River pollution abatement: Fully decentralized solutions implemented through smart contracts*

Jens Gudmundsson[a], Jens Leth Hougaard[a]

[a]*Department of Food and Resource Economics, University of Copenhagen, Denmark*

Abstract

In river systems, costly upstream pollution abatement creates downstream welfare gains. Without a central authority to enforce such abatement, the welfare gains can only be realized if downstream regions compensate for the upstream abatement. We develop a model that makes explicit the impact of pollution on water quality and production benefits. Furthermore, we propose the "Commit and Bargain" mechanism, a decentralized solution that yields welfare-maximizing abatement even in the absence of an informed trusted third party. We show how to implement the mechanism through a smart contract running on a blockchain and discuss the benefits of such a solution.

Keywords: River pollution, decentralized mechanisms, smart contracts, water quality
JEL: C7, D47, D62, Q52, Q25

1. Introduction

With population and industry growth contributing to wastewater production, at least 2 billion people now rely on polluted drinking sources (UN-Water, 2016; WHO, 2019). Difficult trade-offs are in play: pesticide and fertilizer use is a necessity for food security, but its effect on water quality is harmful to human, animal, and plant life (Lai, 2017; FAO, 2018). The conventional economic take on pollution abatement, to rely on regulatory authorities, taxation schemes, and intricate artificial markets (e.g. Duggan and Roberts, 2002; Montero, 2008; Ambec and Ehlers, 2016), may fall flat in practice: there simply may not be such an authority (e.g. for transboundary rivers across jurisdictions), and even if there is one, it may not have the information needed to implement the solutions (e.g. the social cost of pollution in Duggan and Roberts, 2002). Also, more generally, the effectiveness of a central authority has been questioned (Sigman, 2005; Cai et al., 2016). For this reason, we offer a new approach. We propose a decentralized solution for welfare-maximizing abatement that neither requires a trusted third party nor relies on hard-to-get information pertinent to the

*We would like to thank Erik Ansink, Dagim Belay, Frank Jensen, Goytom Abraha Kahsay, and Z. Emel Öztürk as well as seminar participants at University of Copenhagen and Vrije Universiteit Amsterdam 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)

problem.

Key in our proposal is to replace the conventional social planner with a smart contract running on a blockchain (we defer a description of this technology to Section 5). The immediate benefit of doing so is that we no longer need trusted third parties. The contract is transparent and self-executing: instead of trusting someone to implement the mechanism correctly, the contract is hardcoded to do so. Even more, we show that smart contracts allow regions, in a very strong sense, to commit to the end result (e.g. to go through with promised costly abatement). As the affected regions reach agreement without external actors interfering, there is potential to save both time and money. The mechanism that we suggest, denoted *Commit and Bargain*, is easy to implement as a smart contract, relies only on information provided by the regions themselves, and guarantees welfare-maximizing abatement in equilibrium.

The *Commit and Bargain* mechanism operates in two stages and is inspired by earlier work in Gudmundsson et al. (2019, 2023). Regions first simultaneously commit to costly abatement and thereafter bargain on sharing the induced welfare gains. Our main result (Theorem 2) shows that there is a stationary subgame-perfect equilibrium of the game induced by *Commit and Bargain* in which every region commits to welfare-maximizing abatement. The key stepping stone to this is to first pin down the equilibrium expected payoffs of the bargaining stage (Proposition 2). When we turn to the game as a whole, we exploit that a region's welfare-maximizing abatement is independent of the abatement of the others (Theorem 1), a finding driven by that each region's abatement has a constant marginal effect on social welfare (Proposition 1). In equilibrium, costly abatement is fully compensated for by those who benefit from it. On top of this, the welfare gain is shared equally among the regions.

Related literature. Our paper relates to several strands of literature. The model is inspired by the game-theoretic analysis of river sharing problems initiated by the seminal paper of Ambec and Sprumont (2002).[1] 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).[2] Close to our model formulation is Gengenbach et al. (2010), who consider exogeneous pollution levels and model agents' optimal choice of abatement under

[1] 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).

[2] 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. 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, 2021) argue that the model of Ni and Wang (2007) does not account for how pollution is transferred downstream and explicitly introduce the fact that pollution is transferred from upstream to downstream regions at a particular rate. 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 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. However, the analysis is with a completely different aim, as equilibrium abatement is then analyzed under various coalition structures.[3] In comparison, our model also considers optimal choice of abatement level under convex costs, but we further add the aspect of pollution transfers downstream and its effects on water quality (and thereby benefit from water use), taking into account potential inflows of clean water in each region. In this sense, our version of the model provides a bridge to the original river sharing model of Ambec and Sprumont (2002).

