ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Gudmundsson, Jens; Hougaard, Jens Leth

Working Paper

River pollution abatement: Decentralized solutions and smart contracts

IFRO Working Paper, No. 2021/07

Provided in Cooperation with:
Department of Food and Resource Economics (IFRO), University of Copenhagen

Suggested Citation: Gudmundsson, Jens; Hougaard, Jens Leth (2021) : River pollution abatement: Decentralized solutions and smart contracts, IFRO Working Paper, No. 2021/07, University of Copenhagen, Department of Food and Resource Economics (IFRO), Copenhagen

This Version is available at:
https://hdl.handle.net/10419/244335

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DEPARTMENT OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

IFRO Working Paper

River pollution abatement: Decentralized solutions and smart contracts

Jens Gudmundsson
Jens Leth Hougaard

2021 / 07

IFRO Working Paper 2021 / 07

River pollution abatement: Decentralized solutions and smart contracts

Authors: Jens Gudmundsson, Jens Leth Hougaard

JEL-classification: C7, D47, D62, Q52, Q25

Published: November 2021 (This paper is an update of an earlier version with the same number but a slightly different title, published September 2021)

See the full series IFRO Working Paper here:
www.ifro.ku.dk/english/publications/ifro_series/working_papers/

Department of Food and Resource Economics (IFRO)
University of Copenhagen
Rolighedsvej 23
DK 1958 Frederiksberg DENMARK
www.ifro.ku.dk/english/

River pollution abatement: Decentralized solutions and smart contracts*

Jens Gudmundsson

Department of Food and Resource Economics, University of Copenhagen, Denmark

Jens Leth Hougaard

Department of Food and Resource Economics, University of Copenhagen, Denmark

Abstract

In river systems, costly upstream pollution abatement creates downstream welfare gains. Absent adequate agreement on how to share the gains, upstream regions lack incentives to reduce pollution levels. We develop a model that makes explicit the impact of water quality on production benefits and suggest a solution for sharing the gains of optimal pollution abatement, namely the Shapley value of an underlying convex cooperative game. We provide a decentralized implementation through a smart contract to automate negotiations. It ensures a socially optimal agreement supported by fair compensations to regions that turn to cleaner production from those that pollute.

Keywords: River pollution, decentralized mechanism, Shapley value, water quality, smart contract
JEL: C7, D47, D62, Q52, Q25

1. Introduction

Water pollution is a global concern, driven in part by economic development as growing population, agriculture, and industry contribute to wastewater production (UN-Water, 2016). At least 2 billion people rely on polluted drinking sources (WHO, 2019), making waterborne diseases a leading cause of mortality in developing countries (Garg et al., 2018). Difficult trade-offs are in play: for instance, pesticide use may be a necessity for food security, but its effect on water quality is harmful to human, animal, and plant life (see e.g Lai, 2017; FAO, 2018). We focus on the negative externalities that upstream pollution causes on downstream river regions. Pollution abatement comes at a cost to one region but benefits many, leading to an intricate problem in which the latter regions may be willing to, but cannot agree on how to, compensate the former (say by supporting investments in environmentally-friendly production facilities). The effectiveness of relying on a central authority to settle this has been questioned (see e.g. Sigman, 2005; Cai et al., 2016),

*We would like to thank Dagim Belay, Frank Jensen, Goytom Abraha Kahsay and Z. Emel Öztürk for helpful comments. The authors gratefully acknowledge financial support from the Carlsberg Foundation (grant no. CF18-1112).

Email addresses: jg@ifro.ku.dk (Jens Gudmundsson), jlh@ifro.ku.dk (Jens Leth Hougaard)

and oftentimes there simply is no such authority (e.g. for transboundary rivers). Therefore, as in the work of Orstrom and co-authors (see e.g. Orstrom, 1990), we focus on a decentralized solution in which the affected regions negotiate without external involvement. This ensures that regional decision rights are respected and that the resulting unanimously agreed-upon solution is stable. In equilibrium, our solution implements fair compensations to the regions that reduce their pollution.

While there are many ways to assess water quality in practice, they generally have in common that they measure pollutant *concentration* (say the amount of pollutants per liter of water).[1] Existing game-theoretical models on water extraction are explicit on water inflows at various locations of the river (see e.g. Ambec and Sprumont, 2002), but this aspect has been left out in models of river pollution; in effect, these models typically focus on the pollutant quantity rather than its concentration and disregard changes in water quality along the river (Section 2 expands on this). However, these are essential factors in determining optimal abatement as well as in assessing each region's claim on the joint welfare gain from cooperation. For instance, a large downstream inflow, say due to two rivers merging, reduces the importance of upstream abatement. In this work, we develop a novel model that captures water inflows and adjusts production benefits by water quality.

In the model, the most upstream region has an inflow of clean water that it can use for production. Some of the used water naturally dissipates while the rest becomes polluted. The region can choose more environmentally-friendly production to reduce the generated pollution, but this is costly and only benefits the downstream regions. As a baseline, we assume a constant marginal cost of pollution abatement, but hint also on how this can be generalized considerably. The polluted water from production mixes with the clean unused water before entering the next region. In this, the second-most upstream region, there is again an inflow of clean water. Hence, the total water flow typically varies along the river based on dissipation and inflows. Likewise, water quality, defined as the fraction of the available water that is clean, fluctuates based on the abatement decisions. Absent any abatement agreements, we expect that regions do not take the negative externality of their pollution into account. However, welfare improvements can arise if they do: upstream pollution abatement leads to downstream welfare gains. We seek to maximize social welfare as given by the total benefit of the river system net the total abatement costs.

Our first result, Proposition 1, gives a precise expression for the water quality in each region as a function of the abatement decisions. Specifically, we find that region i's abatement affects region k's water quality linearly, albeit in a discounted form based on the production carried out by all intermediate regions j. With resemblance to classic "bang-bang" control (see e.g. Conrad and Clark, 1987), this makes it simple to identify the socially optimal abatement plan. Namely, a region should abate fully (that is, improve downstream water quality as much as possible) if the total

[1] See, for instance, the criteria established by the US Environmental Protection Agency (`https://www.epa.gov/wqc`) and the standards set by the EU (`http://data.europa.eu/eli/dir/2008/105/oj`). A very concrete example is found in Yang et al. (2021).

marginal benefits to the other regions exceed the marginal cost; otherwise, the region is better off fully abstaining from abatement (Theorem 1). Thereafter, we turn to the question of how to share the welfare gain from optimal abatament between the regions. To identify a fair distribution, we introduce the *abatement game*, a cooperative game in which the worth of a coalition is given by the largest welfare gain attainable by the coalition without support from the other regions. This game turns out to have a particular structure; specifically, it is a so-called *activity optimization game with complementarity* (Topkis, 1987, 2011). This implies that the game is convex (Theorem 2), and that the *Shapley value* (Shapley, 1953) of the game provides a particularly appealing way of sharing the gains between the regions (see e.g. Moulin, 1988). This solution could in principle be implemented by a central agency with complete information and control, but we opt instead for a decentralized approach. That is to say, we seek a way for the regions to interact without external involvement for which equilibrium behavior results in payoffs coinciding with the Shapley value.

Specifically, we adapt the *bidding mechanism* of Pérez-Castrillo and Wettstein (2001) when sharing the welfare gain from optimal abatement. The key idea of this mechanism is for one region to propose a solution for the others to accept or reject; in addition, there is a first stage in which each region bids to become the proposer. The (subgame-perfect Nash) equilibrium payoffs of this non-cooperative game coincide with the Shapley value of the abatement game (Theorem 3). Note that the purpose of the mechanism is only to share the welfare gains. Hence, we imagine a situation in which the regions first agree on the stakes involved in terms of individual welfare changes. The bidding mechanism is then used to share the common gain. This differentiates our solution from related mechanisms such as those in Gudmundsson et al. (2019), where the division of the common welfare gain is preceded and affected by a stage in which the stakes are determined strategically. While the mechanisms of Gudmundsson et al. (2019) are developed for the river sharing model, they can be adapted to the present pollution abatement context as well (see Section 9). There are pros and cons with the two approaches: while Gudmundsson et al.'s (2019) mechanisms address the problem as a whole, this comes with significant restrictions on how the welfare gains can be shared. In contrast, the current approach offers more flexibility in sharing the gains and has the further advantage of being inherently finite.

We concretely show that the mechanism is well suited for practical use by providing a prototype for this purpose. The design uses cryptography, smart contracts, and blockchain technology. Together, these provide an ideal environment for decentralized solutions, eliminating the need for trusted third parties and reducing transaction costs. We argue both in general terms for its usefulness in economic design as well as specifically for the application of pollution abatement. Compared to having a central agency impose the solution, it lets the affected regions reach agreement without external actors interfering, saving time and money. Moreover, it resolves issues of trust and verification: with a central agency, the individual regions face the uncertainty of whether the agency acts as intended. In contrast, the smart contract is publicly available, the "rules of the game" are

fixed from the outset, and transactions are executed automatically with minimal delay.

