p} \in \mathbb{R}^N_+$ in which all pollution levels are strictly positive, $\widehat{p}_i > 0$, $i \in N$.*

Proof. When each agent $i \in N$ acts individually, it maximizes its utility $u_i(p) = b_i(p_i) - c_i(q_i(p))$ given the pollution levels p_j, $j < i$, of its upstream agents. We show the uniqueness of the Nash equilibrium pollution levels by induction on the labels of the agents.

The utility of the most upstream agent 1 is independent of the pollution levels of all other agents and is given by $u_1(p) = b_1(p_1) - c_1(p_1)$. Maximizing this with respect to $p_1 \geq 0$ gives the first order condition

$$b_1'(p_1) - c_1'(q_1(p))\frac{\partial q_1(p)}{\partial p_1} = b_1'(p_1) - c_1'(p_1) \leq 0 \perp p_1 \geq 0.$$

By Assumption 2.1 it follows that there exists a unique solution $\widehat{p}_1 > 0$ (note that $\widehat{p}_1 = r_1$). By the same assumption we have that $b_1''(p_1) < 0$ and $c_1''(q_1) = c_1''(p_1) > 0$ for every $p_1 > 0$ and thus $\widehat{p}_1$ satisfies the second order condition $b_1''(p_1) - c_1''(p_1) < 0$ for utility maximization.

Proceeding by induction, assume that for some $1 < i \leq n$, $p_j = \widehat{p}_j > 0$ has been uniquely determined for all $j < i$. The utility of agent i is given by $u_i(p) = b_i(p_i) - c_i(q_i(p))$. Maximizing this utility function with respect to $p_i \geq 0$ gives the first order condition

$$\frac{\partial b_i(p_i)}{\partial p_i} - \frac{\partial c_i(q_i(p))}{\partial q_i}\frac{\partial q_i(p)}{\partial p_i} \leq 0 \perp p_i \geq 0.$$

With $q_i(p) = p_i + \sum_{j=1}^{i-1} \widehat{p}_j$ we obtain the system

$$\begin{aligned} &b_i'(p_i) - c_i'(q_i) \leq 0 \perp p_i \geq 0, \\ &q_i = p_i + \textstyle\sum_{j=1}^{i-1} \widehat{p}_j. \end{aligned} \tag{2.2}$$

By Assumption 2.1.1 b_i' is strictly decreasing in p_i with $b_i'(p_i) \to \infty$ as $p_i \to 0$ and $b_i'(p_i) \to 0$ as $p_i \to \infty$. By Assumption 2.1.2 $c_i'(0) > 0$ and c_i' is strictly increasing in q_i (and therefore strictly increasing in p_i). Hence, for the given pollution levels $\widehat{p}_j$, $j < i$, there exists a unique pollution level $\widehat{p}_i > 0$ that satisfies (2.2). Since, by the same assumptions, b_i'' is negative and c_i'' is positive, it follows that $\widehat{p}_i$ also satisfies the second order condition $b_i''(p_i) - c_i''(q_i) < 0$ for utility maximization. □

Notice that in the Nash equilibrium output all monetary transfers are equal to zero so that the payoffs are given by $z_i(\widehat{p}, \mathbf{0}) = u_i(\widehat{p}) = b_i(\widehat{p}_i) - c_i(\sum_{j=1}^{i} \widehat{p}_j)$, $i \in N$.

2.3 Social welfare and Pareto efficiency

In the river pollution model (N, b, c) the social welfare associated with pollution levels $p \in \mathbb{R}_+^N$ is measured by the difference between the total social benefit $\sum_{i \in N} b_i(p_i)$ and the total social costs $\sum_{i \in N} c_i(\sum_{j=1}^{i} p_j)$. The social welfare function $W\colon \mathbb{R}_+^N \to \mathbb{R}$ assigns to each vector $p \in \mathbb{R}_+^N$ of pollution levels the social welfare[4]

$$W(p) = \sum_{i \in N} b_i(p_i) - \sum_{i \in N} c_i(\sum_{j=1}^{i} p_j).$$

In the next proposition we show that there exist unique and strictly positive pollution levels $\widehat{p}_i$, $i \in N$, that maximize $W(p)$.

Proposition 2.3 *For a river pollution model (N, b, c) that satisfies Assumption 2.1, there exists a unique vector of pollution levels $\widetilde{p} \in \mathbb{R}_+^N$ that maximizes social welfare $W(p)$. In $\widetilde{p}$ all pollution levels are strictly positive.*

Proof. Maximization of $W(p)$ with respect to $p_i \geq 0$, $i \in N$, yields the system of n first order conditions

$$\frac{\partial W(p)}{\partial p_i} = \frac{\partial b_i(p_i)}{\partial p_i} - \sum_{k=i}^{n} \frac{\partial c_k(q_k(p))}{\partial q_k} \frac{\partial q_k(p)}{\partial p_i} \leq 0 \perp p_i \geq 0, \quad i \in N.$$

Since $q_j(p) = \sum_{i=1}^{j} p_i$, we have that $\frac{\partial q_j(p)}{\partial p_i} = 1$ for every $i, j \in N$ with $i \leq j$ and thus the system reduces to

$$\frac{\partial b_i(p_i)}{\partial p_i} - \sum_{k=i}^{n} \frac{\partial c_k(q_k(p))}{\partial q_k} \leq 0 \perp p_i \ \geq 0, \quad i \in N. \tag{2.3}$$

First, observe that at a solution to this system $p_i \leq r_i$ for all $i \in N$ because, for every $p_j \geq 0$, $j < i$, it holds that $b_i'(p_i) < c_i'(\sum_{j=1}^{i-1} p_j + p_i)$ if $p_i > r_i$. Second, at a solution it must hold that $p_i > 0$ for all $i \in N$, because $b_i'(p_i) \to \infty$ as $p_i \to 0$ and $c_i'(\sum_{j=1}^{i} p_j)$ is bounded from above by $c_i'(\sum_{j=1}^{i} r_j)$ for all $p_j \in [0, r_j]$, $j \leq i$. So, any solution of the system (2.3) is strictly positive (and bounded from above by the vector r). To maximize the social welfare $W(p)$ we thus have to find a strictly positive solution to the system

$$\frac{\partial b_i(p_i)}{\partial p_i} - \sum_{k=i}^{n} \frac{\partial c_k(q_k(p))}{\partial q_k} = 0, \quad i \in N. \tag{2.4}$$

For agent n the system yields

$$\frac{\partial b_n(p_n)}{\partial p_n} - \frac{\partial c_n(q_n(p))}{\partial q_n} = 0. \tag{2.5}$$

[4] When we want to stress that we are working within the model (N, b, c) we sometimes write $W_{(N,b,c)}(p)$ for the social welfare function $W(p)$.

For an agent $h < n$ we obtain

$$\frac{\partial b_h(p_h)}{\partial p_h} = \sum_{k=h}^{n} \frac{\partial c_k(q_k(p))}{\partial q_k} = \frac{\partial c_h(q_h(p))}{\partial q_h} + \sum_{k=h+1}^{n} \frac{\partial c_k(q_k(p))}{\partial q_k}.$$

Using that

$$\sum_{k=h+1}^{n} \frac{\partial c_k(q_k(p))}{\partial q_k} = \frac{\partial b_{h+1}(p_{h+1})}{\partial p_{h+1}}$$

we obtain

$$\frac{\partial b_{h+1}(p_{h+1})}{\partial p_{h+1}} = \frac{\partial b_h(p_h)}{\partial p_h} - \frac{\partial c_h(q_h(p))}{\partial q_h}, \quad h < n. \tag{2.6}$$

So, a solution of the welfare maximization problem has to satisfy the system (2.5) and (2.6) of n equations. Now, take some $p_1 > 0$. Since b'_h is strictly decreasing in p_h and c'_h is strictly increasing in q_h for all $h \in N$, it follows that for each value of p_1 there exists a unique positive value p_2 that solves equation (2.6) for $h = 1$, as long as the right hand side of the equation is positive. Further, this value of p_2 is increasing in p_1. Continuing in this way, it follows that for each value of $p_1 > 0$ there is a sequence of unique positive values $p_2, p_3, ..., p_n$ that sequentially solves equation (2.6) for $h = 1, 2, ..., n-1$, as long as the right hand sides of all equations are positive, and that all these values are increasing in p_1. Hence, there exists a unique value of p_1 such that the value p_n obtained from sequentially solving the equations (2.6) for $h = 1, 2, ..., n-1$, solves equation (2.5). It can be concluded that the system (2.5) and (2.6) of n equations has a unique solution $\widetilde{p}_i$, $i \in N$.

It remains to show that $\widetilde{p}$ yields a maximum of the social welfare function $W(p)$. Recall that the components r_i, $i \in N$, of the vector $r \in \mathbb{R}^N_+$ satisfy $b'_i(r_i) = c'_i(r_i)$. Since $\widetilde{p}$ also satisfies the system (2.4), it follows that $\tilde{p}_i < r_i$, $i \in N$. Since the objective function $W(p)$ is continuous in p, it follows by Weierstrass' (extreme value) theorem that $W(p)$ has a maximum on the compact set $\{p \in \mathbb{R}^N \mid 0 \leq p_i \leq r_i,\ i \in N\}$. Since $\frac{\partial W}{\partial p_i} > 0$ if $p_i = 0$ and $\frac{\partial W}{\partial p_i} < 0$ if $p_i = r_i$, $i \in N$, it follows that the maximum is achieved in the interior of this set and thus has to satisfy the first order condition (2.4). Hence, the unique solution to this system yields the maximum. □

The following proposition states that the total pollution in the Pareto efficient outcome is always lower than the total pollution in the Nash equilibrium output. The proof of this proposition is given in Appendix A.

Proposition 2.4 *For the river pollution model (N, b, c), $|N| \geq 2$, satisfying Assumption 2.1, it holds that* $\sum_{i=1}^{n} \widetilde{p}_i < \sum_{i=1}^{n} \widehat{p}_i$.

With slight abuse of notation, in the sequel we denote the highest social welfare that can be obtained in the river pollution model (N, b, c) by $W(N, b, c)$. That is, $W(N, b, c)$ is the social welfare $W(\widetilde{p})$ at the Pareto efficient pollution levels $\widetilde{p} \in \mathbb{R}^N_+$ in the river pollution model (N, b, c). Payoff vector $z(p, t) \in \mathbb{R}^N$ at pollution levels vector $p \in \mathbb{R}^N$ and compensation scheme $t \in \mathbb{R}^N$, is Pareto efficient if there does not exist another pair (p', t') such that $z_i(p', t') \geq z_i(p, t)$ for all $i \in N$ with at least one strict inequality. Clearly, $z(p, t)$ is Pareto efficient if and only if $p = \widetilde{p}$ and $\sum_{i \in N} t_i = 0$, and thus $\sum_{i \in N} z_i(p, t) = W(\widetilde{p}) = W(N, b, c)$. It therefore follows that any Pareto efficient payoff vector $z \in \mathbb{R}^N$ can be implemented by the vector $\widetilde{p} \in \mathbb{R}^N_+$ of efficient pollution levels and the budget balanced compensation scheme $t_i = z_i - u_i(\widetilde{p})$, $i \in N$. We conclude this section with an example, which also will be used to illustrate the discussion in the subsequent sections.

Example 2.5 Let (N, b, c) be a river pollution model with $N = \{1, 2\}$, $b_i(p_i) = \sqrt{p_i}$ and $c_i(q_i) = q_i^2$, $i = 1, 2$. Then the Nash equilibrium pollution levels are given by $\widehat{p}_1 = 0.3969$ and $\widehat{p}_2 = 0.1847$, yielding utilities $u_1(\widehat{p}) = 0.473$ for the upstream agent 1 and $u_2(\widehat{p}) = 0.092$ for the downstream agent 2. The social welfare in the Nash equilibrium is $W(\widehat{p}) = 0.565$. The Pareto efficient pollution levels are $\widetilde{p}_1 = 0.1621$ and $\widetilde{p}_2 = 0.2968$, yielding utilities $u_1(\widetilde{p}) = 0.376$ and $u_2(\widetilde{p}) = 0.334$. Notice that indeed $\widetilde{p}_1 + \widetilde{p}_2 = 0.4589 < 0.5816 = \widehat{p}_1 + \widehat{p}_2$. The maximal social welfare is equal to $W(\widetilde{p}) = 0.710$.