Several papers have studied implementation of welfare-maximizing pollution abatement. However, these solutions rely on the existence of an informed trusted third party, that is, on some degree of centralization. For instance, in the mechanism of Duggan and Roberts (2002), regions report their desired level of emissions as well as their "neighbor's demand". Based on these reports, each region faces a price equal to its marginal contribution to the social cost. Final payment further include penalties for misreporting the neighbor's demand. In the unique equilibrium, firms report welfare-maximizing emission levels. In this way, enforcing correct "market prices" requires a regulator with knowledge of the social cost function. Another example is Montero's (2008) auction (see also Kwerel, 1977), in which participating regions receive a "payback" to incentivize them to reveal important private information truthfully to a regulator, who again is informed of the social cost function. Translating our model into Montero's (2008) setting most naturally leads to a linear social cost function. In this case, Montero's (2008) regulator sets paybacks to zero and "the mechanism is effectively equivalent to an emissions tax set equal to the constant marginal damage" (Requate et al., 2019, p. 136), as in the "polluter-pays" regulation scheme proposed by Ambec and Ehlers (2016). Ambec and Ehlers (2016) show that their scheme implements welfare-maximizing emissions levels in equilibrium. The spirit of the polluter-pays scheme is that no region should be worse off than in the case that there is no pollution, so any harm caused by pollution should be compensated for by the polluter. Our approach is the opposite: any costly abatement should be compensated for by those who benefit from it. Our status quo is that, absent any agreements, each region will act in its own best interest (and not abate); the corresponding status quo in Ambec and Ehlers (2016) is rather that there is a central authority that can actively prevent the regions from polluting.

The *Commit and Bargain* mechanism builds on the ideas developed in Gudmundsson et al.

[3] 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.

(2019, 2023), adjusted to the context of pollution abatement. In the earlier work, the options available to downstream regions depended on the decisions made by upstream regions. The model here is somewhat simpler, and we can use the strong independence result of Theorem 1 to simplify the mechanism. The bargaining stage is adapted from the literature on non-cooperative bargaining with a random proposer, which dates back to work by Binmore et al. (1986), Binmore (1987), and Baron and Ferejohn (1989), among others. Equilibrium existence and uniqueness in such models is analyzed in Eraslan (2002) and Eraslan and McLennan (2013).

Lastly, a good presentation of smart contracts and their potential for economists can be found in Gans (2019). A more detailed discussion of how smart contracts can be used to implement various allocation mechanisms (in particular, auctions) can be found in Mamageishvili and Schlegel (2020). Yet, to our knowledge we are the first to provide and analyze a detailed contract design for bargaining mechanisms with respect to a concrete economic application. The specific implementation of *Commit and Bargain*, using deposits by the regions, is partly inspired by the analysis in Gerber and Wichardt (2009).

Outline. The paper is outlined as follows. In Section 2, we introduce the model. In Section 3, we pin down the welfare-maximizing abatement levels. In Section 4, we introduce the *Commit and Bargain* mechanism and analyze the non-cooperative game it induces. In Section 5, we discuss blockchain technology, its usefulness in the case of pollution abatement, and show how to implement *Commit and Bargain* through a smart contract. We conclude in Section 6. A numerical example, proofs, and technical details are postponed to the Appendix.

2. Model

We consider a stylized model of a river that covers both the benefits of water extraction as well as the ensuing drawbacks as water usage pollutes the water and diminishes its usefulness to downstream regions. Specifically, a river flows through n **regions** $N = \{1, \ldots, n\}$, where 1 is most upstream. Each region i has an **inflow** of $e_i \geq 0$ units of clean water, say due to precipitation; the total amount of water available to region j is $t_j = \sum_{i \leq j} e_j$. Along the river, each region i diverts $y_i \leq t_i$ units of water for **production**, which we take to be exogenous.[4] When used in production, the water gets polluted and is returned to the river. The pollution lowers the water quality as well as the benefits accrued from downstream production. We assume that water quality is measured as the fraction of clean water.[5] Specifically, each "unit of water" is either clean or polluted, and the **quality** $q_i \in [0, 1]$ of the t_i units of water at i's disposal is the fraction that is clean. Hence, $q_1 = 1$ by definition, but $q_2, \ldots, q_n$ may be below 1. Production creates quality-adjusted **benefit**

[4] This allows us to focus exclusively on pollution abatement. Model extensions are discussed in Section 6.