Lastly, we conduct a simulation study to investigate how the Shapley value distributes the welfare gain from socially optimal abatement. That is to say, what regional characteristics typically yield high payoffs? We find that the regions that reap the largest share of the welfare gain are typically located close to the "middle" of the river, as these are in ideal position to both benefit from upstream abatement as well as contribute to downstream gains by abating themselves. Moreover, it is favorable to have high-valued production, to optimally abate fully, and to be situated in such a way that the abatement efforts trickle far downstream.

The paper is outlined as follows. In Section 2, we review the existing literature. In Section 3, we introduce the model. In Section 4, we examine the problem of maximizing social welfare. In Section 5, we define and analyze the cooperative game. In Section 6, we describe the bidding mechanism and how to implement it in practice. In Section 7, we examine an extensive numerical example. In Section 8, we conduct the simulation study. We conclude in Section 9. Proofs and technical details are postponed to the Appendix.

2. Related literature

Our paper relates to several strands of literature, but most obviously to the literature on game-theoretic analysis of river sharing problems initiated by the seminal paper of Ambec and Sprumont (2002). This has been followed by a stream of papers including, for instance, Ambec and Ehlers (2008), Ambec et al. (2013), Ansink and Weikard (2009, 2012), van den Brink et al. (2012), Gudmundsson et al. (2019), and Öztürk (2020). In its original form, the river sharing model concerns welfare-maximizing water extraction along a river shared by multiple regions implemented via associated side payments (see e.g. Beal et al., 2013, for a survey).

A notable branch of this literature focuses on the costs rather than the benefits of water extraction. Ni and Wang (2007) address the problem of sharing pollution costs along a river. In their work, a central agency determines the abatement cost of each region and two methods for allocating the total abatement costs are analyzed. The first method, *Local Responsibility Sharing*, establishes that each region carries its own direct abatement cost. The second method, *Upstream Equal Sharing*, asserts that regions should share equally the abatement cost for each segment comprised of the region itself and all regions situated upstream from it. Although the model of Ni and Wang (2007) constitutes an interesting first step towards an understanding of cost-sharing issues under upstream-downstream externalities, their model is arguably too stylized to capture important aspects of transboundary pollution. Several papers have aimed to address some of the shortcomings. For instance, Alcalde-Unzu et al. (2015) argue that the model of Ni and Wang (2007) does not account for how pollution is transferred downstream: *Local Responsibility Sharing* simply ignores such transfers, while *Upstream Equal Sharing* implicitly assumes equal responsibilities for the pollution at a given region between the region itself and all the upstream regions, which is unlikely to

be the case in practice. Consequently, Alcalde-Unzu et al. (2015, 2021) explicitly introduce the fact that pollution is transferred from upstream to downstream regions at a particular rate. This rate is assumed to be unknown, which creates uncertainties around regional responsibilities represented by responsibility ranges. Alternative cost sharing rules can be found, for instance, in Dong et al. (2012), van den Brink et al. (2018), and Sun et al. (2019). These papers typically take an axiomatic approach to fair allocation.

A more radical change of the model is found in Gengenbach et al. (2010), who consider exogeneous pollution levels and model agents' optimal choice of abatement under a linear damage function and a strictly convex cost function. For every region i, net benefits from abatement therefore equal i's benefit from aggregate abatement by all upstream regions (including i) less the cost of i's own abatement. Equilibrium abatement is then analyzed under various coalition structures. In a somewhat related framework, Steinmann and Winkler (2019) consider optimal abatement choices under strictly increasing and strictly convex costs, but focus on the allocation of welfare gains from full cooperation, providing a new justification for the original "downstream incremental" solution of Ambec and Sprumont (2002). Lastly, van der Laan and Moes (2016) consider a model set-up with optimal pollution choice and payoff functions where region i's benefit only depends on i's own pollution whereas i's costs depend on the pollution levels of i and all upstream regions. They analyze allocation of welfare gains from collaboration among all regions.

In comparison, our model also considers optimal choice of abatement level but under linear costs. We further add the aspect of pollution transfers downstream and its effects on water quality (and thereby benefit from water use), taking into account both dissipation from production as well as potential inflows of clean water in each region. In particular, this is in line with an aspect pointed out by Hou et al. (2019), who argue that the river itself may reduce the pollution as it flows downstream. They model this through a wastewater treatment rate. As in many of the previous papers, we focus on allocation of the welfare gains obtained from full cooperation, but point at how a natural solution (here, the Shapley value) can be implemented by a decentralized mechanism.

3. Preliminaries

We consider a stylized model of a river, studying both the benefits of water extraction and the ensuing drawbacks as water usage pollutes the water, diminishing its usefulness to downstream regions. Costly abatement efforts can, and optimally should, be taken to restore water quality. However, a region has no incentive to abate unless it gets compensated for doing so by the benefiting downstream regions.

3.1. Model

The river flows through **regions** $N = \{1, 2, \ldots, n\}$, where 1 is most upstream. The regions are economic agents who can represent anything from small local farmers to entire countries. Each

region i has an **inflow** of $e_i \geq 0$ units of clean water, say due to precipitation. It also has access to a production facility that it can divert water to. We take each region i's amount $y_i \geq 0$ of water diverted for **production** as exogenous. While this allows us to focus exclusively on pollution abatement, we discuss a weakening of this assumption in Section 9. Once used in production, some of the water dissipates. This naturally depends on the type of production: presumably more water dissipates if used to irrigate fields than if used to operate a hydro power plant. For simplicity, we assume a constant and common **dissipation** rate $\delta \in [0,1]$: with y_i units of water used for production, δy_i units disappear from the river system and the remaining $(1-\delta)y_i$ units get polluted. Thus, the total amount of water available to region j is $t_j \equiv e_j + \sum_{i<j}(e_i - \delta y_i)$.

The benefits accrued from production depend on the water quality. Each "unit of water" is either clean or polluted, and the **quality** $q_i \in [0,1]$ of the water at i's disposal is the fraction that is clean. Hence, $q_1 = 1$, but $q_2, \ldots, q_n$ may be smaller. While there are many forms of water pollution in practice, we have in mind here pollution that is perfectly dissolved and mixed with the clean water. It is important to maintain a high water quality as production creates quality-adjusted **benefit** $q_i b_i$ to region i.[2] Hence, $b_i \geq 0$ is the benefit from using clean water ($q_i = 1$), while fully polluted water ($q_i = 0$) is useless. Again, in practice there are many sources for this benefit and the dependence on high-quality water likely varies considerably: for pure consumptive use, water quality is essential to prevent illnesses; for operating a hydro power plant, it may be less so. For tractability, we opt for the simple form specified above.

The marginal **abatement cost** $c \geq 0$ is constant and common to all regions and implies that cleaning x_i units of water comes at total abatement cost cx_i. Cleanup is limited to the polluted water that exits production, so $x_i \leq (1-\delta)y_i$. We can think of this as the region installing some filter at the end of its production line or, more generally, opting for a more environmentally-friendly production technology (see e.g. Anawar and Chowdhury, 2020). Thus, the set of abatement schemes is $X = \{x \in \mathbb{R}^n_{\geq 0} : x \leq (1-\delta)y\}$. Given $x \in X$, the **social welfare** $W(x) \in \mathbb{R}$ is total quality-adjusted benefits minus total abatement costs:

$$W(x) = \textstyle\sum_i q_i(x) b_i - c \sum_i x_i.$$

In what follows, we seek $x \in X$ to maximize W. This is achieved through costly effort x_i exerted by region i to improve water quality $q_k(x)$ along with quality-adjusted benefits $q_k(x)b_k$ for downstream regions $k > i$. We label the optimal abatement $x^* \in X$.[3] Absent any agreements

[2] We may also model damages d_i alongside benefits b_i. For instance, the total effect on region i may instead be $q_i b_i - (1-q_i)d_i = q_i(b_i + d_i) - d_i$. As the final term is simply an additive constant that has no impact on the decisions, we recover the present model without loss. Hence, we may interpret b_i as both benefit as well as absence of damage of clean water.

[3] There may be a continuum of maximizers of W, but they have a common structure; see Theorem 1. By x^* we mean the (unique) maximizer x of W that maximizes $\sum_i x_i$. In terms of Theorem 1, "indifferences" are broken towards abating.

between the regions, we assume that there will be no pollution abatement; we label this $x^0 = (0, \ldots, 0)$. Throughout, all information is common knowledge. Figure 1 illustrates the model.