Observe that $u_1(\widetilde{p}) = 0.376 < 0.473 = u_1(\widehat{p})$, so that without monetary transfers agent 1 prefers the Nash equilibrium output above the Pareto efficient output. When $t_1 = -t_2$ and $0.097 \leq t_1 \leq 0.242$ both agents have at least the same payoff in the Pareto efficient output $(\widetilde{p}, t)$ as at the Nash equilibrium pollution levels $\widehat{p}$ without monetary compensations. □

3 Distribution of cooperative gains

In the previous section we have seen that the agents in a river pollution model are able to realize the maximum social welfare $W(N, b, c)$ by choosing the Pareto efficient pollution levels $\widetilde{p}_i$, $i \in N$. In this section we discuss, for the two agent case, what compensation schemes $t = (t_1, t_2)$ would allow the agents to sustain these Pareto efficient pollution levels. In particular, in Example 2.5 the Pareto efficient pollution levels $\widetilde{p}_1$ and $\widetilde{p}_2$, together with a monetary compensation scheme $t = (t_1, t_2)$ such that $0.097 \leq t_1 \leq 0.242$ and $t_2 = -t_1$, yield both agents a payoff that is at least equal to its Nash equilibrium payoff. A question that can now be asked is the following: is it reasonable to restrict the value of t_1 between 0.097 and 0.242?

According to Coase (1960) the answer to this question depends on the allocation of property rights. The well-known Coase theorem states that when trade in an externality (pollution caused by the upstream agent to the downstream agent) is possible and there are no transaction costs, bargaining leads to an efficient outcome, regardless of the initial allocation of property rights. It are exactly the property rights that determine how the welfare gain from cooperation is distributed among the agents. For the two-agent river pollution model the Coase theorem implies that cooperation leads to the Pareto efficient pollution levels $p_i = \widetilde{p}_i$, $i = 1, 2$. The transfers t_1 and t_2 then determine how the maximal social welfare $W(N, b, c)$ is distributed over the two agents.

When the upstream agent 1 has the property rights over the river it can cause as much pollution as it pleases, without taking into account the harmful consequences this might have for the downstream agent 2. It thus can be argued that when agent 1 has the property rights over the river it has a legitimate claim to a payoff that is at least equal to the payoff it obtains in the Nash equilibrium output $z_1(\widehat{p}, \mathbf{0}) = u_1(\widehat{p})$. In this case agent 1 would only be willing to cooperate with agent 2, and pollute at its Pareto efficient pollution level, if it receives a monetary compensation t_1 that is at least equal to $u_1(\widehat{p}) - u_1(\widetilde{p}) = \Big(b_1(\widehat{p}_1) - c_1(\widehat{p}_1)\Big) - \Big(b_1(\widetilde{p}_1) - c_1(\widetilde{p}_1)\Big)$. On the other hand, when agent 1 has the property rights over the river, agent 2 knows that without cooperation agent 1 would pollute at its Nash equilibrium level. Then the optimal action of agent 2 is also to pollute at its Nash equilibrium level. Hence, agent 2 would not be willing cooperate with agent 1, and make a monetary transfer, if this would lead to a payoff below its payoff in the Nash equilibrium $z_2(\widehat{p}, \mathbf{0}) = u_2(\widehat{p})$. Thus, the compensation $t_1 = -t_2$ that agent 2 is willing to pay is at most equal to $u_2(\widetilde{p}) - u_2(\widehat{p}) = \Big(b_2(\widetilde{p}_2) - c_2(\widetilde{p}_1 + \widetilde{p}_2)\Big) - \Big(b_2(\widehat{p}_2) - c_2(\widehat{p}_1 + \widehat{p}_2)\Big)$. It can be concluded that when agent 1 has the property rights over the river, the agents are willing to bargain on a transfer t_1 between $\Big(b_1(\widehat{p}_1) - c_1(\widehat{p}_1)\Big) - \Big(b_1(\widetilde{p}_1) - c_1(\widetilde{p}_1)\Big)$ and $\Big(b_2(\widetilde{p}_2) - c_2(\widetilde{p}_1 + \widetilde{p}_2)\Big) - \Big(b_2(\widehat{p}_2) - c_2(\widehat{p}_1 + \widehat{p}_2)\Big)$. In Example 2.5 this bargaining interval is $0.097 \leq t_1 \leq 0.242$.

We now consider the case that the downstream agent 2 has the property rights over the river in the sense that it has the right to claim (and the possibility to enforce) that agent 1 does not cause any pollution, thus that $p_1 = 0$. In this case agent 2 can claim a minimal payoff equal to $z_2((0, r_2), \mathbf{0}) = u_2((0, r_2)) = b_2(r_2) - c_2(r_2)$ (recall that r_i, $i \in N$, is the optimal pollution level of agent i when all other pollution levels are zero). Now, agent 2 is only willing to cooperate with agent 1, and set its Pareto efficient pollution level $\widetilde{p}_2$, if agent 1 pays a monetary transfer $t_2 = -t_1$ that is at least equal to $u_2((0, r_2)) - u_2(\widetilde{p}) = \Big(b_2(r_2) - c_2(r_2)\Big) - \Big(b_2(\widetilde{p}_2) - c_2(\widetilde{p}_1 + \widetilde{p}_2)\Big)$. On the other hand, when agent 2 has the property rights over the river, without cooperation agent 1 has a payoff equal to zero $z_1((0, r_2), \mathbf{0}) = u_1((0, r_2)) = 0$. Agent 1 would therefore not be willing to pay more than $u_1(\widetilde{p}) - u_1((0, r_2)) = b_1(\widetilde{p}_1) - c_1(\widetilde{p}_1)$ to establish cooperation. It can be concluded that when agent 2 has the property rights over the river the agents are willing to bargain on a transfer t_2 between $\Big(b_2(r_2) - c_2(r_2)\Big) - \Big(b_2(\widetilde{p}_2) - c_2(\widetilde{p}_1 + \widetilde{p}_2)\Big)$ and $b_1(\widetilde{p}_1) - c_1(\widetilde{p}_1)$. For Example 2.5 it follows straightforwardly that $b_2(r_2) - c_2(r_2) = 0.473$; so the bargaining interval is $0.139 \leq t_2 \leq 0.376$.

When the property rights over the river are unambiguously defined it follows from the above that, at least in the two agent case, a well-defined bargaining problem is obtained. Every solution to such a bargaining problem results in a distribution of the cooperative gains. However, the bargaining problem is not so obvious when the property rights over the river are not clearly defined. For instance, what would be the output of the two agent river pollution model when both agents claim to have the property rights over the river and neither of the agents accepts the claim of the other agent? In this situation each agent $i \in \{1, 2\}$ claims a payoff that is at least equal to $b_i(r_i) - c_i(r_i)$. In Example 2.5 this would mean that both agents claim at least 0.473. Since the total social welfare (of cooperation) is equal to 0.565, an outcome in which both agents obtain at least their claim is infeasible. This leads to the question how to distribute the deficit that results when each agent claims the property rights over the river. In the following sections we suggest answers to this question by taking into account principles from international water law.

4 Values for the river pollution model

4.1 Preliminaries

In this section we propose and characterize two solutions for the welfare distribution problem within the river pollution model (N, b, c). To do this we use the concept of a value from the theory of cooperative games, see e.g. Shapley (1953). Within this theory, a value is a function on a class of (cooperative) games that assigns to each game in the class a payoff vector, i.e., a vector that specifies a payoff to every player in the game. To apply this notion to polluted rivers, let $\mathcal{RP}^N$ be the class of all river pollution models (N, b, c) with fixed set of agents N satisfying Assumption 2.1. Further, let $\mathcal{RP} = \cup_{N \subset \mathbb{N}} \mathcal{RP}^N$ be the class of all river pollution models over all sets $N \subset \mathbb{N}$. A value now is a function f that assigns to every $(N, b, c) \in \mathcal{RP}$ a payoff vector $f(N, b, c) \in \mathbb{R}^N$. Typically a value is defined axiomatically, that is, a number of desirable axioms (properties) is stated and then it is shown that there exists a unique value that satisfies these axioms.

Ideally, we would base our values for the river pollution model directly on international watercourse law. But, since there currently is no binding international law for managing international rivers, the only guidelines that are available to us are international water doctrines from the (legal) literature (see for instance, Garretson, Hayton and Olmstead (1967) or McCaffrey (2001)). Two of these principles, used by Ambec and Sprumont (2002) for river situations concerning the rival consumption of water and by Ni and Wang (2007) for allocating the costs of cleaning a river from pollution, are the principle of Absolute Territorial Sovereignty (ATS) and the principle of Unlimited Territorial Integrity (UTI). Here, we first apply these two principles to our class $\mathcal{RP}$ of river pollution models.

4.2 The ATS value

The principle of Absolute Territorial Sovereignty (also known as the Harmon doctrine) states that each country (agent) along an international river has absolute sovereignty over the part of the river on its territory (McCaffrey, 2001). For river pollution models the ATS principle favors upstream agents over downstream agents in the sense that it allows an (upstream) agent to choose any pollution level it prefers, without taking into account the consequences for downstream agents. It is not difficult to see that without cooperation, the ATS principle would yield the Nash equilibrium output. We can, however, also apply the ATS principle when the agents along the river do cooperate. As observed in the previous section, the Coase theorem implies that under cooperation all agents pollute at their Pareto efficient pollution level. It are the property rights that determine how the welfare gain from cooperation is distributed among the agents. As in Ambec and Sprumont (2002), we propose that the property rights over an international river are determined by international watercourse principles.

When a group of upstream agents P^i decides to cooperate[5], the ATS principle implies that such a group of agents can pollute as much as it pleases because it has absolute sovereignty over its territory. So, every upstream set of agents P^i can claim a total (combined) payoff under full cooperation (of all agents) that is at least equal to the total welfare that it can attain on its own. If it would not receive at least this welfare level, it would be optimal for the group to cease cooperation with the downstream agents. Let p^i_j, $j \in P^i$, $i \in N$, be a solution to the maximization problem

$$\max_{p_1,\dots,p_i} \sum_{j=1}^{i} \Big(b_j(p_j) - c_j(\sum_{k=1}^{j} p_k)\Big) \tag{4.7}$$

and denote

$$v^i(N,b,c) = \sum_{j=1}^{i} \Big(b_j(p^i_j) - c_j(\sum_{k=1}^{j} p^i_k)\Big).$$

that is, $v^i(N,b,c)$ is the highest welfare that the set of upstream agents P^i can obtain without taking into account the consequences of its pollution to the downstream agents.

[5] Recall from Section 2 that $P^i = \{1,\dots,i\}$, $i \in N$, is the set of all agents upstream of, and including, agent i.

Notice that $v^n(N,b,c) = W(N,b,c)$. The ATS principle thus implies that each group of upstream agents P^i, $i \in N$, can claim at least a total payoff $v^i(N,b,c)$.

We now define the *ATS value*, denoted by ATS, as the function on the class $\mathcal{RP}$ of river pollution models that for every $N \subset \mathbb{N}$ and every $(N,b,c) \in \mathcal{RP}^N$ assigns to every agent $j \in N$ the payoff $ATS_j(N,b,c)$ equal to

$$ATS_j(N,b,c) = v^j(N,b,c) - v^{j-1}(N,b,c),$$

with $v^0(N,b,c) = 0$. So, the ATS value distributes to every upstream set of agents P^i, $i \in N$, a total payoff equal to $\sum_{j=1}^{i} ATS_j(N,b,c) = v^i(N,b,c)$, i.e., every set of upstream agents P^i receives precisely the minimum payoff it can claim according to the ATS principle. The ATS value can be implemented by the Pareto efficient pollution levels $\widetilde{p}_i$, $i \in N$, and a budget balanced compensation scheme t such that $t_i = ATS_i(N,b,c) - u_i(\widetilde{p})$, $i \in N$.