[5] 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.

$q_i b_i$ to region i, where $b_i \geq 0$ is the benefit from using clean water ($q_i = 1$) while fully polluted water ($q_i = 0$) is useless.[6]

Each region i has a strictly convex and differentiable **cost function** $C_i \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $C_i(0) = 0$ and derivative $C_i'(x_i)$ such that $C_i'(0) = 0$. That is, abating x_i units of pollution comes at cost $C_i(x_i)$ to region i. Cleanup is limited to the water used for production.[7] Thus, an **abatement profile** $x \in \mathbb{R}^n_{\geq 0}$ is such that $x \leq y$; let $\mathcal{X} \subseteq \mathbb{R}^n_{\geq 0}$ be the set of abatement profiles. Absent any agreements between the regions, we assume that there is no pollution abatement as abatement is costly but does not affect the abating region. The status quo is $x^0 = (0, \dots, 0) \in \mathcal{X}$. To implement welfare-increasing abatement profiles, downstream regions will provide monetary transfers to compensate upstream, abating regions (say by supporting investments in environmentally-friendly production facilities). Preferences are quasilinear: if region i receives a monetary transfer $\tau_i \in \mathbb{R}$ at the abatement profile $x \in \mathcal{X}$, i's utility is $q_i(x) b_i - C_i(x_i) + \tau_i$. As the monetary transfers cancel out (sum to zero), the **social welfare** $W \colon \mathcal{X} \to \mathbb{R}$ is the total quality-adjusted benefit net abatement cost, $W(x) = \sum_i (q_i(x) b_i - C_i(x_i))$. The welfare difference compared to the status quo is $\Delta(x) = W(x) - W(x^0)$. Throughout, all details of the problem are common knowledge among the regions. Next, in Section 3, we explore the welfare-maximizing abatement profiles. Figure 1 illustrates the model.

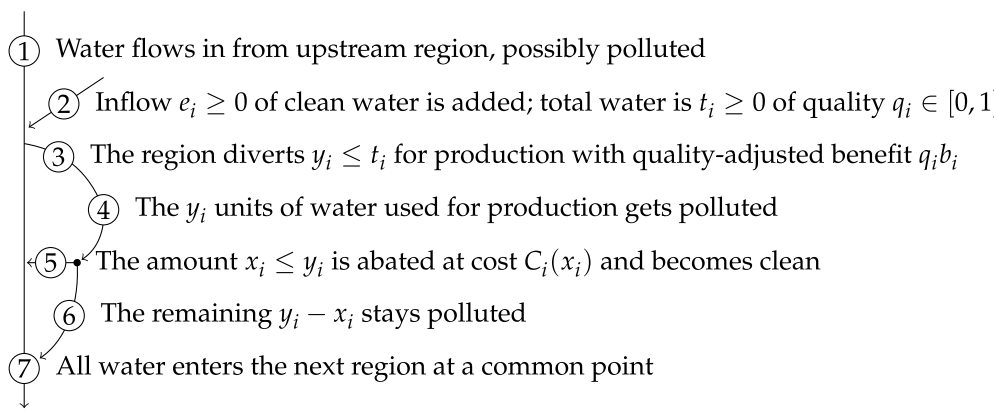


Figure 1: Water flow through a generic region i. Our key point of interest is the abatement decision (5).

[6]We imagine an underlying benefit function $B_i \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ that describes how benefits depend on production. However, as we take production y_i as given, it suffices to include $b_i = B_i(y_i)$ in the model. 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.

[7]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).

3. Efficient pollution abatement

The main result of this section is Theorem 1. It shows that there is a unique welfare-maximizing abatement profile x^*. In particular, x^* is efficient in the strong sense that, for every region i and every abatement profile x_{-i} for the other agents, x_i^* is optimal in terms of maximizing social welfare. That is, even conditional on suboptimal abatement by the others, x_i^* is optimal. All proofs are in the Appendix.

Theorem 1. *There is a unique $x^* \in \mathcal{X}$ such that, for each $i \in N$ and $x = (x_i, x_{-i}) \in \mathcal{X}$ with $x_i \neq x_i^*$,*

$$\Delta(x_i^*, x_{-i}) > \Delta(x_i, x_{-i}).$$

The main argument driving Theorem 1 is given in Proposition 1. Proposition 1 shows that region k's quality is linear in each region i's abatement. This extends immediately to the quality-adjusted benefit $q_k(x)b_k$ as well as to the entire "benefit side" of abatement, $\sum_{k>i} q_k(x)b_k$.[8] Hence, while we impose little on the cost function, the benefit side is far more structured.