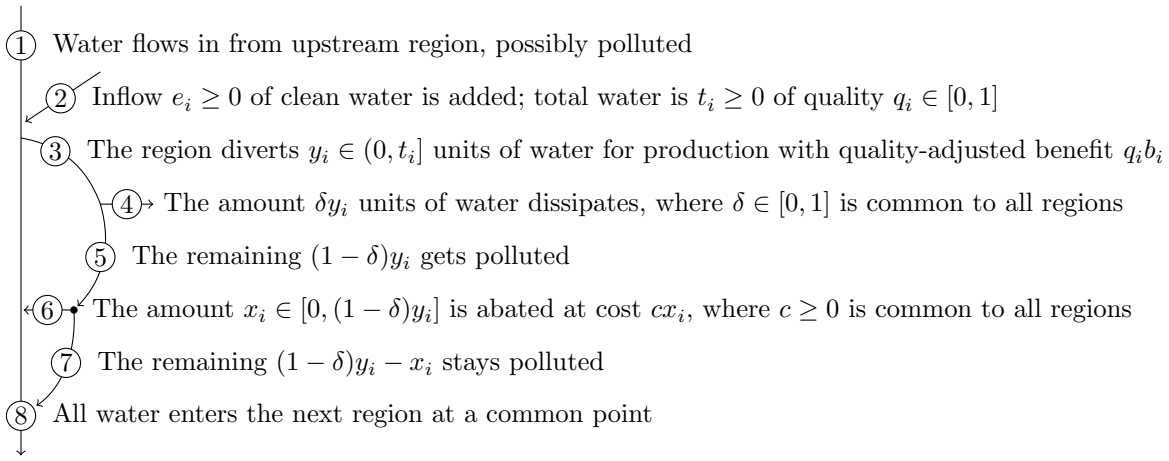

Figure 1: Water flow through a generic region i. Our key point of interest is the abatement decision (black node).

3.2. Simple extensions

It is straightforward to generalize the model in many directions. Rather than a linear river, we could allow a more general structure in which different "subrivers" merge and fork, water and pollution splitting up accordingly. We can weaken the assumptions that the inflow is fully clean and that the post-production water is fully polluted. A practical effect of climate change is that water flows have become more variable and unreliable; this could be captured by letting inflows be stochastic and instead have regions make decisions based on expectations. Moreover, the dissipation rate and abatement costs could vary between regions. These changes only amount to more cumbersome notation and more involved equilibrium expressions. Further extensions are discussed in Section 9.

4. Efficiency: Maximizing social welfare

In this section, we identify the optimal abatements x^*. First, we find a concise expression for the water quality q_k as a function of abatements x. Specifically, Proposition 1 shows that q_k is linear in x_i for $i < k$. Extending on this, we find that also social welfare W is linear in x_i, so region i's optimal decision is a corner solution: either it chooses full abatement or no abatement at all. The deciding thresholds are identified in Theorem 1.

4.1. Computing water quality

Recall that the water quality is the fraction of the total water that is clean. The water cleaned by region i travels through intermediate regions j to eventually affect the water quality in region k.

Specifically, the water cleaned by i that is not used by any intermediate j affects quality q_k positively. In this way, there is a "discounting effect": the further i is from k, the more regions "use up" i's cleaned water to diminish the impact that i's abatement x_i has on k's water quality q_k. Proposition 1 makes this precise.

Proposition 1. *For each scheme $x \in X$ and region $k \in N$,*

$$q_k(x) = \frac{1}{t_k} \sum_{i \leq k} e_i \prod_{i \leq j < k} \left(1 - \frac{y_j}{t_j}\right) + \frac{1}{t_k} \sum_{i < k} x_i \prod_{i < j < k} \left(1 - \frac{y_j}{t_j}\right).$$

The difference in "discounting" arises as j's inflow e_j enters before j's production, while j's cleaned water x_j is only factored in after j's production. To ease notation, we let $\alpha_{ik} \geq 0$ denote the marginal effect of region i's abatement on the quality-adjusted benefit of region $k > i$:

$$\alpha_{ik} \equiv \frac{\partial q_k}{\partial x_i} \cdot b_k = \frac{b_k}{t_k} \prod_{i < j < k} \left(1 - \frac{y_j}{t_j}\right).$$

Upstream regions are not affected by downstream abatement, so $\alpha_{ki} = 0$ for $k \geq i$.

4.2. Socially optimal abatement

Proposition 1 shows that quality q_k is linear in x_i. To be more precise, q_k is typically increasing in x_i for $k > i$.[4] As pollution only flows in one direction, regions upstream of i, so $k < i$, are not affected by x_i. On the other hand, i itself is affected only through the abatement cost cx_i, which again is linear in x_i. Hence, when aggregated over all regions, social welfare is also linear in x_i, so the optimal level x_i^* is found at one of the extremes: either $x_i^* = 0$ or $x_i^* = (1 - \delta)y_i$.

Theorem 1. *The socially optimal scheme x^* is such that, for each region i, either $x_i^* = 0$ or $x_i^* = (1 - \delta)y_i$. Specifically, $x_i^* = (1 - \delta)y_i$ if and only if $\sum_k \alpha_{ik} \geq c$.*

The threshold established in Theorem 1 depends only on the factors exogenous to the model (that is, each α_{ik} is independent of x). Region i's abatement decision boils down to whether the marginal impact it has on all other regions outweigh the marginal abatement costs. Having identified the optimal abatements, we turn to the question of how to share the induced welfare gains.

5. Stability and fairness: Cooperative game

In this section, we introduce the *abatement game*, a cooperative game (N, v) in which $v(S)$ is the largest welfare gain attainable by coalition $S \subseteq N$ when the other regions $N \setminus S$ free-ride on

[4] However, not always. If $y_j = t_j$ for some $i < j < k$, so j's production acts as a "barrier" between i and k, then no clean water is transferred from i to k and q_k is independent of x_i.

the abatement efforts of S and abstain from abating themselves. First, we formally define the characteristic function v. Thereafter, Theorem 2 shows that the abatement game is convex; specifically, it shows that the abatement game is a so called *activity optimization game with complementarity*. We then argue that a natural solution to it is the game's Shapley value.

5.1. Defining the abatement game

Let $W_S(x) \in \mathbb{R}$ denote the welfare of the regions $S \subseteq N$ at abatement scheme $x \in X$:

$$W_S(x) = \textstyle\sum_{i \in S} q_i(x) b_i - c \sum_{i \in S} x_i.$$

Let $X_S \subseteq X$ be the schemes that the regions in S can implement without support from regions outside S. That is to say, $x \in X_S$ sets $x_i = 0$ for all non-members $i \notin S$:

$$X_S = \{x \in X : i \notin S \implies x_i = 0\}.$$

For instance, $x^0 \in X_S$. We are interested in the welfare gain at $x \in X_S$ compared to the status quo x^0. Using Proposition 1 and expressed using the marginal benefits α_{ik}, this takes on a simple form:

$$\begin{aligned} W_S(x) - W_S(x^0) &= \sum_{k \in S} \left(q_k(x) - q_k(x^0) \right) b_k - c \sum_{i \in S} x_i \\ &= \sum_{k \in S} \frac{b_k}{t_k} \sum_{i<k} x_i \prod_{i<j<k} \left(1 - \frac{y_j}{t_j} \right) - c \sum_{i \in S} x_i \\ &= \sum_{i \in S} x_i \sum_{k \in S} \alpha_{ik} - c \sum_{i \in S} x_i. \end{aligned}$$

Absent external support from $N \setminus S$, the regions in S can guarantee welfare gain $v(S) \geq 0$:

$$v(S) = \max_{x \in X_S} W_S(x) - W_S(x^0) = \max_{x \in X_S} \sum_{i \in S} x_i \sum_{k \in S} \alpha_{ik} - c \sum_{i \in S} x_i.$$

Let $x^S \in X_S$ be the optimal abatement scheme for S. Proposition 2 is analogous to Theorem 1 and its proof is omitted. The difference is that the welfare gain only includes the regions in S (meaning that the sum is taken over $k \in S$), while the "discounting effect" remains unchanged (so the product present in α_{ik} remains over all $i < j < k$).

Proposition 2. *For each coalition $S \subseteq N$, the optimal abatement scheme x^S is such that, for each region $i \in S$, either $x_i^S = 0$ or $x_i^S = (1 - \delta) y_i$. Specifically, $x_i^S = (1 - \delta) y_i$ if and only if $\sum_{k \in S} \alpha_{ik} \geq c$.*

Proposition 2 implies that abatement incentives increase with the size of the coalition. If region i optimally abates fully given $S \subseteq N$, then i does the same for all $T \supseteq S$. Going the other way,

if i abstain from abatement in T, then i abstain in $S \subseteq T$ as well. In particular, if i abstain from abatement in N—that is, if it is socially optimal for i to abstain—then i abstain in all coalitions.

5.2. Activity optimization games with complementarity, convexity, and the Shapley value

In the proof of Theorem 2 below, we show that the abatement game is an instance of a so-called "activity optimization game with complementarity" (Topkis, 1987, 2011). Indeed, this holds for a considerably broader class of cost functions than the one with constant marginal costs studied so far: the proof is provided for any continuous and submodular cost functions $C(x, S)$.[5] This includes all functions separable across regions such as $C(x, S) = \sum_{i \in S} C_i(x_i)$ for functions $C_i \colon \mathbb{R}_{\geq 0} \to \mathbb{R}$. It also allows interesting cases in which abatement is cheaper the cleaner the water is, say $C(x, S) = c \sum_{i \in S}(1 - q_i(x))x_i$. Note that this is not separable as $q_i(x)$ depends on the upstream regions' abatements.