In the sequel, for any river pollution model $(N,b,c) \in \mathcal{RP}$ and some agent $i \in N$, let $(P^i, b^{1,i}, c^{1,i})$ denote the river pollution model restricted to the upstream set of agents P^i. Then, $(P^i, b^{1,i}, c^{1,i})$ is a river problem in $\mathcal{RP}^{P_i}$ with set of agents P^i, benefit functions $b_j^{1,i} = b_j$, $j \in P^i$, and cost functions $c_j^{1,i} = c_j$, $j \in P^i$. Notice that for every $i \in N$,

$$W(P^i, b^{1,i}, c^{1,i}) = v^i(N,b,c),$$

i.e., the worth $v^i(N,b,c)$ that the agents in P^i can guarantee themselves under the ATS principle within the river pollution model (N,b,c) is equal to the total social welfare that P^i can attain within the (sub)river problem $(P^i, b^{1,i}, c^{1,i})$. Hence the ATS value satisfies

$$\sum_{j=1}^{i} ATS_j(N,b,c) = W(P^i, b^{1,i}, c^{1,i}), \text{ for all } i \in N. \tag{4.8}$$

Using this it follows that the ATS value is characterized by an *efficiency* and an *upstream autonomy* axiom.

Axiom 4.1 Efficiency
A value f on the class of river pollution models $\mathcal{RP}$ is efficient if it holds for every $(N,b,c) \in \mathcal{RP}$ that $\sum_{i \in N} f_i(N,b,c) = W(N,b,c)$.

In cooperative game theory efficiency axioms are considered as such basic axioms that some authors include them in the definition of a value. In our model efficiency follows from the Coase theorem. As stated before, the Coase theorem implies that all agents pollute at their Pareto efficient pollution levels and the property rights determine how the maximum social welfare $W(N,b,c)$ is distributed over the agents.

Axiom 4.2 Upstream autonomy
A value f on the class of river pollution models $\mathcal{RP}$ satisfies upstream autonomy if for every $(N,b,c) \in \mathcal{RP}$ and any $i \in N$ it holds that $f_i(N,b,c) = f_i(P^i, b^{1,i}, c^{1,i})$.

When all agents downstream of i are not present, upstream autonomy implies that agent i receives the same payoff as it would receive when these agents are present. So, it states that the payoff of an agent does not depend on its downstream agents. We now can state and prove the following characterization theorem for the ATS value.

Theorem 4.3 *A value f on the class of river pollution models $\mathcal{RP}$ satisfies efficiency and upstream autonomy if and only if f is the ATS value.*

Proof. We first show that the ATS value satisfies the two axioms. Efficiency follows straightforwardly from the definition of ATS, since $\sum_{i \in N} ATS_i(N,b,c) = W(N,b,c)$. Upstream autonomy follows straightforwardly from equation (4.8), because for every $i \in N$,

$$ATS_i(N,b,c) = \sum_{j=1}^{i} ATS_j(N,b,c) - \sum_{j=1}^{i-1} ATS_j(N,b,c) =$$
$$W(P^i, b^{1,i}, c^{1,i}) - W(P^{i-1}, b^{1,i-1}, c^{1,i-1}) = ATS_i(P^i, b^{1,i}, c^{1,i}).$$

Next, take $(N,b,c) \in \mathcal{RP}$ and assume that f satisfies efficiency and upstream autonomy. We prove uniqueness by induction on the labels of the agents, starting with the most upstream agent 1. For $i = 1$ we have by upstream autonomy that $f_1(N,b,c) = f_1(P^1, b^{1,1}, c^{1,1})$, thus the payoff of agent 1 in the $|N|$-agent river problem (N,b,c) is equal to the payoff of agent 1 in the 1-agent river problem $(P^1, b^{1,1}, c^{1,1})$. By efficiency we have that $f_1(P^1, b^{1,1}, c^{1,1}) = W(P^1, b^{1,1}, c^{1,1})$. So, $f_1(N,b,c) = ATS_1(N,b,c)$. Now, assume by induction that $f_k(N,b,c) = ATS_k(N,b,c)$ for all $k < i \leq n$. Then

$$f_i(N,b,c) = f_i(P^i, b^{1,i}, c^{1,i}) = W(P^i, b^{1,i}, c^{1,i}) - \sum_{k=1}^{i-1} f_k(P^i, b^{1,i}, c^{1,i}),$$

where the first equality follows from upstream autonomy and the second from efficiency. Since, again by upstream autonomy, $f_k(P^i, b^{1,i}, c^{1,i}) = f_k(N,b,c)$ it follows by the induction hypotheses and equation (4.8) that

$$f_i(N,b,c) = W(P^i, b^{1,i}, c^{1,i}) - \sum_{k=1}^{i-1} f_k(N,b,c) = W(P^i, b^{1,i}, c^{1,i}) - \sum_{k=1}^{i-1} ATS_k(N,b,c) =$$
$$W(P^i, b^{1,i}, c^{1,i}) - W(P^{i-1}, b^{1,i-1}, c^{1,i-1}) = ATS_i(N,b,c).$$

□

For the two agent river pollution model (N,b,c) with $N = \{1,2\}$ the ATS value gives the payoffs

$$ATS_1(N,b,c) = W(P^1, b^{1,1}, c^{1,1}) = b_1(r_1) - c_1(r_1) = u_1(r)$$

and

$$ATS_2(N,b,c) = W(N,b,c) - ATS_1(N,b,c) = W(N,b,c) - u_1(r).$$

So, upstream agent 1 receives a payoff equal to its Nash equilibrium payoff and all gains from cooperation go to the downstream agent 2. The ATS value in this instance corresponds to the outcome discussed in Section 3 in which agent 1 has the property rights over the river and agent 2 pays the minimum possible transfer to agent 1 in order for it to be compensated for its loss in utility when switching from the Nash equilibrium to the Pareto efficient pollution level. In Example 2.5 the ATS value would mean that agent 2 pays $t_1 = 0.097$ to agent 1.

For the case with more than two agents the upstream autonomy axiom implies that property rights are assigned subsequently from upstream to downstream along the river. First agent 1 has the right to choose its optimal pollution level, regardless of the other agents. Then agents 1 and 2 cooperate and have the right to choose their joint optimal pollution levels, without considering the other agents, and so on. The ATS value assigns at each step the gain of cooperation between the agents in P^{i-1} and the next agent i to agent i, $i = 2, ..., n$. So, each time an agent i joins its set of upstream agents P^{i-1} all the gain of cooperation goes to agent i and the upstream agents are just compensated to keep their payoffs equal.

The next theorem states that the ATS value gives each agent $i \in N$ a payoff that is at least equal to the payoff it would receive in the Nash equilibrium output. Each agent in a river pollution model therefore weakly prefers its payoff according to the ATS value to its payoff in the Nash equilibrium output.

Theorem 4.4 *Let $(N,b,c) \in \mathcal{RP}$ be a river pollution model satisfying Assumption 2.1. Then, for any $i \in N$, $ATS_i(N,b,c) \geq z_i(\widehat{p}, \mathbf{0}) = u_i(\widehat{p})$.*

Proof. For agent 1 the theorem is true by definition of the ATS value. Next consider some agent $\ell \geq 2$ and take $i = \ell - 1$. Note that $W(P^i, b^{1,i}, c^{1,i}) = \sum_{j=1}^{i} \Big(b_j(p_j^i) - c_j(\sum_{k=1}^{j} p_k^i)\Big)$, where p_j^i, $j \in P^i$, is a solution to the maximization problem (4.7). Let $\bar{p}_\ell$ be the optimal pollution level of agent ℓ, given that all its upstream agents $j \leq \ell - 1$ choose p_j^i. This yields utility $\bar{u}_\ell = b_\ell(\bar{p}_\ell) - c_\ell(\sum_{k=1}^{i} p_k^i + \bar{p}_\ell)$ to agent ℓ. By definition of the ATS value it follows that

$$ATS_\ell(N,b,c) = W(P^\ell, b^{1,\ell}, c^{1,\ell}) - W(P^{\ell-1}, b^{1,\ell-1}, c^{1,\ell-1}) \geq \bar{u}_\ell.$$

Further, applying Proposition 2.4 to the river pollution model $(P^i, b^{1,i}, c^{1,i})$ it follows that $\sum_{j=1}^{i} p_j^i < \sum_{j=1}^{i} \widehat{p}_j$. Hence

$$ATS_\ell(N,b,c) \geq \bar{u}_\ell > b_\ell(\widehat{p}_\ell) - c_\ell(\sum_{k=1}^{\ell} \widehat{p}_k) = u_\ell(\widehat{p}).$$

□

To conclude this subsection we would like to mention that for a river model with rival consumption of the river water, Ambec and Sprumont (2002) propose a solution similar to the ATS value, called the downstream incremental distribution (or value). This name refers to the fact that all the gains of cooperation between an upstream group of agents P^i, $i \in N$, and the subsequent agent along the river $i+1$ are distributed to the agent $i+1$. For the same model, Herings, van der Laan and Talman (2007) and van den Brink, van der Laan and Vasil'ev (2007) alternatively propose the upstream incremental solution (or value). The upstream incremental solution is also based on the ATS principle but distributes the gains of cooperation between the set $Q_i = \{i, \ldots, n\}$, $i > 1$, of downstream agents and the preceding agent along the river $i-1$ to the agent $i-1$. This approach cannot be followed for our river pollution model with non-rival use of the water because the welfare that a downstream group of agents can obtain without cooperating with its upstream agents is not unambiguously defined (it depends on the behavior of the upstream agents). Instead, we therefore consider the principle of Unlimited Territorial Integrity to define a counterpart of the ATS value.

4.3 The UTI value

The principle of Unlimited Territorial Integrity states that each country (agent) along an international river has the right to demand the natural flow of the river into its territory that is both undiminished in quantity and unchanged in quality by the countries (agents) upstream to it (McCaffrey, 2001). For river pollution models the UTI principle favors downstream agents over upstream agents in the sense that an (upstream) agent is only allowed to pollute the river if it has the explicit consent of all agents downstream to it. When the downstream agents in Q_i decide to cooperate, the UTI principle implies that such a group of agents can claim a completely clean river. This means that none of the agents upstream of the group Q_i is allowed to cause any pollution. Thus, in a river pollution model $(N, b, c) \in \mathcal{RP}$ any group of downstream agents Q_i can claim a total (combined) payoff under full cooperation (of all agents) that is at least equal to the total welfare that Q_i can attain under the condition that all upstream agents $j < i$ set pollution level $p_j = 0$. If it would not receive at least this welfare level under full cooperation it would be optimal for the group of downstream agents to cease cooperation with the upstream agents and invoke the UTI principle. Let γ_j^i, $j \in Q_i$, $i \in N$, be a solution to the maximization problem

$$\max_{p_i, \ldots, p_n} \sum_{j=i}^{n} \Big(b_j(p_j) - c_j(\sum_{k=i}^{j} p_k) \Big) \tag{4.9}$$

and denote

$$w^i(N,b,c) = \sum_{j=i}^{n} \Big(b_j(\gamma^i_j) - c_j(\sum_{k=i}^{j} \gamma^i_k)\Big).$$

That is, $w^i(b,c)$ is the highest welfare that the downstream group Q_i can obtain under the condition that the pollution levels of all the upstream agents are equal to zero. Notice that $w^1(N,b,c) = W(N,b,c)$. The UTI principle implies that each group of downstream agents Q_i, $i \in N$, can claim at least a total payoff $w^i(N,b,c)$.