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

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

Returning to Theorem 1, the key insight is that, if we can align the individual incentives with that of maximizing social welfare, then there are compelling reasons to expect that regions will abate efficiently. That is to say, if we can identify monetary transfers between the regions such that (i) each abating region is fully reimbursed for its abatement and, on top of this, (ii) each region obtains a share of the gain Δ, then each region has an interest in maximizing the gain Δ. By Theorem 1, abating x_i^* is then optimal regardless of what i expects the others to do. Next, in Section 4, we introduce a mechanism that accomplishes precisely this.

4. The "Commit and Bargain" mechanism

Inspired by our prior work (Gudmundsson et al., 2019, 2023), we introduce a two-stage mechanism that ensures efficient equilibrium abatement. Compared to the mechanisms explored previously, the strong incentives due to Theorem 1 allow us here to simplify the mechanism.[9] The

[8] This is in line with, for instance, how region i's pollution causes constant marginal damage to region k in Ambec and Ehlers (2016).

[9] In particular, the first stage of the mechanisms in Gudmundsson et al. (2019, 2023) features sequential moves to rule out potential undesirable equilibria. Moreover, the earlier work explores more general "recognition functions" that condition the second-stage proposer on the first-stage announcements.

Commit and Bargain mechanism first has each region commit to a particular investment level c_i corresponding to abatement x_i with $c_i = C_i(x_i)$; thereafter, the regions bargain (potentially indefinitely) on how to share the induced welfare gain $\Delta(x)$.

Definition 1 (*Commit and Bargain*)**.** The mechanism operates in the following steps:

1. Regions simultaneously commit to abatement costs, each region i choosing $c_i \geq 0$.
2. With equal probabilities, one region i is randomly selected to be the proposer.
3. The proposer i suggests monetary transfers $\tau \in \mathcal{T} = \{\tau \in \mathbb{R}^n \mid \sum_k \tau_k = 0\}$.
4. Regions are ordered at random and then sequentially decide to accept or reject the proposal:
    (a) If all accept, then the proposal is implemented and the game ends.
    (b) If some region rejects, then we return to Step 2 and select a (possibly new) proposer. ◦

In this way, the bargaining stage is an infinitely repeated game in which (i) the status quo prevails until agreement is reached, and (ii) once agreement is reached, the agreed levels are fixed (for the remainder of the infinitely repeated game). Payoffs are discounted with common discount factor $\delta \in (0,1)$. Thus, region i derives an infinite stream of utilities $(u_i^t)_{t=0}^{\infty}$ with average discounted payoff

$$(1-\delta)\sum_{t=0}^{\infty} \delta^t u_i^t.$$

If $\tau_i > 0$, then region i is to receive funds; if $\tau_i < 0$, then i has to pay (to compensate another region for abating). If commitments c (corresponding to abatement profile x) and transfers τ are agreed upon in round $r \in \{0,1,\dots\}$, then

$$u_i^t = \begin{cases} q_i(x^0)b_i & \text{for } t < r \\ q_i(x)b_i - c_i + \tau_i & \text{for } t \geq r. \end{cases}$$

Equivalently, for $t \geq r$, $u_i^t = q_i(x^0)b_i + q_i(x)b_i - c_i + \tau_i - q_i(x^0)b_i$. In this way, the average discounted payoff simplifies to

$$\begin{aligned} &q_i(x^0)b_i + (1-\delta)\sum_{t=r}^{\infty} \delta^t (q_i(x)b_i - c_i + \tau_i - q_i(x^0)b_i) \\ &= q_i(x^0)b_i + \delta^r (q_i(x)b_i - c_i + \tau_i - q_i(x^0)b_i). \end{aligned}$$

If agreement never is reached (corresponding to the limit $r \to \infty$), this simplifies further to $q_i(x^0)b_i$.

We restrict to subgame-perfect equilibria. Within the bargaining stage, we restrict further to stationary equilibria. As in Eraslan (2002), we say that a strategy profile is stationary if it does not depend on the current round or the history of play leading up to the current round. Hence, for given commitments c, region i always proposes the same transfer vector $\tau^i \in \mathcal{T}$ and always accepts the same set of proposals $A^i_j \subseteq \mathbb{R}^n$ made by region j.