An important property of activity games with complementarity is that they are convex (Topkis, 2011, Theorem 5.4.1).[6] In convex games, each player j's marginal contribution, $v(S \cup \{j\}) - v(S)$, weakly increases with the size of the coalition S; a simple example is when each economic agent holds a unique input and the inputs are complementary (Section 5 in Topkis, 2011, offers more examples). Theorem 2 summarizes the discussion thus far.

Theorem 2. *The abatement game (N, v) is convex.*

The **core** $\mathcal{C}(N, v) \subseteq \mathbb{R}^N$ of a cooperative game (N, v) consists of all payoffs such that each coalition S is adequately compensated. In other words, an agreement based on core payoffs ensures that the welfare-maximizing solution of the grand coalition N cannot be overthrown by any coalition $S \subseteq N$ acting in their self interest:

$$\mathcal{C}(N, v) = \left\{ x \in \mathbb{R}^N : \sum_{i \in N} x_i = v(N) \text{ and, for each } S \subseteq N, \sum_{i \in S} x_i \geq v(S) \right\}.$$

For convex games, Shapley (1971) showed that the core has a particular structure and is non-empty. Thus, by Theorem 2, we can always allocate the welfare gains in such a way that the optimal solution is sustained. One way of doing so is through the celebrated **Shapley value** (Shapley, 1953), which always provides a core allocation for convex games (see e.g. Moulin, 1988; Peleg and Sudhölter, 2007, for further normative underpinnings). The solution awards region i a payoff according to i's weighted marginal contribution:

$$\phi_i(N, v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left(v(S \cup \{i\}) - v(S) \right).$$

[5] We redefine W_S naturally as $W_S(x) = \sum_{i \in S} q_i(x) b_i - C(x, S)$. Otherwise the game is unchanged.

[6] The game (N, v) is **convex** if, for each $j \in N$ and $S \subseteq T \subseteq N$, $v(S \cup \{j\}) - v(S) \leq v(T \cup \{j\}) - v(T)$.

Absent a central planner to enforce the Shapley value distribution between the regions, we will next seek a decentralized implementation of it. That is to say, we wish to design a game to be played between the regions for which equilibrium play yields Shapley value payoffs to the regions.

6. Decentralized implementation: The bidding mechanism

In this section, we introduce the *bidding mechanism* of Pérez-Castrillo and Wettstein (2001). This mechanism is appealing for several reasons. From a technical viewpoint, it implements the Shapley value payoffs in subgame-perfect equilibrium. That is to say, in every such equilibrium of this non-cooperative game, the players obtain the desired payoffs (compared to, say, obtaining these payoffs in expectation). The mechanism has the advantage that there is no pre-specified order on the players, say to fix the order in which they take action. Moreover, it is well suited for practical implementation as it is finite.

The bidding mechanism runs in multiple stages. It ends in one region proposing a solution for the others to sequentially accept or reject. In equilibrium, the proposal will be carefully constructed to award the others just enough that they accept, leaving as much as possible to the proposer. Thus, the proposer holds an advantageous position. For this reason, the proposal stage is preceded by a bidding stage in which the regions bargain for the proposal right. The mechanism very concretely ensures that the proposer makes a fair proposal: each region has veto power, and if the proposal is rejected by anyone, then the proposer gets excluded from the negotiations (which reset with a new bidding round). Intuitively, the region leaves a proposal for the others to contemplate and exits the negotiations not to be let in again—if all regions accept the proposal, it is implemented; if some region does not, then the negotiations proceed in the same format but the current proposer no longer participates.

Next, we define the bidding mechanism. Thereafter, we analyze a particular equilibrium of the induced game. Finally, we provide a practical implementation of the mechanism through a smart contract.

6.1. Formalizing the bidding mechanism

There are two stages in the bidding mechanism. The first always includes payments between some regions while the second only does so if the proposal is accepted (in which case the mechanism terminates thereafter).

Bidding stage. Each region i bids $\beta_j^i \in \mathbb{R}$ to region j. This is interpreted as i being willing to pay β_j^i to j for i to get the proposal right. We may have $\beta_j^i < 0$, but we will see that equilibrium bids are non-negative in the present setting.

For each region i, we compute the net bid $B^i = \sum_j \beta_j^i - \sum_j \beta_i^j$. If positive, then i is willing to pay more to the others for the proposal right than they are willing to pay i. In this way, the larger B^i, the more "valuable" the proposal right seemingly is to i. The proposer is selected as the region

with the highest net bid. If there are several to choose from (which happens in the theoretical equilibrium), then any selection from these can be done (for instance, a uniformly random draw appears appealing). At this point, all bids by the proposer are paid out. Hence, with i denoting the proposer, region $j \neq i$ receives β_j^i from i. All other bids are discarded.

Proposal stage. The proposer i leaves an offer for the others to consider and exits the negotiations. Specifically, the proposer offers $\gamma_j \in \mathbb{R}$ to each region $j \neq i$. In sequence, the regions evaluate the proposal and choose whether to accept or reject it. If all accept, then the proposal is implemented: each region $j \neq i$ then exits with a total payoff $\beta_j^i + \gamma_j$ while the proposing region i keeps the rest, $v(N) - \sum_{j \neq i} \left(\beta_j^i + \gamma_j\right)$. On the other hand, if some region rejects the proposal, then we return to the bidding stage and i is excluded from the negotiations. The bids β_j^i already paid out are not refunded, so the proposer has a strong incentive to make a fair offer that all regions accept. At the same time, once i is excluded, the amount at stake changes (typically reduces) to $v(N \setminus \{i\})$, which limits the incentives for the other regions to reject the proposal.

6.2. Equilibrium behavior

We restate the result of Pérez-Castrillo and Wettstein (2001) adapted to the present context. It shows that the bidding mechanism provides a decentralized implementation of the Shapley value. That is to say, as long as the regions agree on the negotiation protocol—that is, to use the bidding mechanism—no "social planner" or third parties are needed to achieve the desired outcome. Once the rules of the game are in place, the interaction between the regions, each acting individually and in their own self interest, yields a fair division of the optimal welfare gains.

Theorem 3. *(Pérez-Castrillo and Wettstein, 2001, Theorem 1) The bidding mechanism implements the Shapley value of the abatement game in subgame-perfect Nash equilibrium.*

Pérez-Castrillo and Wettstein (2001) also provide an intuitive equilibrium that we will reconstruct here. Let $\phi(N) \equiv \phi(N, v)$ denote the Shapley value of the abatement game (suppressing v), where $\phi_j(N)$ is region j's payoff. Let also $\phi(N \setminus \{i\})$ denote the Shapley value of the reduced game that arises when region i is excluded from partaking (which occurs if i makes a proposal that gets rejected). Then the equilibrium is as follows: In the bidding stage, each region i bids $\beta_j^i = \phi_j(N) - \phi_j(N \setminus \{i\})$ to region $j \neq i$. In the proposal stage, with i denoting the proposer, region $j \neq i$ is offered $\phi_j(N \setminus \{i\})$. When evaluating the proposal, region $j \neq i$ accepts the offer if and only if j is awarded at least $\phi_j(N \setminus \{i\})$. In this equilibrium, the negotiations end in the first "round"—no proposal is ever rejected. The proposer i offers precisely the amount to j that j is willing to accept. Final payoffs coincide with the Shapley value: for each region $j \neq i$, $\beta_i^j + \gamma_j = \phi_j(N) - \phi_j(N \setminus \{i\}) + \phi_j(N \setminus \{i\}) = \phi_j(N)$. The payoffs of the game distribute the same amount as the Shapley value does, namely $v(N)$, so the proposer i obtains $\phi_i(N)$ as desired.

Taking a closer look at the equilibrium bids, $\beta_j^i = \phi_j(N) - \phi_j(N \setminus \{i\})$, the "balanced contributions property" (Myerson, 1980) implies that they will be symmetric. Hence, the bid that i makes to j coincides with the bid that j makes to i. For this reason, the net bids B^i add to zero for every region i—that is, in equilibrium, anyone can be chosen as the proposer.

In Pérez-Castrillo and Wettstein's (2001) original formulation, they require the underlying game merely to be *zero-monotonic*.[7] Here, Theorem 2 showed that the abatement game is convex, which implies zero-monotonicity. For convex games, Sprumont (1990) shows that the Shapley value is *population monotonic*. This means that, the more regions that cooperate, the higher the award every region. That is to say, not only is there an increase in what is at stake (that is, the sum of payoffs), but the Shapley value even ensures that every region sees their payoff go up. Most importantly, $\phi_j(N) \geq \phi_j(N \setminus \{i\})$. This implies that i's equilibrium bid to j, $\phi_j(N) - \phi_j(N \setminus \{i\})$, is non-negative. The same holds true for i's proposed offer $\phi_j(N \setminus \{i\})$ to j.

6.3. Blockchains and smart contracts

We will now argue that the bidding mechanism is well suited for practical implementation. In what follows, we provide a simplified introduction to the relevant technology consisting of blockchains and smart contracts.[8] This is intended to show the usefulness of smart contracts for practical mechanism design in general; in Subsection 6.4, we turn to the particular case of the bidding mechanism.