We now define the *UTI value*, denoted by UTI, as the function on the class $\mathcal{RP}$ of river pollution models that for every $N \subset \mathbb{N}$ and every $(N,b,c) \in \mathcal{RP}^N$ assigns to every agent $j \in N$ payoff $UTI_j(N,b,c)$ equal to

$$UTI_j(N,b,c) = w^j(N,b,c) - w^{j+1}(N,b,c),$$

with $w^{n+1}(N,b,c) = 0$. So, the UTI value distributes to every downstream set of agents Q_i, $i \in N$, a total payoff equal to $\sum_{j=i}^{n} UTI_j(N,b,c) = w^i(N,b,c)$, i.e., every set of downstream agents Q_i receives precisely the minimum payoff it can claim according to the UTI principle. The UTI value can be implemented by the Pareto efficient pollution levels $\widetilde{p}_i$, $i \in N$, and a budget balanced compensations scheme t such that $t_i = UTI_i(N,b,c) - u_i(\widetilde{p})$, $i \in N$.

In the sequel, for any river pollution model $(N,b,c) \in \mathcal{RP}$ and some agent $i \in N$, let $(Q_i, b^{i,n}, c^{i,n})$ denote the river pollution model restricted to the downstream set of agents Q_i. Then, $(Q_i, b^{i,n}, c^{i,n})$ is a river problem in $\mathcal{RP}^{Q_i}$ with set of agents Q_i, benefit functions $b^{i,n}_j = b_j$, $j \in Q_i$, and cost functions $c^{i,n}_j = c_j$, $j \in Q_i$.[6] Notice that for every $i \in N$,

$$W(Q_i, b^{i,n}, c^{i,n}) = w^i(N,b,c),$$

i.e., the worth $w^i(N,b,c)$ that the agents in Q_i can guarantee themselves under the UTI principle within the river pollution model (N,b,c) is equal to the total social welfare that Q_i can attain within the (sub)river problem $(Q_i, b^{i,n}, c^{i,n})$. Hence the UTI value satisfies

$$\sum_{j=i}^{n} UTI_j(N,b,c) = W(Q_i, b^{i,n}, c^{i,n}), \text{ for all } i \in N. \qquad (4.10)$$

Using this it follows that the UTI value is characterized by the efficiency axiom and an *downstream autonomy* axiom.

Axiom 4.5 Downstream autonomy
A value f on the class of river pollution models $\mathcal{RP}$ satisfies downstream autonomy if for every $(N,b,c) \in \mathcal{RP}$ and any $i \in N$ it holds that $f_i(N,b,c) = f_i(Q_i, b^{i,n}, c^{i,n})$.

[6] Otherwise than assumed until now, in this river problem the agents are numbered from i to n.

When all agents upstream of i are not present, downstream autonomy implies that agent i receives the same payoff as it would receive when these agents are present. So, downstream autonomy states that the payoff of an agent does not depend on its upstream agents. We now can state and prove the following characterization theorem for the UTI value.

Theorem 4.6 *A value f on the class of river pollution models $\mathcal{RP}$ satisfies efficiency and downstream autonomy if and only if f is the UTI value.*

Proof. We first show that the UTI value satisfies the two axioms. Efficiency follows straightforwardly from the definition of UTI, since $\sum_{i\in N} UTI_i(N,b,c) = W(N,b,c)$. Downstream autonomy follows straightforwardly from equation (4.10), because for every $i \in N$,

$$UTI_i(N,b,c) = \sum_{j=i}^{n} UTI_j(N,b,c) - \sum_{j=i+1}^{n} UTI_j(N,b,c) =$$

$$W(Q_i, b^{i,n}, c^{i,n}) - W(Q_{i+1}, b^{i+1,n}, c^{i+1,n}) = UTI_i(Q_i, b^{i,n}, c^{i,n}).$$

Next, take $(N,b,c) \in \mathcal{RP}$ and assume that f satisfies efficiency and downstream autonomy. We prove uniqueness by induction on the labels of the agents, starting with the most downstream agent n. For $i = n$ we have by downstream autonomy that $f_n(N,b,c) = f_n(Q_n, b^{n,n}, c^{n,n})$. So the payoff of agent n in the $|N|$-agent river problem (N,b,c) is equal to the payoff of agent n in the 1-agent river problem $(Q_n, b^{n,n}, c^{n,n})$. By efficiency we have that $f_n(Q_n, b^{n,n}, c^{n,n}) = W(Q_n, b^{n,n}, c^{n,n})$ and thus $f_n(N,b,c) = UTI_n(N,b,c)$. Now, assume by induction that $f_k(N,b,c) = UTI_k(N,b,c)$ for all $k > i \geq 1$. Then

$$f_i(N,b,c) = f_i(Q_i, b^{i,n}, c^{i,n}) = W(Q_i, b^{i,n}, c^{i,n}) - \sum_{k=i+1}^{n} f_k(Q_i, b^{i,n}, c^{i,n}),$$

where the first equality follows from downstream autonomy and the second from efficiency. Since, again by downstream autonomy, $f_k(Q_i, b^{i,n}, c^{i,n}) = f_k(N,b,c)$ it follows by the induction hypothesis and equation (4.10) that

$$f_i(N,b,c) = W(Q_i, b^{i,n}, c^{i,n}) - \sum_{k=i+1}^{n} f_k(N,b,c) = W(Q_i, b^{i,n}, c^{i,n}) - \sum_{k=i+1}^{n} UTI_k(N,b,c) =$$

$$W(Q_i, b^{i,n}, c^{i,n}) - W(Q_{i+1}, b^{i+1,n}, c^{i+1,n}) = UTI_i(N,b,c).$$

□

For the two agent river pollution model (N,b,c) with $N = \{1,2\}$ the UTI value gives the payoffs

$$UTI_2(N,b,c) = W(Q_2, b^{2,2}, c^{2,2}) = b_2(r_2) - c_2(r_2) = u_2((0,r_2))$$

and

$$UTI_1(N,b,c) = W(N,b,c) - UTI_2(N,b,c) = W(N,b,c) - u_2((0,r_2)).$$

So, downstream agent 2 receives a payoff equal to the minimal payoff it can achieve when it can claim (and enforce) that agent 1 does not cause any pollution. The UTI value in this instance corresponds to the outcome discussed in Section 3 in which agent 2 has the property rights over the river and agent 1 pays the minimum possible transfer to agent 2 in order for it to be compensated for its loss in utility when agent 2 gives up its right to a clean river and agrees to cooperate on the Pareto efficient pollution levels. In Example 2.5 the UTI value would mean that agent 1 pays $t_2 = 0.139$ to agent 2.

For the case with more than two agents the downstream autonomy axiom implies that the property rights are assigned subsequently from downstream to upstream along the river. First agent n has the right to claim clean water and choose its optimal pollution level under the restriction that all upstream levels are zero. Then the agents n and $n-1$ together cooperate and have the right to their joint optimal pollution levels under the restriction that all upstream levels are zero, and so on. The UTI value assigns at each step the gain in welfare when the agents in Q_{i+1} share their UTI rights with the upstream neighboring agent i, to agent i, $i = 1, ..., n-1$. So, each time an agent i joins its set of downstream agents Q_{i+1} all the increase in total welfare goes to agent i and the downstream agents are just compensated to keep their payoffs equal.

In Theorem 4.4 we have seen that the ATS value assigns to each agent a payoff that is at least equal to the utility it would receive in the Nash equilibrium output. This does not hold for the UTI value, as can be seen from Example 2.5. The UTI value, however, does satisfy a property that is not satisfied by the ATS value: it guarantees that all agents receive a non-negative payoff. To see that the ATS value does not guarantee non-negative payoffs, consider a two agent river pollution model and suppose that agent 2 has much higher costs of pollution than agent 1. Then it could be that $W(N,b,c) = [b_1(\widetilde{p}_1) - c_1(\widetilde{p}_1)] + [b_2(\widetilde{p}_2) - c_2(\widetilde{p}_1 + \widetilde{p}_2)]$ is smaller than $W(P_1, b^{1,1}, c^{1,1}) = b_1(r_1) - c_1(r_1)$ which would mean that $ATS_2 = W(N,b,c) - W(P_1, b^{1,1}, c^{1,1}) < 0$. Agent 2, however, would still be willing to cooperate with agent 1 because its ATS payoff is at least equal to its Nash equilibrium payoff. The next theorem shows the all UTI payoffs are non-negative.

Theorem 4.7 *Let $(N,b,c) \in \mathcal{RP}$ be a river pollution model satisfying Assumption 2.1. Then $UTI_i(N,b,c) \geq 0$ for every $i \in N$.*

Proof. For agent n the theorem is true, because $UTI_n(N,b,c) = b_n(r_n) - c_n(r_n) > 0$. Next consider some agent $i \leq n-1$. According to the UTI value this agent receives $UTI_i(N,b,c) = W(Q_i, b^{i,n}, c^{i,n}) - W(Q_{i+1}, b^{i+1,n}, c^{i+1,n})$. Since the pollution levels $p_i = 0$

and $p_j = \gamma_j^{i+1}$ for $j > i$ are feasible for the maximization problem (4.9) with respect to agent i, and the levels $p_j = \gamma_j^{i+1}$, $j > i$ are a solution for the maximization problem (4.9) with respect to agent $i+1$, it follows that $W(Q_i, b^{i,n}, c^{i,n}) \geq W(Q_{i+1}, b^{i+1,n}, c^{i+1,n})$, so that $UTI_i(N, b, c) \geq 0$. □

As already mentioned at the end of Section 3 for the two agent case, the ATS and UTI values are incompatible. In Example 2.5 both the ATS claim $W(P_1, b^{1,1}, c^{1,1})$ of agent 1 and the UTI claim $W(Q_2, b^{2,2}, c^{2,2})$ of agent 2 are equal to 0.473, while the social welfare is equal to 0.565. This means that it is impossible to satisfy both claims simultaneously. In general, it holds for every $i \in \{1, ..., n-1\}$ that the sum of the ATS claim $W(P_i, b^{1,i}, c^{1,i})$ of the upstream set of agents P^i and the UTI claim $W(Q_{i+1}, b^{i+1,n}, c^{i+1,n})$ of its downstream complement Q_{i+1} exceeds the maximal total available welfare $W(N, b, c)$. In the next section we therefore discuss compromise solutions.

5 TIBS values

In the previous section we have seen that the ATS principle allocates the property rights over a river to the upstream agents and that the UTI principle allocates them to the downstream agents. In this section we propose values for river pollution models that force both the upstream and the downstream agents along the river to make concessions. To introduce these values, consider a river pollution model $(N, b, c) \in \mathcal{RP}$ and an agent $j \in N$. Suppose that all agents along the river pollute at their Pareto efficient level $\widetilde{p}_i$, $i \in N$, and that each agent upstream of agent j is given its ATS value payoff while each agent downstream of agent j is given its UTI value payoff. Since the agents along the river are maximally able to divide the maximum social welfare $W(N, b, c)$ among themselves, if one would like to obtain an efficient payoff distribution for the river pollution model (N, b, c), it must be that agent j receives (pays) the entire surplus (deficit) $W(N, b, c) - \sum_{k \in P^{j-1}} ATS_k(N, b, c) - \sum_{k \in Q_{j+1}} UTI_k(N, b, c)$. More formally, for all $i, j \in N$ let $t_i^j(N, b, c)$ be defined as

$$t_i^j(N,b,c) = \begin{cases} ATS_i(N,b,c) & \text{if } i < j, \\ W(N,b,c) - \sum_{k \in P^{i-1}} ATS_k(N,b,c) - \sum_{k \in Q_{i+1}} UTI_k(N,b,c) & \text{if } i = j, \\ UTI_i(N,b,c) & \text{if } i > j. \end{cases}$$

In this way each $j \in N$ induces the value t^j on the class $\mathcal{RP}$ of river pollution models, that assigns to each $(N, b, c) \in \mathcal{RP}$ the payoff vector $t^j(N, b, c) \in \mathbb{R}^N$.[7] Note that for $j = 1$ it holds that $t^1(N, b, c) = UTI(N, b, c)$ and for $j = n$ that $t^n(N, b, c) = ATS(N, b, c)$. It is

[7]The value t^j resembles the value for games on cycle-free graph structures of Demange (2004) in which agent j is the top agent in a hierarchy.

not difficult to see that the value t^j can result in a very large negative payoff $t^j_j(N,b,c) = W(N,b,c) - \sum_{k\in P^{j-1}} ATS_k(N,b,c) - \sum_{k\in Q_{j+1}} UTI_k(N,b,c)$ for agent j. In the sequel we are going to consider weighted averages of the values t^j, $j \in N$.