The main result of this section is Theorem 2. It shows that *Commit and Bargain* accomplishes to align incentives and that each region commits to efficient abatement in equilibrium.[10]

Theorem 2. *There is a stationary subgame-perfect equilibrium of the game induced by Commit and Bargain in which every $i \in N$ commits $c_i = C_i(x^*_i)$.*

The key stepping stone to Theorem 2 is to pin down the expected equilibrium payoffs in the bargaining stage. In the most relevant case that the commitment stage results in x with $\Delta(x) \geq 0$, the bargaining stage has a unique stationary subgame-perfect equilibrium in which the gain $\Delta(x)$ is shared equally among all regions (on top of their status-quo payoff $q_i(x^0)b_i$). The equal split is due to the symmetric treatment of the regions (e.g. the uniform draw of proposer). In this equilibrium, agreement is reached immediately in the first round of bargaining. Proposition 2 also covers the less interesting case $\Delta(x) < 0$ in which agreement never is reached.

Proposition 2. *In all stationary subgame-perfect equilibria of the bargaining stage, every $i \in N$ has expected payoff*

$$\mathbb{E}\pi_i = q_i(x^0)b_i + \max\{\Delta(x), 0\}/n.$$

To implement *Commit and Bargain* in practice, several questions still need to be resolved. Most importantly, we need to ensure that regions follow through on their commitment to efficient abatement. Moreover, in terms of timing, abating regions may wish to be reimbursed upfront; on the other hand, reimbursing regions may prefer to see the quality increase (due to pollution abatement) before making their promised monetary transfers. The question is how this can be resolved—in particular in the context of international rivers across legislations and without trusted third parties. Next, in Section 5, we propose a solution to this that explores the new blockchain technology and smart contracts.

5. Practical implementation through smart contracts

In this section, Subsection 5.1 first provides a simplified introduction to the relevant technology consisting of blockchains and smart contracts.[11] Thereafter, in Subsection 5.2, we sketch how

[10]We contend that the equilibrium identified in Theorem 2 is the interesting one. Technically, due to the simultaneous-move commitment stage, there can also be "bad" equilibria in which two (or more) regions "misbehave", each because the other one does. Define $\bar{x} \in \mathbb{R}^n_{\geq 0}$ with $\Delta(\bar{x}_i, x^*_{-i}) = 0$. That is, i "overabates" to the point that it is impossible to improve on the status quo. One can easily create instances where $\bar{x}_i \leq y_i$, so $\bar{x}_i$ is feasible. Every $x \in X$ with $x_i \geq \bar{x}_i$ for at least two regions i can be obtained in equilibrium.

[11]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`.

to connect smart contracts to pollution abatement. Finally, in Subsection 5.3, we turn to the implementation of *Commit and Bargain* and in particular discuss some of the difficulties that arise when turning the theoretical mechanism into a working contract.

5.1. Background

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).[12] 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. The variables are initialized when the contract is "deployed" on the blockchain; the contract then obtains a unique address and its code is fixed. In this way, 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.[13]

5.2. Smart contracts and pollution abatement

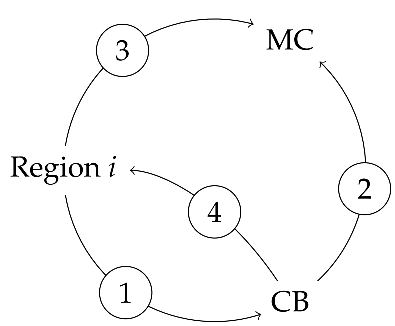
We envision a scenario in which pollution abatement is directly tied to and controlled by some software (say relying on some sensor data to optimize a filtration process). Similar to how Tesla's

[12] 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.

[13] There are costs associated to interacting with a smart contract as well. For the Ethereum blockchain, each instruction has a "gas" cost. In our implementation in Subsection 5.3, each region essentially transacts three times with the contract to alter different variables in its storage. On average, each such transaction uses around 100 000 gas. Each unit of gas costs a number of gwei (a basic unit of Ethereum's cryptocurrency) to execute. At the time of writing (January 2023), the gas price was roughly 25 gwei. In this way, the three transactions amount to 75 million gwei, or 0.0075 ether. The exchange rate to USD at the time was 1 300 USD for one ether. Hence, executing the contract would cost each region 10 USD. The most expensive operation is the one-time deployment of the contract on the blockchain, which uses roughly 1.2 million gas and thus costs around 40 USD.