A *smart contract* is a piece of code that governs a set of variables and provides functions to modify these variables. The code is publicly available and can be inspected by all parties before use to ensure that it works as intended. Interactions with the contract occur through *transactions*, which may specify functions (in the contract) to run as well as inputs to run them on. An elementary feature is that a transaction may transfer value between accounts through an associated cryptocurrency. This can for instance be from the user to the contract (say as a deposit) or the other way around (say by calling a "refund" function within the contract that returns the deposit from the contract's account).[9] For efficiency purposes, transactions are grouped together and ran sequentially in *blocks*. The blocks are cryptographically chained in the sense that each block contains a pointer to the block it extends on. This permits a consistent, global view of the current state of the contract: anyone can rerun all transactions from the contract's inception to the most recent block and thereby determine the current values of the contract's variables. Once the contract is deployed on the blockchain, it obtains a unique address and its code is forever fixed. In this way,

[7] The game (N, v) is **zero-monotonic** if, for each $S \subseteq N$ and $i \in N$, $v(\{i\}) + v(S) \leq v(S \cup \{i\})$. This is satisfied for the abatement game: the game is monotonic, $v(S \cup \{i\}) \geq v(S)$, and $v(\{i\}) = 0$.

[8] Many excellent sources cover these topics in greater detail; we refer the interested reader to Nakamoto (2008), Ferguson et al. (2010), Katz and Lindell (2014), Damgård et al. (2020), and `http://ethereum.org`.

[9] This provides a simple way to incentivize users: all may be required to make a deposit at the outset, but only those who act as intended get refunded in the end.

users are safe in knowing that no one can "override" the contract and make it do something beyond its intended functionalities—no one can for instance empty the contract's balance unless there is a function specifically for this purpose.

We will argue that smart contracts pose an ideal decentralized replacement for the social planners, auctioneers, and centralized clearinghouses prevalent in economic theory. For instance, tasks typically assigned a trusted auctioneer—receiving bids, identifying winners, transferring funds—can be automated through the contract. This not only eliminates the need for trust but also significantly reduces transaction costs.[10] While it will become apparent how smart contracts easily can be used to run sequential-move games, we will next explain how to go even further to also capture simultaneous-move games and random events.

In many economic interactions, it is desirable that agents act "simultaneously", that they individually make choices without information on the others' choices. While transactions in smart contracts inherently are "sequential" rather than "simultaneous", this can still be resolved through elementary cryptography (see e.g. Damgård et al., 2020). Specifically, a *cryptographic hash function* H maps inputs α of any size to outputs of a fixed size in such a way that computing $H(\alpha)$ is easy while reverse-engineering an input α from an output $H(\alpha)$ is hard. "Simultaneous" choices can then be achieved through a "commitment" and a "reveal" phase: first, each user submits her encrypted action, $H(\alpha)$; once all such transactions have been processed, each user submits her actual action α. The contract checks that the action matches the encrypted commitment. For instance, to run "rock, paper, scissors", each player selects a number $\alpha \in \mathbb{N}$, where $\alpha \bmod 3$ determines the associated action (say $\alpha \bmod 3 = 0$ is "rock", 1 is "paper", and 2 is "scissors"). Each player sends $H(\alpha)$—from which the other cannot infer α—and thereafter sends α. The contract can then determine the winner according to the rules of the game and potentially transfer funds. A similar approach can be used to generate "random" numbers. Again, each user commits to $H(\alpha)$, reveals α, and then the α's combine to generate a "random" number $H(\alpha_1\alpha_2\dots) \in \{0, 1, \dots, n\}$. For instance, with the secure hashing algorithm SHA-256, there are $n = 2^{256} - 1$ possible outputs. In this way, smart contracts provide a great way to run both simultaneous, sequential, and probabilistic mechanisms. Our implementation of the bidding mechanism employs many of these ideas.[11]

[10] There are costs associated to interacting with a smart contract as well. For the Ethereum blockchain, each instruction has a "gas" cost. For instance, for our implementation presented in Subsection 6.4, each region essentially transacts five times with the contract to alter different variables in its storage. On average, each such transaction costs roughly 100 000 gas. Each unit of gas costs a number of gwei (a basic unit of Ethereum's cryptocurrency) to execute. This price is set by the user. Setting it high ensures that high-priority transactions get registered within seconds. For the current application, it would seem a reasonable time frame is rather several hours, so users would get by with a considerably lower price. The transaction would then get registered during times of low network demand. All numbers below fluctuate considerably over time; the ones presented were observed in September, 2021. Gas prices are in the range of 20 gwei, so the five transactions of 100 000 gas each cost 10 million gwei, which is 0.01 ether. The exchange rate to USD at the time was 3 600 USD for one ether. Hence, executing the contract would cost each region roughly 36 USD. The code has not been optimized to any degree, so there are presumably ways to reduce this cost.

[11] We refer to `https://github.com/jensgudmundsson/AbatementSmartContract` for a complete prototype of the

6.4. Smart contract implementation of the bidding mechanism

In what follows, we describe how to practically implement the bidding mechanism through a smart contract in several phases. To make it fully operational in practice, some extra steps are added (for instance, the set of players is clear from the outset in the bidding mechanism, but we include a registration phase first). In the contract, a variable called `state` keeps track of the current phase and ensures that only valid operations are executed throughout. We detail the phases below.

Deployment of the contract. When initialized, the contract expects three parameters, namely the number of participating regions n, a possible deadline after which deposited funds can be refunded, and a fixed deposit F that each region will provide. Fixing the number of regions from the outset is necessary to know when to move from the registration to the deposit phase below (this could also be achieved by having a registration deadline). A deadline after which refunds are possible is to ensure that the regions' deposits do not get stuck in case someone makes a mistake when interacting with the contract and it fails to terminate. When regions first deposit to the contract, we take as given that they deposit the fixed deposit F and their welfare change from cooperation compared to the status quo, that is, region k deposits $F + \sum_i x_i \alpha_{ik} - c x_k$. The latter part may be negative, so F must be set large enough to ensure that deposits are non-negative. Once the mechanism terminates, each region is refunded the fixed deposit F.

Registration phase. Each region calls the contract's `register` function to link the region to its address on the blockchain. Intuitively, think of this as the region opening an "internal bank account" within the contract. Once all partaking regions have done so, we proceed to the next phase.

Deposit phase. Each region deposits an amount to the contract, here taken as the fixed deposit F and the region's welfare change from cooperation compared to the status quo. In this way, once all regions have made their deposits, the contract's balance is $nF + v(N)$. When calling the `deposit` function, the transacted funds are added to the "internal bank account" connected to the region. In this way, if the mechanism does not finish before the set deadline, anyone can call the `abort` function to refund all regions according to their "internal balances".

Commitment phase. Bidding stage, part 1: Each region submits their encrypted bid (corresponding to β). Specifically, to commit to the bid $[\beta_1^i, \beta_2^i, \dots]$, the region computes the hash value of the array $[\beta_1^i, \beta_2^i, \dots]$ and submits it to the `commit` function.[12]

Reveal phase. Bidding stage, part 2: Each region submits their actual bid together with enough funds to cover the bid. For instance, the transaction value for the bid $[\beta_1^i, \beta_2^i, \dots]$ needs to be at least $\beta_1^i + \beta_2^i + \dots$. The contract stores the bid if it matches the encrypted commitment. Once all

smart contract for the Ethereum blockchain.

[12] For convenience, our implementation provides a `DEBUGgetHash` to obtain this value.

regions have revealed their committed bids, the contract computes the net bids to determine the proposer. The proposer's bids are executed through transfers between the "internal bank accounts".

Proposal phase. Proposal stage, part 1: The proposer submits their proposal in the same way as when revealing their actual bid in the previous phase. Again, the transaction value needs to cover the proposal and the balance of the proposer's internal account gets increased accordingly.

Evaluation phase. Proposal stage, part 2: Each region accepts or rejects the proposal. If all accept, then the proposal is implemented. This is done first internally: the proposed amounts are deducted from the proposer's account and added to the other regions' accounts, while the initial deposits go the other way (that is, the proposer keeps $v(N)$ less the bids paid out). Once all internal balances are updated, the actual transfers back to the regions take place, and then the contract terminates. If instead some region rejects the proposal, then the contract resets in the following sense. First, all regions are paid out according to their internal accounts—this means that the rejected proposer makes a loss as her bids were paid out at the end of the reveal phase. Thereafter, the proposer is turned inactive for the remainder of the mechanism and the contract returns to the deposit phase.

For practical purposes, it may be desirable to postpone payment to the regions until they provide proof that they have abated as agreed. This can be achieved using tools of the "Internet-of-Things": the regions can position multiple sensors along the river, which report data directly to the smart contract. A region is then paid only if the reported water quality is adequate. To give an example, Singh et al. (2020) suggest an IoT sensor-based blockchain framework for temperature monitoring.