Let $\alpha \in \mathbb{R}^N_+$ with $\sum_{i\in N} \alpha_i = 1$ and α_j the *weight* of agent $j \in N$. Given such a vector of weights α, the *TIBS*$^\alpha$ *value*, denoted by $TIBS^\alpha$, is defined as the function on the class $\mathcal{RP}$ of river pollution models that for every $N \subset \mathbb{N}$ and every $(N,b,c) \in \mathcal{RP}^N$ assigns to every agent $i \in N$ payoff $TIBS^\alpha_i(N,b,c)$ equal to

$$TIBS^\alpha_i(N,b,c) = \sum_{j\in N} \alpha_j t^j_i(N,b,c).$$

We call the α-weighted average of the values t^j, $j \in N$, the TIBS$^\alpha$ value because it can be seen as reflecting a water sharing principle known as the principle of Territorial Integration of all Basin States (TIBS).

The TIBS principle states that the water in an international river belongs to all riparian states combined, no matter where it enters the river, and that each riparian state is entitled to a reasonable and equitable share in the optimal use of the river water (McCaffrey, 2001). In the (legal) literature on the subject it is also referred to as the principle of community (of interests) in the waters, the principle of common management or the drainage basin approach. For river pollution models the TIBS principle does not seem to favor upstream agents over downstream agents, or vice versa. It only requires that the agents make optimal use of the river (pollute at their Pareto efficient levels) and share the social welfare that results in a 'reasonable and equitable' manner.

Obviously, the terms 'reasonable' and 'equitable' are not precise. To make them precise, let f be a value on the class $\mathcal{RP}$ of all river problems and compare the following two situations for some agent $i \in N\setminus\{n\}$. In the first situation the agents in P^i are cooperating together and the agents in Q_{i+1} are cooperating together, but the two sets of agents are not cooperating with each other. That is, the agents in P^i are cooperating and claim to have the right to pollute according to the ATS principle. In this situation each agent $j \in P^i$ obtains payoff $f_j(P^i, b^{1,i}, c^{1,i})$ assigned by the value f to the upstream (sub)river problem $(P^i, b^{1,i}, c^{1,i})$. On the other hand, the agents in Q_{i+1} cooperate and claim the right of clean water according to the UTI principle. When this claim can be enforced each agent $j \in Q_{i+1}$ obtains payoff $f_j(Q_{i+1}, b^{i+1,n}, c^{i+1,n})$ assigned by the value f to the downstream (sub)river problem $(Q_{i+1}, b^{i+1,n}, c^{i+1,n})$. Next, suppose that the agents in P^i and Q_{i+1} are forced to cooperate. In this second situation, each agent $j \in N$ obtains payoff $f_j(N,b,c)$ assigned by the value f to the river problem (N,b,c).

The question can be asked which property the value f should satisfy with respect to the change in payoffs between the first situation and the second situation? Stated differently, how should the difference between the total available payoff in the second situation and the total (claimed) payoff in the first situation be distributed amongst the two groups P^i and Q_{i+1}? The TIBS principle would imply that for each group the change in payoffs is 'reasonable and equitable'. Assuming that a vector of weights $\alpha \in \mathbb{R}^N_+$ with $\sum_{i \in N} \alpha_i = 1$ is given, we let these 'reasonable and equitable' amounts be represented by the total weights $\sum_{j \in P^i} \alpha_j$ and $\sum_{j \in Q_{i+1}} \alpha_j$ of the two groups. So, the change in total payoff under full cooperation relative to the ideal situation that both groups can realize their claims is attributed to the two sets of agents P^i and Q_{i+1} proportional to the total weight of each set. We thus require for a value f on the class of river pollution models $\mathcal{RP}$ that

$$\frac{\sum_{j \in P^i} \Big(f_j(N,b,c) - f_j(P^i, b^{1,i}, c^{1,i}) \Big)}{\sum_{j \in Q_{i+1}} \Big(f_j(N,b,c) - f_j(Q_{i+1}, b^{i+1,n}, c^{i+1,n}) \Big)} = \frac{\sum_{j \in P^i} \alpha_j}{\sum_{j \in Q_{i+1}} \alpha_j},$$

provided that both $\sum_{j \in P^i} \alpha_j$ and $\sum_{j \in Q_{i+1}} \alpha_j$ are nonzero. The above discussion holds for all $i \in N \setminus \{n\}$. This results in the following axiom for a value f on the class of river pollution models (that also allows for the case that some weights are zero).

Axiom 5.1 α-TIBS fairness
Given $\alpha \in \mathbb{R}^N_+$ with $\sum_{i \in N} \alpha_i = 1$, a value f on the class of river pollution models $\mathcal{RP}$ satisfies α-TIBS fairness if for every $(N,b,c) \in \mathcal{RP}$ and any $i \in N \setminus \{n\}$ it holds that

$$\sum_{j \in Q_{i+1}} \alpha_j \Big[\sum_{j \in P^i} \Big(f_j(N,b,c) - f_j(P^i, b^{1,i}, c^{1,i}) \Big) \Big] =$$

$$\sum_{j \in P^i} \alpha_j \Big[\sum_{j \in Q_{i+1}} \Big(f_j(N,b,c) - f_j(Q_{i+1}, b^{i+1,n}, c^{i+1,n}) \Big) \Big]. \quad (5.11)$$

Efficiency and this α-TIBS fairness axiom characterize the TIBS$^\alpha$ value.

Theorem 5.2 *Given $\alpha \in \mathbb{R}^N_+$ with $\sum_{i \in N} \alpha_i = 1$, a value f on the class of river pollution models $\mathcal{RP}$ satisfies efficiency and α-TIBS fairness if and only if f is the $TIBS^\alpha$ value.*

Proof. We prove in Appendix B that the TIBS$^\alpha$ value satisfies the two axioms. Here we prove that there exists a unique value that satisfies efficiency and α-TIBS fairness by induction on the number of agents. Let (K,b,c) be a one-agent river problem with $K = \{k\}$ for some $k \in \mathbb{N}$, i.e., k is the single agent in K. Then by efficiency we have that $f_k(K,b,c) = W(K,b,c)$, where $W(K,b,c) = b_k(r_k) - c_k(r_k)$ with b_k and c_k the benefit and cost functions of k and r_k the optimal level of pollution.

Now, assume by induction that $f(K,b,c)$ is determined uniquely by efficiency and α-TIBS fairness for every river pollution model (K,b,c) with number of agents $k = |K| < n$, and let (N,b,c) be a river pollution model with $n = |N|$ agents. For every $i \in N \setminus \{n\}$, the (sub)river models $(P^i, b^{1,i}, c^{1,i})$ and $(Q_{i+1}, b^{i+1,n}, c^{i+1,n})$ have at most $n-1$ agents and so the payoff vectors $f(P^i, b^{1,i}, c^{1,i}) \in \mathbb{R}^{P^i}$ and $f(Q_{i+1}, b^{i+1,n}, c^{i+1,n}) \in \mathbb{R}^{Q_{i+1}}$ have been determined already. Efficiency of f implies on the (sub)river problem $(P^i, b^{1,i}, c^{1,i})$ that

$$\sum_{j \in P^i} f_j(P^i, b^{1,i}, c^{1,i}) = W(P^i, b^{1,i}, c^{1,i})$$

and on the (sub)river problem $(Q_{i+1}, b^{i+1,n}, c^{i+1,n})$ that

$$\sum_{j \in Q_{i+1}} f_j(Q_{i+1}, b^{i+1,n}, c^{i+1,n}) = W(Q_{i+1}, b^{i+1,n}, c^{i+1,n}).$$

So, the α-TIBS fairness property reduces to

$$\sum_{j \in Q_{i+1}} \alpha_j \Big[\sum_{j \in P^i} f_j(N, b, c) - W(P^i, b^{1,i}, c^{1,i}) \Big] =$$

$$\sum_{j \in P^i} \alpha_j \Big[\sum_{j \in Q_{i+1}} f_j(N, b, c) - W(Q_{i+1}, b^{i+1,n}, c^{i+1,n}) \Big] \tag{5.12}$$

for all $i \in N \setminus \{n\}$. Since there are $n-1$ equations of type (5.12) and also it must hold by efficiency that $\sum_{i \in N} f_i(N, b, c) = W(N, b, c)$, we have n linearly independent equations in n unknowns. Hence, the payoffs $f_i(N, b, c)$, $i \in N$, are uniquely determined. □

The class of TIBS$^\alpha$ values is very rich in the sense that it encompasses a lot of other values. First of all, notice that by definition of the payoff vectors t^j, the $TIBS^\alpha(N, b, c) = ATS(N, b, c)$ if $\alpha_n = 1$ and that $TIBS^\alpha(N, b, c) = UTI(N, b, c)$ if $\alpha_1 = 1$. So, the case that all weight is given to the last agent reflects the ATS principle and every upstream coalition receives the payoff that it can obtain when it has the right to pollute. Reversely, the case that all weight is given to the first agent reflects the UTI principle and every downstream coalition receives the payoff that it can obtain when it has the right to claim no pollution by its upstream agents. More generally, when $\alpha_j = 0$ for all $j \leq i$, then $TIBS^\alpha(N, b, c)$ is a weighted average of the vectors $t^j(N, b, c)$, $j \geq i+1$ and all agents in the upstream set P^i receive their ATS-payoff. Reversely, when $\alpha_j = 0$ for all $j \geq i+1$, then $TIBS^\alpha(N, b, c)$ is a weighted average of the vectors $t^j(N, b, c)$, $j \leq i$ and all agents in the downstream set Q_{i+1} receive their UTI-payoff.

Some particular solutions are the following. For the weight vector $\alpha^e \in \mathbb{R}^N_+$ with $\alpha^e_1 = \alpha^e_n = \frac{1}{2}$, the α-TIBS fairness property (5.11) reduces for every $i \in N \setminus \{n\}$ to

$$\sum_{j \in P^i} \Big(f_j(N, b, c) - f_j(P^i, b^{1,i}, c^{1,i}) \Big) = \sum_{j \in Q_{i+1}} \Big(f_j(N, b, c) - f_j(Q_{i+1}, b^{i+1,n}, c^{i+1,n}) \Big).$$

Noticing that for $f = TIBS^{\alpha^e}$ we have that $\sum_{j \in P^i} TIBS^{\alpha^e}_j(P^i, b^{1,i}, c^{1,i}) = W(P^i, b^{1,i}, c^{1,i})$ and $\sum_{j \in Q_{i+1}} TIBS^{\alpha^e}_j(Q_{i+1}, b^{i+1,n}, c^{i+1,n}) = W(Q_{i+1}, b^{i+1,n}, c^{i+1,n})$, the α^e-TIBS fairness axiom states that, for every $i < n$, the total (combined) loss that the agents in P^i experience when they are forced to cooperate with the agents in Q_{i+1} should be equal to the total (combined) loss that the agents in Q_{i+1} experience. Together with the efficiency axiom the α^e-TIBS fairness axiom characterizes the value

$$TIBS_i^{\alpha^e}(N,b,c) = \frac{ATS_i + UTI_i}{2},$$

being the average of the ATS and UTI values.[8] More generally, every weight vector $\alpha \in \mathbb{R}_+^N$ with $\alpha_1 + \alpha_n = 1$ gives a weighted average of the ATS and UTI values.