Autopilot software comes installed with the purchase of a car but has to be activated through a monthly subscription, we imagine that the abatement equipment is supplied to the regions but inoperable until activated. It is straightforward for the equipment manufacturer to use smart contract for its license management; that is, purchasing and activating the software license can be done on-chain (e.g., once the software is activated, it sends a transaction to register this in the manufacturer's contract).

The benefit of this approach is that "our" contract, implementing *Commit and Bargain*, can interact with the manufacturer's contract. In this way, when region i commits to investing c_i in our contract, they will also be required to deposit this amount; once agreement is reached, our contract uses the deposited funds to purchase the corresponding licenses from the manufacturer's contract. This automated, self-executing feature makes the commitments credible to a degree that would be very difficult to attain through a conventional trusted third party.[14] The same applies to the proposal evaluation stage: for a region j to accept a proposal τ in which it is supposed to pay some funds, $\tau_j < 0$, the region has to deposit enough to cover this. That is, once a proposal is unanimously accepted, it can be implemented immediately. A region set to receive a monetary transfer (to compensate for abatement) will only do so once it has activated its licenses; again, this is achieved by interaction between the two contracts. Assuming that licenses cannot be repurchased, the incentives are straightforward: if the license is not activated, the region gets nothing; if activated, it gains τ_i. Figure 2 sketches the timeline.



1. Regions interact through the *Commit and Bargain* (CB) contract to make commitments, proposals, deposits, etc.
2. Once agreement is reached, the CB contract uses the deposits to purchase licenses from the manufacturer (MC)
3. Once the licenses are activated by the region, the software registers this within the MC contract
4. This finally triggers the CB contract to pay out the monetary transfers agreed on to the regions

Figure 2: Timeline of how the regions interact with the *Commit and Bargain* contract (CB) and the manufacturer's contract (MC).

As an added safeguard, we can employ tools of the "Internet-of-Things": the regions can position multiple sensors along the river, which report data directly to the smart contract. The transfers are then paid out only once the reported water quality is adequate. To give an example, Singh et al.

[14] Gans (2019) makes a related observation regarding this commitment: "... because so much can be included in smart contracts running off the blockchain, there is potential to innovate and deploy economic mechanisms that could create incentives for humans to perform and report on obligations in a truthful manner. In other words, the commitment engendered by the blockchain could substitute for a lack of trust in the real world."

(2020) suggest an IoT sensor-based blockchain framework for temperature monitoring.

5.3. *Smart contract implementation of Commit and Bargain*

A key component of our implementation of *Commit and Bargain* are so called *cryptographic hash functions*. Such a (deterministic) function H maps arbitrary inputs to outputs of a fixed size (e.g. a 256-bit hash function has $2^{256} \approx 10^{77}$ different outputs). In particular, the inputs are mapped "evenly" over the output range with every output generated with roughly the same probability. Moreover, small input variations lead to "unpredictable" output changes: that is, a sequence of outputs $H(1), H(2), \ldots$ appears as if drawn uniformly from $\{0, 1, \ldots, 2^{256} - 1\}$. While it is easy to compute $H(x)$, it is difficult to reverse-engineer an input x from the encrypted output $H(x)$ or to find a different input y with the same output, $H(x) = H(y)$.

To make *Commit and Bargain* operational in practice, some extra steps need to be taken.[15] In the theoretical model, the set of regions is fixed from the outset; now, we include a "registration" phase to link regions to their addresses on the blockchain. Furthermore, due to the sequential nature of the blockchain, the simultaneous-move commitment stage is divided into two steps.

Definition 2 (Smart contract implementation of *Commit and Bargain*)**.** The hash function H and the parameter $M \geq 0$ is fixed from the outset. The mechanism operates in the following steps:

- 0. Regions get linked to their respective addresses through a "registration" function.
- 1a. Each region i submits $H(c_i')$ to the "encrypted commitment" function, where $c_i' \bmod M = c_i$.
- 1b. Each region i submits c_i' to the "actual commitment" function and deposits at least c_i.
- 2. Regions are (uniformly) randomly ordered; the first in the order becomes the proposer.
- 3. The proposer i submits $\tau \in \mathcal{T}$ to the "proposal" function. If i is to pay some amount, $\tau_i < 0$, then i also has to deposit at least $-\tau_i$.
- 4. In order, regions sequentially decide to accept or reject the proposal by submitting 1 or 0 to the "evaluate proposal" function. If region j is to pay some amount, $\tau_j < 0$, then j also has to deposit at least $-\tau_j$.
  - (a) If all accept, then the proposal is implemented and the game ends.
  - (b) If some region rejects, then we return to Step 2 and select a possibly new proposer. $\circ$