Taken together, the "time line" of the game is as follows. First, the regions conclude "what is at stake", that is, the individual welfare changes from optimal abatement. Thereafter, the regions interact through the smart contract. Hence, abatement levels remain at the status quo x^0 until agreement on how to share the gains is reached. Once concluded, each region is guaranteed some net payments from the contract: this can either be positive (so the region gets more refunded from the contract than it pays into it; this typically happens for abating countries) or negative. These funds get released once the region has abated according to the optimal plan (say made operational by measuring water quality along the river). For the abating regions, the positive net payments from the contract then cover the abatement costs as well as leaves the region some share of the joint welfare gain (namely their Shapley value payoff). For the other regions, the potentially negative net payments from the contract get offset by the benefit increase due to improved water quality; again, the end net effect equals their Shapley value payoff.

7. Numerical example

To illustrate all aspects of the preceding sections, we consider a four-region river with inflows $e = (8, 2, 2, 0)$. Let $\delta = 1/2$, so half the water used for production dissipates. The production plan is

$y = (4, 4, 8, 4)$, which yields water totals $t = (8, 8, 8, 4)$. Hence, the two last regions use all available water for production. Benefits are $b = (12, 48, 24, 32)$ and the unit cost is $c = 4$.

As a baseline, we can examine the status quo $x^0 = (0, 0, 0, 0)$: we obtain $q(x^0) = (1, 3/4, 5/8, 0)$, $u^0 \equiv q(x^0)b - cx^0 = (12, 36, 15, 0)$, and $W(x^0) = 63$. This is improved considerably when all but the last region abate fully, so $x = (2, 2, 4, 0)$, for which we obtain $q(x) = (1, 1, 1, 1)$ and $W(x) = 84$. However, the socially optimal scheme is $x^* = (2, 0, 4, 0)$, for which $q(x^*) = (1, 1, 3/4, 1)$, $u^* \equiv q(x^*)b - cx^* = (4, 48, 2, 32)$, and $W(x^*) = 86$. Thus, welfare changes for the individual regions by $u^* - u^0 = (-8, 12, -13, 32)$ and increases overall by $W(x^*) - W(x^0) = 23$. Hence, optimal abatement and water quality need not be monotonic as exerting costly abatement only is meaningful if it creates sufficient downstream benefit.

For the abatement game, the non-zero elements of the characteristic function are $v(\{1, 2\}) = v(\{1, 2, 4\}) = 4$, $v(\{1, 2, 3\}) = 7$, $v(\{3, 4\}) = v(\{1, 3, 4\}) = v(\{2, 3, 4\}) = 16$, and $v(N) = 23$. Table 1 shows the Shapley value $\phi(N)$ of the abatement game as well as those of the reduced games, $\phi(N \setminus \{i\})$. From this, we can deduce each region i's equilibrium bid $\beta_j^i = \phi_j(N) - \phi_j(N \setminus \{i\})$ and proposal $\gamma_j = \phi_j(N \setminus \{i\})$ to the other regions j.

Region	1	2	3	4
$\phi(N)$	3	3	9	8
$\phi(N \setminus \{1\})$		0	8	8
$\phi(N \setminus \{2\})$	0		8	8
$\phi(N \setminus \{3\})$	2	2		0
$\phi(N \setminus \{4\})$	3	3	1	

Table 1: The first row gives the Shapley value of the abatement game. Later rows present the Shapley values of the reduced games following one region's exclusion.

When initializing the smart contract, we set the fixed deposit F large enough to make all regions deposit a positive amount. Recall that the individual welfare changes are $(-8, 12, -13, 32)$. Hence, the largest individual loss created by the change from x^0 to x^* is 13 (region 3 pays for full abatement, 16, but only gains 3 from the quality increase due to region 1's abatement). Here, we set $F = 16$, but any $F \geq 13$ suffices. Hence, the regions will deposit $F + (-8, 12, -13, 32) = (8, 28, 3, 48)$. We refer to Table 2 for a summary of the "internal balances" throughout the phases of the contract.

To find region i's bid, we compute $\phi(N) - \phi(N \setminus \{i\})$ through Table 1. For region 1, we obtain $\beta^1 = (-, 3, 1, 0)$. In Table 2, row 3, we assume that each region i transfers exactly $\sum_j \beta_j^i$ in the reveal phase, but also larger deposits would be completely risk-free and simply held by the contract until it terminates and the excess is returned.

Recall, equilibrium bids are symmetric and net bids zero. Hence, any region can be chosen to be the proposer; in what follows, we choose region 1. In Table 2, row 4, internal transfers matching region 1's bids are conducted. From Table 1, we immediately find region 1's equilibrium

Region	1	2	3	4
1. Registration	0	0	0	0
2. Deposit of $F + \sum_i x_i\alpha_{ik} - cx_k$ for each region k	+ 8	+ 28	+ 3	+ 48
3. Deposit to cover committed bids	+ 4	+ 4	+ 10	+ 8
4. Proposer's bids carried out	− 4	+ 3	+ 1	
5. Deposit to cover proposal by region 1	+ 16			
6. Proposal accepted and carried out	− 16		+ 8	+ 8
7. Proposer gets $\sum_i x_i\alpha_{ik} - cx_k$ from each region k	+ 31	− 12	− (−13)	− 32

Table 2: Throughout the different phases of the contract, each region's "internal account" is updated following deposits and transfers.

proposal, namely $\gamma_1 = \phi(N \setminus \{1\}) = (-, 0, 8, 8)$. In Table 2, row 5, we assume that region 1 transfers precisely the amount needed to cover the proposal, namely $0 + 8 + 8 = 16$; again, larger amounts would work as well. If any region would reject the proposal, then all would be refunded their balances at this point. Region 1 would make a loss of 4 (the bids just transferred) while regions 2 and 3 respectively gain 3 and 1. However, the amount at stake would be reduced from $v(N) = 23$ to $v(N \setminus \{1\}) = 16$. If instead the proposal is unanimously accepted, then the proposed amounts are transferred away from the proposer, Table 2, row 6, while the individual welfare changes that make out $v(N)$ are transferred from the others to the proposer, row 7. The final balances then get refunded to the regions. The net effect of the interaction with the contract is derived by looking at the steps in which internal transfers are conducted, namely rows 4, 6, and 7 in Table 2. These amount to $(-4, 3, 1, 0) + (-16, 0, 8, 8) + (31, -12, 13, -32) = (11, -9, 22, -24)$.[13] Once the negotiations conclude, the actual abatement takes place: the abating regions 1 and 3 incur abatement costs while the others derive benefit from improved water quality. Specifically, recall that the individual welfare gains are $(-8, 12, -13, 32)$. Thus, once optimal abatement is implemented, the total welfare gain compared to the status quo is $(11, -9, 22, -24) + (-8, 12, -13, 32) = (3, 3, 9, 8)$, which is precisely the Shapley value payoffs.

8. Simulation study

To get a better understanding of who the "winners" are when distributing the common welfare gain according to the Shapley value of the abatement game, we conduct a simulation study. Instances are randomly generated with some fixed parameters. Specifically, we consider $n = 12$ regions, marginal costs $c = 4/5$, and dissipation rate $\delta = 1/2$. Water totals are kept constant at $t_i = 12$, so

[13] Alternatively, we can sum all entries in Table 2 to obtain $(39, 23, 35, 32)$. This is the amount paid back to the regions from the contract. From this, we subtract the amount paid into the contract by the regions, namely rows 2, 3, and 5, which add up to $(28, 32, 13, 56)$.

$e_1 = 12$ and otherwise $e_i = \delta y_{i-1}$. That is to say, i's inflow is just enough to cover the dissipation in $i-1$. Production y and benefits b are drawn uniformly and independently from $\{1, 2, \ldots, 12\}$. (The two may of course be correlated in practice. Here, they are independent, so we can distinguish the two effects.) We examine 10 000 instances, restricting to non-trivial instances with $v(N) > 0$. We describe in Appendix B how to quickly compute the characteristic function v for each instance.

Table 3 summarizes a regression on the output of the simulations. It shows how the Shapley value of region i is affected by a variety of variables.

Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
Location i	0.121	Benefit b_{i-1}	0.020	Production y_{i-1}	−0.029
Squared location i^2	−0.012	Benefit b_i	0.109	Production y_i	−0.018
Abatement	0.213	Benefit b_{i+1}	0.062	Production y_{i+1}	−0.067

Table 3: Regression on 100 000 observations (10 countries and 10 000 instances) with dependent variable being the Shapley value of region i. Regions 1 and n, which lack one neighbor, have been excluded. All variables are always significant, with even the smallest t-statistic above 40.

For the location, the results show a concave, quadratic relationship: being closer to the "middle" of the river is beneficial.[14] This is intuitive: these regions are relatively "close" to everyone and can benefit from cooperation with regions both up and down the stream. On the other hand, a coalition such as $\{1, n\}$ is rare to create any welfare gain, so the most up- and downstream regions fall short in this respect. Next, the variable *abatement* takes on value 0 or 1 depending on whether the region chooses full abatement in the social optimum. The results show that abatement has a large, positive impact. That is to say, the Shapley value favors regions that optimally abate. Indeed, the average payoff of an abating region is approximately 2.8 times the average payoff of a region that abstains from abatement.