When we take weight vector $\alpha^a \in \mathbb{R}_+^N$ with $\alpha_i^a = \frac{1}{n}$ for all $i \in N$, the α-TIBS fairness property (5.11) reduces for every $i \in N \setminus \{n\}$ to

$$\frac{1}{i}\Big[\sum_{j\in P^i}\Big(f_j(N,b,c) - f_j(P^i, b^{1,i}, c^{1,i})\Big)\Big] = \frac{1}{n-i}\Big[\sum_{j\in Q_{i+1}}\Big(f_j(N,b,c) - f_j(Q_{i+1}, b^{i+1,n}, c^{i+1,n})\Big)\Big].$$

Thus the α^a-TIBS fairness axiom states that, for every $i < n$, the average loss that the agents in P^i experience when they are forced to cooperate with the agents in Q_{i+1} should be equal to the average loss that the agents in Q_{i+1} experience. Together with the efficiency axiom the α^a-TIBS fairness axiom characterizes the value

$$TIBS_i^{\alpha^a}(N,b,c) = \frac{1}{n}\sum_{j\in N} t_i^j(N,b,c),$$

which is the average of all values $t^j(N,b,c)$, $j \in N$.[9]

The discussion above shows that the vector of weights $\alpha \in \mathbb{R}_+^N$ can be seen as containing information on the property rights over the river. In fact, the weights can be seen as some sort of counterparts of the property rights; they show how the loss of welfare resulting from enforced cooperation between an upstream set P^i and its downstream complement Q_{i+1} is distributed between the two groups, relative to the most ideal situations for both groups. When $\alpha_1 = 1$, all the loss is taken by P^i, when $\alpha_n = 1$, all loss is taken by Q_{i+1}. Further, when $\alpha = \alpha^e$, both groups equally share the loss and when $\alpha = \alpha^a$, the average loss of the agents in both groups is equal. We could say that $\sum_{j\in P^i} \alpha_j$ and $\sum_{j\in Q_{i+1}} \alpha_j$ reflect the responsibilities of both groups of not polluting the water. The higher $\sum_{j\in P^i} \alpha_j$ is, the larger the loss that the group of agents P^i has to take relative to its total payoff $\sum_{j\in P^i} ATS_j(N,b,c)$ in its most ideal situation; respectively the higher $\sum_{j\in Q_{i+1}} \alpha_j$ is, the larger the loss that the group of agents Q_{i+1} has to take relative to its total payoff $\sum_{j\in Q_{i+1}} UTI_j(N,b,c)$ in its most ideal situation. Although in this paper the weights

[8] The value $TIBS^{\alpha^e}$ resembles the equal gain splitting solution for sequencing problems of Curiel (1988).

[9] The value $TIBS^{\alpha^a}$ resembles the average tree solution for cycle-free graph games of Herings, van der Laan and Talman (2008).

α_i, $i \in N$, are exogenous (they can be seen as reflecting, for instance, existing power structures among countries) we could also envision a model in which they are the subject of negotiation between the agents. In that case, agents bargain over weights α_i, $i \in N$, that in combination with efficiency and α-TIBS fairness would lead to a unique solution for river pollution models.

6 Rivers with multiple springs and multiple sinks

In this section we generalize the river pollution model (N, b, c) in which agents are located along the single-stream river from upstream to downstream to river pollution models with multiple springs and/or multiple sinks, i.e., rivers that have multiple tributaries and/or multiple distributaries. We also define and characterize $TIBS^\alpha$ values for such river systems. We describe a river system with multiple springs and sinks by a directed graph (N, D), where the set of nodes of the graph $N = \{1, ..., n\}$ corresponds to the set of agents along the river, and $D \subseteq \{(i,j)|i,j \in N,\ i \neq j\}$ is a collection of directed links that represents the flow of water between the agents. That means that a directed link $(i,j) \in \{(i,j)|i,j \in N,\ i \neq j\}$ is in the set D if and only if j is a downstream neighbor of i along the river (and thus i is an upstream neighbor of j). Each spring and each sink of the river is identified by an agent, i.e., an agent $i \in N$ is a spring when it has no upstream neighbors and an agent $i \in N$ is a sink if it has no downstream neighbors. We denote by $K^i \subset N$ the set of all neighbors (upstream and downstream) of i. Notice that for a river system (N, b, c) with a single spring, a single sink, $N = \{1, \dots, n\}$ and agents numbered successively from upstream to downstream, the collection of directed links is given by $D = \{(i, i+1)|i \in N \setminus \{n\}\}$.

We only consider river systems that are represented by connected graphs[10], i.e., for each two different agents i and j there is sequence of k different agents $(i_1, ..., i_k)$ such that $i_1 = i$, $i_k = j$ and, for every $h = 1, \dots, k-1$, either $(i_h, i_{h+1}) \in D$ or $(i_{h+1}, i_h) \in D$. We call such a sequence a path, i.e., starting from i agent j can be reached by traveling on the river visiting subsequently i_h, $h = 2, \dots, k-1$. Notice that $k = 2$ when i and j are neighbors and that the journey goes downstream when $(i_h, i_{h+1}) \in D$ and upstream when $(i_{h+1}, i_h) \in D$. We say that agent $j \neq i$ is upstream of i (and i downstream of j), when along the full path the journey from j to i is downstream. For $i \in N$, let P^i denote the set of agents upstream of, and including, agent i in (N, D).

Finally, for (N, D) to represent a river we require that (N, D) is cycle-free, i.e., for each pair i and j there is a unique path connecting i and j. A connected cycle-free directed graph (N, D) gives the most general possible definition of a river, except that it does not

[10] Otherwise we have several river systems, which can be treated separately.

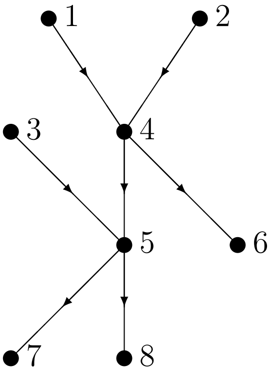

Figure 1: River (N, D) from Example 6.1.

allow for anabranches (parts of a river where it splits into two or more separate streams that merge again further downstream).

Example 6.1 Let (N, D) represent a river with $N = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and $D = \{(1, 4), (2, 4), (3, 5), (4, 5), (4, 6), (5, 7), (5, 8)\}$, see Figure 1. Then the set of springs is $\{1, 2, 3\}$ and the set of sinks is $\{6, 7, 8\}$. The two streams originating at 1 and 2 merge together at agent 4. There the river immediately splits again into two streams, one to agent 5 and one to agent 6. The stream at agent 5 is joined by a stream originating at agent 3. The resulting stream, in turn, splits into two streams, one to agent 7 and one to agent 8. For $i = 5$, we have that $K^5 = \{3, 4, 7, 8\}$ where 3 and 4 are upstream neighbors and 7 and 8 are downstream neighbors. Further, $P^5 = \{1, 2, 3, 4, 5\}$ is the set of agents upstream of 5, including 5 itself. Notice that agent 6 is not in P^5, because along the path from 6 to 5, one has to travel upstream when going from 6 to 4. □

A river pollution model with multiple springs and/or multiple sinks is now given by $((N, D), b, c)$ with (N, D) the river system, and as before, $b = \{b_i | i \in N\}$ the collection of benefit functions and $c = \{c_i | i \in N\}$ the collection of cost functions.

At vector $p \in \mathbb{R}^N_+$ of pollution levels, the total pollution experienced by an agent i is given by $q_i(p) = \sum_{j \in P^i} p_j$; this is the total pollution of i itself and all its upstream agents. Notice that the pollution caused by some agent i hurts all its downstream agents, thus pollution of for instance agent 4 affects agent 4 itself and its downstream agents 5, 7 and 8. Again, the output of the model is a pair (p, t) of pollution levels and monetary transfers, yielding to every $i \in N$ payoff

$$z_i(p, t) = u_i(p) + t_i = b_i(p_i) - c_i(\sum_{j \in P^i} p_j) + t_i.$$

Let $\widetilde{p}$ be a solution to the welfare maximization problem

$$\max_{p \in \mathbf{R}^N_+} \sum_{i \in N} \Big(b_i(p_i) - c_i(\sum_{j \in P^i} p_j) \Big) \tag{6.13}$$

and denote $W((N,D),b,c) = \sum_{i \in N} u_i(\widetilde{p})$ as the highest social welfare that can be obtained.

The class of all river pollution models with multiple springs and/or multiple sinks is denoted by $\mathcal{RPM}$ and a value is a function f on $\mathcal{RPM}$ that assigns to every $((N,D),b,c) \in \mathcal{RPM}$ a payoff vector $f((N,D),b,c) \in \mathbb{R}^N$. We now generalize the efficiency and α-TIBS fairness axioms and the corresponding TIBS$^\alpha$ value to the class $\mathcal{RPM}$.

Axiom 6.2 Efficiency on $\mathcal{RPM}$
A value f on the class of river pollution models $\mathcal{RPM}$ is efficient if it holds for every $((N,D),b,c) \in \mathcal{RPM}$ that $\sum_{i \in N} f_i((N,D),b,c) = W((N,D),b,c)$.

To state the α-TIBS fairness axiom on $\mathcal{RPM}$, consider a connected and cycle-free river system (N,D) and suppose that a link $(i,j) \in D$ is deleted, i.e., there is no water flow from agent i to its downstream neighbor j. We then have two separate connected cycle-free directed graphs that, individually, again represent (part of) a river. For instance, if we delete $(4,5)$ from D in Example 6.1 we obtain two separate river systems, namely $(\{1,2,4,6\},\{(1,4),(2,4),(4,6)\})$ and $(\{3,5,7,8\},\{(3,5),(5,7),(5,8)\})$.[11] Let (N,D) be a river system and suppose that either $(i,j) \in D$ or $(j,i) \in D$. Then we denote by $(N^{i|j}, D^{i|j})$, respectively $(N^{j|i}, D^{j|i})$ the two subriver systems that result when deleting this link from D, where $(N^{i|j}, D^{i|j})$ represents the part that contains agent i and $(N^{j|i}, D^{j|i})$ the part that contains j. Denote $b^{N^{i|j}}$ as the set of benefit functions $b_k^{N^{i|j}} = b_k$, $k \in N^{i|j}$. Analogously, denote $c^{N^{i|j}}$ as the set of cost functions for $k \in N^{i|j}$ and $b^{N^{j|i}}$ and $c^{N^{j|i}}$ as the sets of benefit and cost functions for $k \in N^{j|i}$. We are now ready to state the α-TIBS fairness axiom on $\mathcal{RPM}$.

Axiom 6.3 α-TIBS fairness on $\mathcal{RPM}$
Given $\alpha \in \mathbb{R}_+^N$ with $\sum_{i \in N} \alpha_i = 1$, a value f on the class of river pollution models $\mathcal{RPM}$ satisfies α-TIBS fairness if for every $((N,D),b,c) \in \mathcal{RPM}$ and any $(i,j) \in D$ it holds that

$$\sum_{k \in N^{j|i}} \alpha_j \Big[\sum_{k \in N^{i|j}} \Big(f_k((N,D),b,c) - f_k((N^{i|j}, D^{i|j}), b^{N^{i|j}}, c^{N^{i|j}}) \Big) \Big] =$$
$$\sum_{k \in N^{i|j}} \alpha_j \Big[\sum_{k \in N^{j|i}} \Big(f_k((N,D),b,c) - f_k((N^{j|i}, D^{j|i}), b^{N^{j|i}}, c^{N^{j|i}}) \Big) \Big].$$

Similar as on the class $\mathcal{RP}$ of river problems with one spring and one sink, the two axioms characterize a unique value on the class $\mathcal{RPM}$. This value is the generalization of the $TIBS^\alpha$ value to $\mathcal{RPM}$. To state this value, let $i,j \in N$ be two different agents and let h_j^i be the first agent on the unique path in (N,D) from i to j that is reached after leaving i.[12] For instance, in Example 6.1 if we take $i=4$ and $j=3$ then $h_j^i = 5$. As in Section 5,

[11] Observe that deleting for instance $(1,4)$ from D in Example 6.1 we obtain the river system $(\{1\},\emptyset)$ with 1 the only agent in it, and the river system $(\{2,3,4,5,6,7,8\},\{(2,4),(3,5),(4,5),(4,6),(5,7),(5,8)\})$.