When region i wants to commit to $c_i \geq 0$, they represent it by some $c_i' \geq 0$ such that $c_i' \bmod M = c_i$ and submit $H(c_i')$ to the "encrypted commitment" function. This is to ensure that no other region can infer c_i from the encrypted commitment. If regions instead were to submit $H(c_i)$ directly and there were reasons to expect c_i to be bounded by some $K \geq 0$, then it would suffice to compute $H(x)$ for $x \leq K$ to learn c_i. Unless K is very large, this can be done quickly. In contrast,

[15] We refer to `https://github.com/jensgudmundsson/AbatementContract` for a complete prototype of the smart contract for the Ethereum blockchain.

our approach gives many more ways of "representing" c_i through c_i', so "checking all of them" is no longer viable.[16] In Step 1a, even though region j may have observed $H(c_i')$ when choosing its encrypted commitment, it cannot infer c_i and condition its choice on it; in Step 1b, even though region j may have observed c_i when revealing its actual commitment, it will be unable to find a different input than c_j' to match its encrypted commitment. Hence, even though regions inevitably act sequentially, the information structure is equivalent to that of a simultaneous-move game. As M is publicly known and the profile c' is observable, the profile c of commitments is known to all regions at the conclusion of Step 1b.

A final complication is to randomly order the regions as there is no "random number generator" in the smart contract.[17] This is resolved by again relying on the hash function, and specifically that it maps inputs "evenly" over the output range with every hash value generated with roughly the same probability. For instance, $H(\text{blockNumber}) \bmod 2$ would produce 0 for half the blocks and 1 otherwise while $H(\text{blockNumber}) \bmod n$ corresponds to a uniform distribution over $\{0, \ldots, n-1\}$. Once this source of randomness is in place, the Fischer-Yates shuffle is used to generate the random order (see e.g. Knuth, 1997).

We refer to Appendix A for a short numerical example that illustrates the model as well as our smart contract implementation of the *Commit and Bargain* mechanism.

6. Concluding remarks

We have studied fair allocation of the welfare gains of efficient river pollution abatement. For this purpose, we examined a 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. We proposed a decentralized implementation of this solution through the *Commit and Bargain* mechanism, inspired by Gudmundsson et al. (2019, 2023). A fully decentralized solution is especially useful in cases where there is no natural central authority, for instance for transboundary rivers across jurisdictions. In addition, we showed how *Commit and Bargain* can be implemented in practice through a smart contract, allowing the regions to credibly commit to pollution abatement and to negotiate without external involvement. By not requiring a trusted third party and by being able to automate payments, transaction costs are reduced.

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

[16]With say $M = 2^{128} \approx 10^{38}$, region i can commit to any $c_i \in \{0, \ldots, 2^{128} - 1\}$ in $2^{256}/2^{128} = 2^{128}$ different ways.

[17]Mamageishvili and Schlegel (2020) propose a slightly different approach to this that is in the same spirit.

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. These changes only amount to more cumbersome notation and more involved equilibrium expressions.

There are also more ambitious alternative specifications that we leave for future research. First, our study separates emissions into pollution and abatement with corresponding benefits and costs, but focuses only on the abatement decision. A potential extension is to let the regions decide on both the production and abatement levels (compare footnotes 4 and 6) or to integrate it into one "end emission" level. We conjecture that the *Commit and Bargain* mechanism would have to be modified to deal with this setting. Specifically, for upstream region i to know how much to abate, it must be able to compute the welfare-effect of its abatement x_i, which depends on downstream production y_j. At the same time, the ideal level of production y_j would depend on the water quality q_j, which depends on upstream abatement x_i. We leave this topic for future research.

Second, in contrast to the approach of for instance Kwerel (1977), Duggan and Roberts (2002), and Montero (2008), we examine a model of linearly ordered regions. A potential extension is the "unordered" case in which there is a "social damage" function that depends on aggregate pollution. The key to ensure the strong incentives in the commitment stage (Theorem 2) is that a region's marginal damage is constant (Proposition 1 and Theorem 1). As long as this remains, the *Commit and Bargain* mechanism implements the efficient allocation in equilibrium.

Third, 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 contrast to the "polluter pays principle" (e.g. Ambec and Ehlers, 2016), 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 $(0, \dots, 0)$. In this way, x_i^0 would be i's mandatory abatement: in equilibrium, i presumably would cover this part, while any abatement on top of x_i^0 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*.[18] Another 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 $\bar{q}$.