Turning to the benefits, the results show that the awards increase with the value of the region's production. Indeed, we find that region i's award is increasing also in the benefits of the neighboring $i-1$ and $i+1$. It is increasing in b_{i+1} as i's abatement efforts then come with a greater welfare gain. A potential way to explain that it also increases in b_{i-1} is that the higher b_{i-1}, the easier it is to convince regions before $i-1$ to abate when i is added to the coalition. Lastly, production moves in the opposite direction. Recall that production and benefits are independent here: therefore, increased production merely increases the "discounting effect" of abatement. The higher y_i, the less abatement by regions upstream of i affects regions downstream of i. Hence, the larger y_i, the smaller the abatement incentives for the upstream regions. The intuition is the same for y_{i-1}, albeit with a smaller impact. Finally, the larger y_{i+1}, the smaller i's incentives for pollution abatement.

[14] The effect is not fully represented in the regression as the two extreme regions are skipped; compare Figure 2.

This effect is the largest of the three, as it pertains to i's abatement decision (which in itself was found to have a large effect).

To summarize, the regions that reap the largest share of the welfare gains are typically close to the middle of the river. They optimally adopt full abatement, have high-valued production (high b_i), and their abatement efforts trickle further downstream (small y_{i+1}).

Finally, Figure 2 illustrates some further observations from the simulation study. Regions are on the horizontal axis. On the left vertical, we measure the share of the Shapley value, corresponding to the solid curve. Specifically, in each instance, we compute each region's share of the total gain; these values are then averaged over all 10 000 instances. It confirms the intuition developed above: regions in the middle are best off. On the right vertical, we measure percentages. Specifically, the dashed curve shows how often the region abates in the socially optimal solution. This is at a fairly constant level before dropping off at the end. Intuitively, region i needs to affect at least a few downstream regions for abatement to be worthwhile (explaining the drop at the end), but the discount effect is so large that it does not make a big difference if it affects even more regions (explaining the constant level at the start). Lastly, the dotted curve shows the fraction of contribution vectors in which region i makes a positive contribution. Recall, the Shapley value is the average contribution that i makes in the $n!$ different orders. Here, we do not take the value of the contribution into account, we simply check whether the contribution is positive or not. This follows a similar trajectory as the Shapley value (the solid curve) except towards the end. We can interpret this as that the most downstream regions, in relative terms, contribute more often but in smaller amounts.

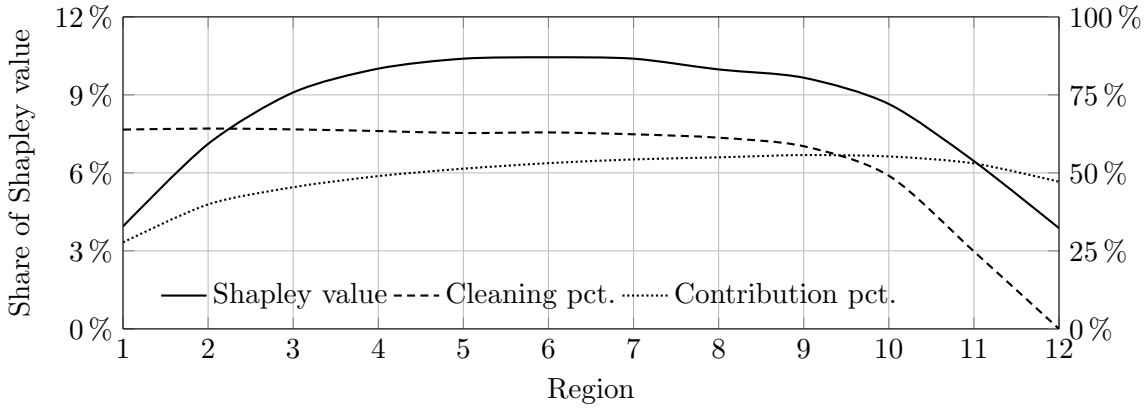

Figure 2: On the left axis, the region's share of the Shapley value is measured (averaged over 10 000 instances). For instance, region 3 gets on average 9% of the total gain. On the right axis, we measure (dashed) how frequently the region abates at the social optimum and (dotted) the fraction of contribution vectors in which the region makes a positive contribution.

9. Concluding remarks

We have studied fair allocation of the welfare gains of efficient river pollution abatement. For this purpose, we developed a novel model in which costly upstream pollution abatement increases downstream quality-adjusted benefits. We identified the optimal abatement levels and then suggested a fair way to share the welfare gain between the regions. Specifically, we defined an associated cooperative game to which the Shapley value was a natural solution. Thereafter, we proposed a decentralized implementation of this solution. In particular, we showed how Pérez-Castrillo and Wettstein's (2001) *bidding mechanism* can be implemented in practice through a smart contract, allowing the regions to negotiate without external involvement and reducing transaction costs. Lastly, we explored through a simulation study how the Shapley value distributed the welfare gain among the regions.

In addition to the extensions mentioned at the end of Section 3, there are several promising alternative specifications that we leave for future research.

First, our study does not concern the pollution decisions—the production plan is held fixed—but rather the thereon following abatement decisions. An interesting extension is to let the regions decide on both the production and abatement levels. Benefits would then depend increasingly on production (say b_i is a strictly increasing and concave function in the production y_i). If all water used for production dissipates, $\delta = 1$, so there in effect is no pollution, then we recover the river sharing model of Ambec and Sprumont (2002). In this particular setting, Gudmundsson et al. (2019) suggest a decentralized mechanism that involves announcements of production levels followed by a bargaining stage on welfare gains. This type of a mechanism extends naturally even when there is pollution to take into account. Specifically, in the first stage of Gudmundsson et al.'s (2019) mechanism, the regions would announce both production and abatement levels. Details on how the first-stage announcements affect the second-stage bargaining rights may need to be adjusted to the richer pollution setting. As with the bidding mechanism, it can be implemented through a smart contract.

Second, the status quo has been assumed to be that no region abates whatsoever. This has a decisive impact on the equilibrium payoffs: for instance, the most upstream region is compensated fully for all pollution abatement it undertakes. This is in some contrast to the "polluter pays principle", under which the polluter is held fully responsible. An alternative is to speak more generally of the status quo $x^0 \in \mathbb{R}^n_{\geq 0}$ rather than set it specifically to $(0, \ldots, 0)$. In this way, x^0_i would be i's mandatory abatement: in equilibrium, i presumably would cover this part, while any abatement on top of x^0_i would be covered by the downstream regions that benefit from the cleaner water. Interpreted in terms of international water law, such a change would move us towards the principle of *limited territorial sovereignty* and away from *absolute territorial sovereignty*.[15] Another

[15] As described by Salman (2007), ATS means that "a state is free to dispose, within its territory, of the waters of

alternative is to fix a minimum water quality $\bar{q} \in [0, 1]$ (say pertaining to a tipping point at which the river is beyond rescue) throughout. That is to say, pollution abatement has to be such that the water entering (or exiting) each region i is at least $\bar{q}$. This may lead to an interesting difference when computing the value of a coalition S for the cooperative game. Specifically, S now needs to take into account that the it abates, the less the non-members $N \setminus S$ are required to do.

References

Alcalde-Unzu, J., Gómez-Rúa, M., Molis, E., 2015. Sharing the costs of cleaning a river: the Upstream Responsibility rule. Games and Economic Behavior 90, 134–150.

Alcalde-Unzu, J., Gómez-Rúa, M., Molis, E., 2021. Allocating the costs of cleaning a river: expected responsibility versus median responsibility. International Journal of Game Theory 50, 185–214.

Ambec, S., Dinar, A., McKinney, D., 2013. Water sharing agreements sustainable to reduced flows. Journal of Environmental Economics and Management 66, 639–655.

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

Ambec, S., Sprumont, Y., 2002. Sharing a River. Journal of Economic Theory 107, 453–462.

Anawar, H.M., Chowdhury, R., 2020. Remediation of Polluted River Water by Biological, Chemical, Ecological and Engineering Processes. Sustainability 12, 7017.

Ansink, E., Weikard, H.P., 2009. Contested water rights. European Journal of Political Economy 25, 247–260.

Ansink, E., Weikard, H.P., 2012. Sequential sharing rules for river sharing problems. Social Choice and Welfare 38, 187–210.

Beal, S., Ghintran, A., Remila, E., Solal, P., 2013. The River Sharing Problem: A Survey. International Game Theory Review 15.

van den Brink, R., Heb, S., Huang, J.P., 2018. Polluted river problems and games with a permission structure. Games and Economic Behavior 108, 182–205.

Cai, H., Chen, Y., Gong, Q., 2016. Polluting thy neighbor: Unintended consequences of China's pollution reduction mandates. Journal of Environmental Economics and Management 76, 86–104.

an international river in any matter it deems fit, without concern for the harm or adverse impact that such use may cause to other riparian states". In contrast, LTS asserts that "every riparian state has a right to use the waters of the international river, but is under a corresponding duty to ensure that such use does not harm other riparians".