[12] Notice that $h_j^i = j$ if j is a direct neighbor of i in (N,D).

again we define for each $j \in N$ a payoff vector $t^j((N,D),b,c)$. First, the payoff to j itself is defined as

$$t^j_j((N,D),b,c) = W((N,D),b,c) - \sum_{k \in K^j} W((N^{k|j}, D^{k|j}), b^{N^{k|j}}, c^{N^{k|j}}))$$

and for $i \neq j$ the payoff to i is defined as

$$t^j_i((N,D),b,c) = W((N^{i|h^i_j}, D^{N^{i|h^i_j}}), b^{N^{i|h^i_j}}), c^{N^{i|h^i_j}}))- \sum_{k \in K^i \setminus \{h^i_j\}} W((N^{k|i}, D^{k|i}), b^{N^{k|i}}, c^{N^{k|i}}).$$

Example 6.4 We illustrate the payoff vector t^j for some agents in the river pollution model given in Figure 1. Taking $j = 5$, for the agents $i = 3,\ 7,\ 8$ we have $h^i_5 = 5$ and i does not have other neighbors. So, $N^{i|h^i_j} = \{i\}$ and it follows that each of these agents receives a payoff equal the payoff that it can achieve by maximizing $b_i(p_i) - c_i(p_i)$. Since 3 is a spring it follows that $t^5_3((N,D),b,c)$ is equal to the utility that 3 can attain under the ATS principle, i.e., not taking into account the effect of its pollution on its downstream agents. On the other hand, agents 7 and 8 are sinks and their payoffs $t^5_i((N,D),b,c)$, $i = 7, 8$ are equal to the utilities these agents can attain under the UTI principle, i.e., it is the payoff they can obtain under the condition that their upstream agents set zero pollution levels. For the agents $i = 1,\ 2,\ 6$ we have $h^i_5 = 4$ and i does not have other neighbors. Also for these agents we have $N^{i|h^i_j} = \{i\}$ and it follows that each of these agents receives a payoff equal to the payoff it can achieve by maximizing $b_i(p_i) - c_i(p_i)$. Since 1 and 2 are springs it follows that $t^5_i((N,D),b,c)$, $i = 1, 2$, are equal to the utilities that these agents can attain under the ATS principle. Agents 6 is a sink and receives payoff $t^5_6((N,D),b,c)$ equal to the utility it can attain under the UTI principle. For $i = 4$ we have $h^4_5 = 5$ and $K^4 \setminus \{h^4_5\} = \{1, 2, 6\}$. It follows that the payoff $t^5_4((N,D),b,c)$ is equal to the total welfare that the agents in $N^{4|5} = \{1, 2, 4, 6\}$ can attain on their own (so under the ATS principle, not taking into account the effect of their pollution on their downstream agents 5, 7 and 8), minus the total payoff to the agents 1, 2, and 6. Finally, agent 5 receives a payoff $t^5_5((N,D),b,c)$ equal to the total welfare minus the sum of the payoffs to the other agents.

Taking $j = 8$ the payoffs are the same as above for all agents, except agents 5 and 8. For agent 5 we now have $h^5_8 = 8$ and $K^5 \setminus \{h^5_8\} = \{3, 4, 7\}$. It follows that its payoff $t^8_5((N,D),b,c)$ is equal to the total welfare that can be obtained by the agents $N^{5|8} = N \setminus \{8\}$ under the ATS principle (not taking into account its effect on 8) minus the sum of the payoffs to the other players in $N^{5|8}$. Agent 8 receives payoff $t^8_8((N,D),b,c)$ equal to the total welfare minus the sum of the payoffs to the other agents. □

It should be noticed that, given an agent j for each $k \in K^j$ it holds that

$$\sum_{h \in N^{k|j}} t^j_h((N,D),b,c) = W((N^{k|j}, D^{k|j}), b^{N^{k|j}}, c^{N^{k|j}}),$$

thus each set of agents $N^{k|j}$ in the part of the river containing k and that results from deleting (k, j) from D (when k is upstream of j) or deleting (j, k) from D (when k is downstream of j), receives the total payoff it can attain on its own, ignoring the others. When k is an upstream neighbor of j, it means that the agents in $N^{k|j}$ realize their welfare under the ATS principle (not taking into account the effect of their pollution on their downstream agents) and when k is a downstream neighbor of j it means that the agents in $N^{k|j}$ realize their welfare under the UTI principle (claiming zero pollution by their upstream agents). Stated differently, the payoff vector $t^j((N, D), b, c)$ assigns to upstream sets $N^{k|j}$, $k \in K^j$, their ATS claims and to downstream sets $N^{k|j}$, $k \in K^j$, their UTI claims.

For a given weight vector $\alpha \in \mathbb{R}^N_+$ with $\sum_{i \in N} \alpha_i = 1$, the $TIBS^\alpha$ value assigns to each river pollution model in $\mathcal{RPM}$ the weighted average of the payoff vectors $t^j((N, D), b, c)$, thus

$$TIBS^\alpha((N, D), b, c) = \sum_{j \in N} \alpha_j t^j((N, D), b, c).$$

The next theorem states that Efficiency and α-TIBS fairness on $\mathcal{RPM}$ characterize the $TIBS^\alpha$ value on $\mathcal{RPM}$. The proof goes along the same lines as the proof of Theorem 5.2 and is therefore omitted.

Theorem 6.5 *Given $\alpha \in \mathbb{R}^N_+$ with $\sum_{i \in N} \alpha_i = 1$, a value f on the class $\mathcal{RPM}$ of river pollution models with multiple springs and sinks satisfies efficiency and α-TIBS fairness on $\mathcal{RPM}$ if and only if f is the $TIBS^\alpha$ value.*

Also within the class $\mathcal{RPM}$ the vector of weights $\alpha \in \mathbb{R}^N_+$ can be seen as some sort of counterparts of the property rights over the river. When a link (i, j) (thus j is downstream of i) is deleted from D, the weights determine how the loss of welfare resulting from enforced cooperation between the upstream set $N^{i|j}$ and the downstream set $N^{j|i}$ is distributed between the two groups, relative to the most ideal situations for both groups. The higher $\sum_{k \in N^{i|j}} \alpha_k$ is, the larger the loss that the group of agents $N^{i|j}$ has to take relative to the total welfare it can attain without taking care of its downstream group $N^{j|i}$; respectively the higher $\sum_{k \in N^{j|i}} \alpha_k$ is, the larger the loss that the group of agents $N^{j|i}$ has to take relative to the total welfare it can attain under the claim of zero pollution by its upstream agents.

7 Concluding remarks

In this paper we have introduced a model for international river pollution problems in which the agents can choose pollution levels and are able to make monetary compensations to each other. We found that the total pollution level is always lower under cooperation (if agents coordinate their pollution levels) than under individual action. The gain in social welfare that results when the agents would switch from their individually optimal to the socially optimal pollution levels can be distributed among the agents through monetary transfers. Following Coase (1960) these monetary transfers only depend on the initial allocation of the property rights. Since an initial allocation of the property rights is not given, we referred to the ATS, UTI and TIBS principles from international water law to provide an allocation of the property rights and therefore a distribution of the cooperative gains. The ATS value allocated the property rights over the river to the upstream agents but the gains of cooperation to the downstream agents. The UTI value, conversely, distributed the property rights over the river to the downstream agents but the gains of cooperation to the upstream agents. The TIBS$^{\alpha}$ value allowed for compromises between the ATS and UTI values by using an exogenous vector of weights and distributed the property rights over the river in accordance with this vector.

References

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

Ambec, S. and L. Ehlers (2008), Sharing a river among satiable agents, *Games and Economic Behavior* 64, 35-50.

Ansink, E. and H.P. Weikard (2011), Sequential sharing rules for river sharing problems, *Social Choice and Welfare*, doi:10.1007/s00355-010-0525-y.

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

Brink, J.R. van den, G. van der Laan and N. Moes (2011), Fair agreements for sharing international rivers with multiple springs and externalities, *Journal of Environmental Economics and Management*, doi:10.1016/j.jeem.2011.11.003.

Brink, J.R. van den, A. Estévez-Fernández, G. van der Laan and N. Moes (2011), Independence axioms for water allocation, Tinbergen Institute Discussion Paper 2011-128/1, Tinbergen Institute, Amsterdam.

Brink, J.R. van den, G. van der Laan and V. Vasil'ev (2007), Component efficient solutions in line-graph games with applications, *Economic Theory* 33, 349-364.

Coase, R.H. (1960), The problem of social cost, *Journal of Law and Economics* 3, 1-44.

Curiel, I. (1988), *Cooperative Game Theory and Applications*, Ph.D. Thesis, Katholieke Universiteit Nijmegen.

Demange, G. (2004), On group stability in hierarchies and networks, *Journal of Political Economy* 112, 754-778.

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

Fernandez, L. (2002), Solving water pollution problems along the US-Mexico Border, *Environment and Development Economics* 7, 715-732.

Fernandez, L. (2009), Wastewater pollution abatement across an international border, *Environment and Development Economics* 14, 67-88.

Garretson, A.H., R.D. Hayton and C.J. Olmstead (1967), *The Law of International Drainage Basins*, Oceana Publications, Inc., Dobbs Ferry, New York.

Gengenbach, M.F., H.P. Weikard and E. Ansink (2010), Cleaning a river: an analysis of voluntary joint action, *Natural Resource Modeling* 23, 565-589.

Gray, W.B. and R.J. Shadbegian (2004), 'Optimal' pollution abatement - whose benefits matter, and how much?, *Journal of Environmental Economics and Management* 47, 510-534.

Herings, P.J.J., G. van der Laan and A.J.J. Talman (2007), The socially stable core in structured transferable utility games, *Games and Economic Behavior* 59, 85-104.

Herings, P.J.J., G. van der Laan and A.J.J. Talman (2008), The average tree solution for cycle-free graph games, *Games and Economic Behavior* 62, 77-92.

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

Kilgour, D.M. and A. Dinar (1995), Are stable agreements for sharing international river waters now possible?, Policy Research Working Paper 1474, World Bank, Washington D.C.

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

Lipscomb M. and A.M. Mobarak (2007), Decentralization and water quality: evidence from the re-drawing of county boundaries in Brazil, BREAD working paper no. 159, Bureau for Research and Economic Analysis of Development, Durham, North Carolina.

Mäler, K.G. (1990), International environmental problems, *Oxford Review of Economic Policy* 6, 80-108.

McCaffrey, S.C. (2001), *The Law of International Watercourses*, Oxford University Press, Oxford.

Ni, D. and Y. Wang (2007), Sharing a polluted river, *Games and Economic Behavior* 60, 176-186.

Parrachino, I., A. Dinar and F. Patrone (2006), Cooperative game theory and its application to natural, environmental and water resource issues: 3. Application to water resources, Policy Research Working Paper 4074, World Bank, Washington.

Shapley, L.S. (1953), A value for n-person games, in *Contributions to the Theory of Games, Vol. II*, ed. H.W. Kuhn and A.W. Tucker, Princeton University Press, Princeton, 307-317.

Sigman, H. (2002), International spillovers and water quality in rivers: do countries free-ride?, *The American Economic Review* 92, 1152-1159.