Conrad, J.M., Clark, C.W., 1987. Natural Resource Economics: Notes and Problems. Cambridge University Press.

Damgård, I., Nielsen, J.B., Orlandi, C., 2020. Distributed Systems and Security. `https://cs.au.dk/~orlandi/dsikdist`. Accessed 2020-11-04.

Dong, B., Ni, D., Wang, Y., 2012. Sharing a Polluted River Network. Environmental and Resource Economics 53, 367–387.

FAO, 2018. More people, more food, worse water? A global review of water pollution from agriculture. URL: `http://www.fao.org/3/CA0146EN/ca0146en.pdf`.

Ferguson, N., Schneier, B., Kohno, T., 2010. Cryptography Engineering: Design Principles and Practical Applications. John Wiley & Sons.

Garg, T., Hamilton, S.E., Hochard, J.P., Kresch, E.P., Talbot, J., 2018. (Not so) gently down the stream: River pollution and health in Indonesia. Journal of Environmental Economics and Management 92, 35–53.

Gengenbach, M.F., Weikard, H.P., Ansink, E., 2010. Cleaning a river: An analysis of voluntary joint action. Natural Resource Modeling 23, 565–590.

Gudmundsson, J., Hougaard, J.L., Ko, C.Y., 2019. Decentralized Mechanisms for River Sharing. Journal of Environmental Economics and Management 94, 67–81.

Hou, D., Lardon, A., Sun, P., Xu, G., 2019. Sharing a pollution river under waste flow control. Working paper .

Katz, J., Lindell, Y., 2014. Introduction to Modern Cryptography. 2 ed., Chapman & Hall.

van der Laan, G., Moes, N., 2016. Collective decision making in an international river pollution model. Natural Resource Modeling 29, 374–399.

Lai, W., 2017. Pesticide use and health outcomes: Evidence from agricultural water pollution in China. Journal of Environmental Economics and Management 86, 93–120.

Moulin, H., 1988. Axioms of Cooperative Decision Making. Cambridge University Press.

Myerson, R.B., 1980. Conference structures and fair allocation rules. International Journal of Game Theory 9, 169–182.

Nakamoto, S., 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. `https://bitcoin.org/bitcoin.pdf`. Accessed 2020-11-04.

Ni, D., Wang, Y., 2007. Sharing a polluted river. Games and Economic Behavior 60, 176–186.

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

Öztürk, Z.E., 2020. Fair social orderings for the sharing of international rivers: A leximin based approach. Journal of Environmental Economics and Management 101, 102302.

Peleg, B., Sudhölter, P., 2007. Introduction to the Theory of Cooperative Games. Springer.

Pérez-Castrillo, D., Wettstein, D., 2001. Bidding for the Surplus: A Non-cooperative Approach to the Shapley Value. Journal of Economic Theory 100, 274–294.

Salman, S.M.A., 2007. The Helsinki Rules, the UN Watercourses Convention and the Berlin Rules: Perspectives on International Water Law. Water Resources Development 23, 625–640.

Shapley, L.S., 1953. A value for n-person games. Annals of Mathematics Study , 307–317.

Shapley, L.S., 1971. Cores of convex games. International Journal of Game Theory 1, 11–26.

Sigman, H., 2005. Transboundary spillovers and decentralization of environmental policies. Journal of Environmental Economics and Management 50, 82–101.

Singh, R., Dwivedi, A.D., Srivastava, G., 2020. Internet of Things Based Blockchain for Temperature Monitoring and Counterfeit Pharmaceutical Prevention. Sensors 20, 3951.

Sprumont, Y., 1990. Population monotonic allocation schemes for cooperative games with transferable utility. Games and Economic Behavior 2, 378–394.

Steinmann, S., Winkler, R., 2019. Sharing a River with Downstream Externalities. Games 10, 23.

Sun, P., Hou, D., Sun, H., 2019. Responsibility and sharing the cost of cleaning a polluted river. Mathematical Methods of Operations Research 89, 143–156.

Topkis, D.M., 1987. Activity optimization games with complementarity. European Journal of Operational Research 28, 358–368.

Topkis, D.M., 2011. Supermodularity and Complementarity. Princeton University Press.

UN-Water, 2016. Towards a Worldwide Assessment of Freshwater Quality. URL: `https://www.unwater.org/publications/towards-worldwide-assessment-freshwater-quality/`.

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

WHO, 2019. Factsheet on drinking-water. URL: `https://www.who.int/news-room/fact-sheets/detail/drinking-water`.

Yang, S., Liang, M., Qin, Z., Qian, Y., Li, M., Cao, Y., 2021. A novel assessment considering spatial and temporal variations of water quality to identify pollution sources in urban rivers. Nature: Scientific Reports 8714.

Appendix A. Proofs

Appendix A.1. Proof of Proposition 1

Let $c_k \geq 0$ denote the amount of clean water available to region k (not to be confused with the cost c, which is irrelevant here), so $q_k = c_k/t_k$. The clean water comes from three sources: k's own clean inflow e_k, the water cleaned by $k-1$, namely x_{k-1}, and the clean water that entered but was not used for production in $k-1$. Specifically, $k-1$ used the fraction y_{k-1}/t_{k-1} of its available water; hence, the amount of clean water it did not use is

$$\left(1 - \frac{y_{k-1}}{t_{k-1}}\right) c_{k-1}.$$

Using $c_1 = e_1$ and defining $x_0 \equiv 0$,

$$\begin{aligned} c_k &= e_k + x_{k-1} + \left(1 - \frac{y_{k-1}}{t_{k-1}}\right) c_{k-1} \\ &= e_k + x_{k-1} + \left(1 - \frac{y_{k-1}}{t_{k-1}}\right)(e_{k-1} + x_{k-2}) + \left(1 - \frac{y_{k-1}}{t_{k-1}}\right)\left(1 - \frac{y_{k-2}}{t_{k-2}}\right) c_{k-2} \\ &\vdots \\ &= \sum_{i \leq k} (e_i + x_{i-1}) \prod_{i \leq j < k} \left(1 - \frac{y_j}{t_j}\right). \end{aligned}$$

Divide by t_k to obtain the desired expression for $q_k = c_k/t_k$. □

Appendix A.2. Proof of Theorem 1

Recall that $W(x) = \sum_i q_i(x) b_i - c \sum_i x_i$. Differentiate with respect to x_i and reformulate using the marginal benefits α_{ik}:

$$\frac{\partial W}{\partial x_i} = \sum_k \frac{\partial q_k}{\partial x_i} \cdot b_k - c = \sum_k \alpha_{ik} - c.$$

If the expression is positive, then W is increasing in x_i, so $x_i^* = (1-\delta) y_i$. If negative, then $x_i^* = 0$. □

Appendix A.3. Proof of Theorem 2

We will show that the abatement game is an activity optimization game with complementarity. Specifically, we show how our model can be relabeled to fit the one of Topkis (2011, Section 5.4).

Each player (region) has a single private activity (abatement) and there are no public activities. The set of feasible activity levels is X. The return function is $g(x,S) = W_S(x) - W_S(x^0) = \sum_{i\in S} x_i \sum_{k\in S} \alpha_{ik} - C(x,S)$. The cost function C is submodular (so $-C$ is supermodular) in (x,S) on $\{(x,S) : S \subseteq N, x \in X_S\}$ and upper semicontinous in x on X_S for each $S \subseteq N$. The first term,

$$\sum_{i\in S} x_i \sum_{k\in S} \alpha_{ik},$$

is supermodular (as well as submodular), so g is the sum of supermodular functions and itself supermodular. The characteristic function is here labeled v rather than f.

Convexity of the abatement game now follows from Theorem 5.4.1 in Topkis (2011). □

Appendix B. Supplementary material to simulation study

Here, we will describe a simple way to compute the characteristic function v for the abatement game. Recall from Proposition 2 that

$$v(S) = \sum_{i\in S} x_i^S \sum_{k\in S} \alpha_{ik}, \text{ where } x_i^S = \begin{cases} (1-\delta)y_i & \text{if } \sum_{k\in S} \alpha_{ik} \geq c \\ 0 & \text{otherwise.} \end{cases}$$

We proceed as follows. First, compute $A \in \mathbb{R}^{N\times N}$ with generic entry $a_{ik} = (1-\delta)y_i\alpha_{ik}$ for $i \neq k$ and $a_{ii} = -(1-\delta)y_i c$ on the diagonal. This key step needs only to be done once. Second, fix a coalition $S \subseteq N$ and copy A to $A^S \equiv A$. "Zero out" the rows and columns that pertain to non-members of S: for each $\{i,k\} \not\subseteq S$, set $a_{ik}^S = 0$. Hence, in the updated matrix, $a_{ik}^S \neq 0 \implies \{i,k\} \subseteq S$. Third, sum the entries of row $i \in S$: if negative (that is, if $\sum_{k\in S} \alpha_{ik} < c$), zero out also row i. The value $v(S)$ is the sum of the entries in the final matrix A^S. In this way, we can quickly compute all of v by repeatedly modifying the original matrix A.