Sigman, H. (2005), Transboundary spillovers and decentralization of environmental policies, *Journal of Environmental Economics and Management* 50, 82-101.

Wang, Y. (2011), Trading water along a river, *Mathematical Social Sciences* 61, 124-130.

Appendix A

Proof of Proposition 2.4. We prove this proposition by induction on the number of agents n. We first consider a river model (N, b, c) with $n = 2$. As noticed in the proofs of Proposition 2.2 and Proposition 2.3, when $n = 2$ we have both in the Nash equilibrium and in the Pareto efficient output that agent 2 sets it pollution level p_2 so that

$$\frac{\partial b_2(p_2)}{\partial p_2} - \frac{\partial c_2(q_2)}{\partial q_2}\frac{\partial q_2(p)}{\partial p_2} = \frac{\partial b_2(p_2)}{\partial p_2} - \frac{\partial c_2(q_2)}{\partial q_2} = 0.$$

Because $\frac{\partial c_2}{\partial q_2}$ is continuous and strictly increasing it has an inverse $\frac{\partial c_2}{\partial q_2}^{-1}$. It follows that in both cases it must hold that

$$q_2 = p_1 + p_2 = \frac{\partial c_2}{\partial q_2}^{-1}\frac{\partial b_2(p_2)}{\partial p_2}. \tag{7.14}$$

It also follows from the proof of Proposition 2.3 that $\widetilde{p}_1$ and $\widetilde{p}_2$ are such that

$$\frac{\partial b_1(\widetilde{p}_1)}{\partial p_1} - \frac{\partial c_1(\widetilde{p}_1)}{\partial q_1} = \frac{\partial b_2(\widetilde{p}_2)}{\partial p_2}.$$

Since $b_2' > 0$ at every $p_2 > 0$ and $\widetilde{p}_2 > 0$ (see proof Proposition 2.3) it must be that

$$\frac{\partial b_1(\widetilde{p}_1)}{\partial p_1} > \frac{\partial c_1(\widetilde{p}_1)}{\partial q_1}.$$

By Assumption 2.1 it follows that $\widetilde{p}_1 < \widehat{p}_1$. When agent 1 chooses $p_1 = \widetilde{p}_1 < \widehat{p}_1$ and agent 2 would pollute at (or below) its Nash equilibrium pollution level, $p_2 \leq \widehat{p}_2$, we would have that

$$u_1((p_1, p_2)) + u_2((p_1, p_2)) < b_1(\widehat{p}_1) - c_1(\widehat{p}_1) + b_2(\widehat{p}_2) - c_2(\widehat{p}_2).$$

This would mean that agent 1 and 2 would obtain a higher social welfare in the Nash equilibrium than in the Pareto efficient output, a contradiction. We therefore must have that $\widetilde{p}_2 > \widehat{p}_2$. Now, since $\widetilde{p}_2 > \widehat{p}_2$ and b_2' is strictly decreasing in p_2 it follows that $\frac{\partial b_2(\widetilde{p}_2)}{\partial p_2} < \frac{\partial b_2(\widehat{p}_2)}{\partial p_2}$. Because $\frac{\partial c_2}{\partial q_2}^{-1}$ is strictly increasing in its argument it follows from equation (7.14) that

$$\widetilde{p}_1 + \widetilde{p}_2 = \frac{\partial c_2}{\partial q_2}^{-1} \frac{\partial b_2(\widetilde{p}_2)}{\partial p_2} < \frac{\partial c_2}{\partial q_2}^{-1} \frac{\partial b_2(\widehat{p}_2)}{\partial p_2} = \widehat{p}_1 + \widehat{p}_2.$$

We now denote the vectors of the unique Nash equilibrium and social welfare maximizing pollution levels for a tuple (K, b, c) with $k = |K|$ agents by $\widehat{p}^k$, respectively $\widetilde{p}^k$. Proceeding by induction, assume that

$$\sum_{i=1}^{k} \widetilde{p}_i^k < \sum_{i=1}^{k} \widehat{p}_i^k. \tag{7.15}$$

for every river pollution model (K, b, c) with $k = |K| < n$. Now, for some (N, b, c) with $|N| = n$, let $(N \setminus \{n\}, b, c)$ be the model in which the last agent n is deleted. By definition, $\widetilde{p}_i^{n-1}$, $i \in N \setminus \{n\}$ is the solution to the welfare maximization problem

$$\max_{p_1, \ldots, p_{n-1}} \sum_{i=1}^{n-1} b_i(p_i) - \sum_{i=1}^{n-1} c_i\Big(\sum_{j=1}^{i} p_j\Big) \tag{7.16}$$

and $\widetilde{p}_i^n$, $i \in N$ is the solution to the welfare maximization problem

$$\max_{p_1, \ldots, p_n} \sum_{i=1}^{n-1} b_i(p_i) - \sum_{i=1}^{n-1} c_i\Big(\sum_{j=1}^{i} p_j\Big) + \Big[b_n(p_n) - c_n\Big(\sum_{j=1}^{n-1} p_j + p_n\Big)\Big]. \tag{7.17}$$

Since $c_n' > 0$ at every $q_n = \sum_{j=1}^{n-1} p_j + p_n$, it follows from comparing problem (7.16) with problem (7.17) that

$$\sum_{i=1}^{n-1} \widetilde{p}_i^n \leq \sum_{i=1}^{n-1} \widetilde{p}_i^{n-1}. \tag{7.18}$$

On the other hand we have that

$$\sum_{i=1}^{n-1} \widehat{p}_i^{n-1} = \sum_{i=1}^{n-1} \widehat{p}_i^n, \tag{7.19}$$

because the unique Nash equilibrium pollution levels of the agents $1, \ldots, n-1$ do not depend on the action (or presence) of agent n. From inequality (7.15) with $k = n-1$, and the (in)equalities (7.18) and (7.19) it follows that

$$\sum_{i=1}^{n-1} \widetilde{p}_i^n < \sum_{i=1}^{n-1} \widehat{p}_i^n.$$

As noticed in the proofs of Proposition 2.2 and Proposition 2.3, both in the Nash equilibrium and in the Pareto efficient output agent n sets it pollution level p_n so that

$$\frac{\partial b_n(p_n)}{\partial p_n} - \frac{\partial c_n(\sum_{j=1}^n p_j)}{\partial q_n} = 0.$$

Because b_n' is strictly decreasing, c_n' is strictly increasing and $\sum_{i=1}^{n-1} \widetilde{p}_i^n < \sum_{i=1}^{n-1} \widehat{p}_i^n$, we would have that

$$\frac{\partial b_n(p_n)}{\partial p_n} - \frac{\partial c_n(\sum_{i=1}^{n-1} \widetilde{p}_i^n + p_n)}{\partial q_n} > 0.$$

for any $p_n \leq \widehat{p}_n$. So, it must be that $\widetilde{p}_n^n > \widehat{p}_n^n$. Further, because c_n' is continuous and strictly increasing it has an inverse $\frac{\partial c_n}{\partial q_n}^{-1}$ that is also strictly increasing in its argument. Analogously as for the case $n = 2$ it now follows that

$$\sum_{i=1}^{n} \widetilde{p}_i^n = \frac{\partial c_n}{\partial q_n}^{-1} \frac{\partial b_n(\widetilde{p}_n^n)}{\partial p_n} < \frac{\partial c_n}{\partial q_n}^{-1} \frac{\partial b_n(\widehat{p}_n^n)}{\partial p_n} = \sum_{i=1}^{n} \widehat{p}_i^n.$$

□

Appendix B

Proof that the TIBS$^\alpha$ value satisfies Efficiency and α-TIBS fairness. Efficiency follows straightforwardly from the definition of $TIBS^\alpha$, since

$$\sum_{i \in N} TIBS_i^\alpha(N, b, c) = \sum_{i \in N} \sum_{j \in N} \alpha_j t_i^j(N, b, c) = \sum_{j \in N} \alpha_j \sum_{i \in N} t_i^j(N, b, c) =$$

$$\sum_{j\in N}\alpha_j\ W(N,b,c)=W(N,b,c).$$

To show the second axiom, consider an agent $i\in N\setminus\{n\}$. Then

$$\sum_{j\in P^i} TIBS^{\alpha}_j(N,b,c)=\sum_{j\in P^i}\sum_{k\in N}\alpha_k t^k_j(N,b,c)=$$
$$\sum_{j\in P^i}\Big(\sum_{k\in P^i}\alpha_k t^k_j(N,b,c)+\sum_{k\in Q_{i+1}}\alpha_k t^k_j(N,b,c)\Big)=$$
$$\sum_{k\in P^i}\alpha_k\sum_{j\in P^i}t^k_j(N,b,c)+\sum_{k\in Q_{i+1}}\alpha_k\sum_{j\in P^i}t^k_j(N,b,c)=$$
$$\sum_{k\in P^i}\alpha_k\Big(W(N,b,c)-\sum_{j\in Q_{i+1}}t^k_j(N,b,c)\Big)+\sum_{k\in Q_{i+1}}\alpha_k\sum_{j\in P^i}t^k_j(N,b,c)=$$
$$\sum_{k\in P^i}\alpha_k\Big(W(N,b,c)-\sum_{j\in Q_{i+1}}UTI_j(N,b,c)\Big)+\sum_{k\in Q_{i+1}}\alpha_k\sum_{j\in P^i}ATS_j(N,b,c),\qquad(7.20)$$

where the last two equalities follow from the definition of the payoff vectors $t^j(N,b,c)$, $j\in N$. Substituting equations (4.8) and (4.10) into equation (7.20) yields

$$\sum_{j\in P^i}TIBS^{\alpha}_j(N,b,c)=$$
$$\sum_{k\in P^i}\alpha_k\Big(W(N,b,c)-W(Q_{i+1},b^{i+1,n},c^{i+1,n})\Big)+\sum_{k\in Q_{i+1}}\alpha_k W(P^i,b^{1,i},c^{1,i}).\qquad(7.21)$$

By efficiency of $TIBS^{\alpha}$ in the (sub)river model $(P^i,b^{1,i},c^{1,i})$ we have that

$$\sum_{j\in P^i}TIBS^{\alpha}_j(P^i,b^{1,i},c^{1,i})=W(P^i,b^{1,i},c^{1,i})=\sum_{k\in N}\alpha_k W(P^i,b^{1,i},c^{1,i}).\qquad(7.22)$$

Subtracting equation (7.22) from equation (7.21) we obtain

$$\sum_{j\in P^i}\Big(TIBS^{\alpha}_j(N,b,c)-TIBS^{\alpha}_j(P^i,b^{1,i},c^{1,i})\Big)=$$
$$\sum_{k\in P^i}\alpha_k\Big(W(N,b,c)-W(Q_{i+1},b^{i+1,n},c^{i+1,n})\Big)+\Big(\sum_{k\in Q_{i+1}}\alpha_k-\sum_{k\in N}\alpha_k\Big)W(P^i,b^{1,i},c^{1,i})=$$
$$\sum_{k\in P^i}\alpha_k\Big(W(N,b,c)-W(Q_{i+1},b^{i+1,n},c^{i+1,n})-W(P^i,b^{1,i},c^{1,i})\Big).\qquad(7.23)$$

Analogously it follows for the agents in Q_{i+1} that

$$\sum_{j\in Q_{i+1}}\Big(TIBS^{\alpha}_j(N,b,c)-TIBS^{\alpha}_j(Q_{i+1},b^{i+1,n},c^{i+1,n})\Big)=$$
$$\sum_{k\in Q_{i+1}}\alpha_k\Big(W(N,b,c)-W(P^i,b^{1,i},c^{1,i})-W(Q_{i+1},b^{i+1,n},c^{i+1,n})\Big).\qquad(7.24)$$

Multiplying equation (7.23) with $\sum_{k\in Q_{i+1}}\alpha_k$ and equation (7.24) with $\sum_{k\in P^i}\alpha_k$ shows that the α-TIBS fairness property (5.11) in Axiom 5.1 is satisfied. □