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

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., Ehlers, L., 2016. Regulation via the Polluter-pays Principle. The Economic Journal 126, 884–906.

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.

Baron, D.P., Ferejohn, J.A., 1989. Bargaining in legislatures. The American Political Science Review 83, 1181–1206.

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

Binmore, K., 1987. Perfect equilibria in bargaining models, in: Binmore, K., Dasgupta, P. (Eds.), The Economics of Bargaining. Blackwell, Oxford. chapter 5, pp. 77–105.

Binmore, K., Rubinstein, A., Wolinsky, A., 1986. The Nash bargaining solution in economic modelling. The RAND Journal of Economics 17, 176–188.

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.

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.

Duggan, J., Roberts, J., 2002. Implementing the Efficient Allocation of Pollution. American Economic Review 92, 1070–1078.

Eraslan, H., 2002. Uniqueness of Stationary Equilibrium Payoffs in the Baron-Ferejohn Model. Journal of Economic Theory 103, 11–30.

Eraslan, H., McLennan, A., 2013. Uniqueness of stationary equilibrium payoffs in coalitional bargaining. Journal of Economic Theory 148, 2195–2222.

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.

Gans, J.S., 2019. The Fine Print in Smart Contracts. Working Paper 25443. NBER.

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

Gerber, A., Wichardt, P., 2009. Providing public goods in the absence of strong institutions. Journal of Public Economics 93, 429–439.

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

Gudmundsson, J., Hougaard, J.L., Platz, T.T., 2023. Decentralized Task Coordination. European Journal of Operational Research 304, 851–864.

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

Knuth, D.E., 1997. The Art of Computer Programming, Volume 2: Seminumerical Algorithms. 3 ed., Addison-Wesley Longman Publishing Co., Inc.

Kwerel, E., 1977. To Tell the Truth: Imperfect Information and Optimal Pollution Control. The Review of Economic Studies 44, 595–601.

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.

Mamageishvili, A., Schlegel, J.C., 2020. Mechanism Design and Blockchains. `https://arxiv.org/abs/2005.02390`.

Montero, J.P., 2008. A Simple Auction Mechanism for the Optimal Allocation of the Commons. American Economic Review 98, 496–518.

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.

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.

Requate, T., Camacho-Cuena, E., Siang, C.K., Waichman, I., 2019. Tell the truth or not? The Montero mechanism for emissions control at work. Journal of Environmental Economics and Management 95, 133–152.

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.

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.

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.

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`.

Appendix A. Numerical example

In this section, we consider an extended numerical example to illustrate the model, the *Commit and Bargain* mechanism, and the smart contract. The regions $N = \{1, 2\}$ have inflows $e = (1, 1)$, water totals $t = (1, 2)$, production $y = (1, 1)$, benefits $b = (64, 64)$, and cost functions $C_i(x_i) = 64x_i^2$. At the status quo $x^0 = (0, 0)$, we have $q(x^0) = (1, 1/2)$ and utilities are 64 and 32, respectively. For allocation $x \in \mathcal{X} = [0, 1]^2$, we have $W(x) = 64(1 + (1 + x_1)/2 - (x_1^2 + x_2^2))$. Hence, the welfare-maximizing abatement is $x^* = (1/4, 0)$ with costs $c = (4, 0)$. This induces quality $q(x^*) = (1, 5/8)$ and utilities 60 and 40, respectively. The welfare gain is $\Delta(x^*) = 4$. The expected equilibrium payoffs are 66 and 34, respectively. With discount factor $\delta \in (0, 1)$, the equilibrium acceptance thresholds are $a_1 = 4 + 2\delta$ and $a_2 = 2\delta - 8$. Region i's equilibrium monetary transfer is $\tau^i = (\tau_i^i, \tau_j^i) = (-a_j, a_j)$. For instance, with $\delta = 1/2$, $\tau_1^1 = 7$ and $\tau_1^2 = 5$. The larger δ, the less the "proposer's rent" and the closer the proposals get to $(6, -6)$.

For the smart contract, set the parameter $M = 2^{128}$. To commit $c = (4, 0)$, the region can choose, for instance, $c' = (4, 2^{196})$, as this gives c_i' mod $M = c_i$. Hence, in Step 1a, they submit $H(c_i')$ to the "encrypted commitment" function, where $H(4) = 8a35a\ldots$ and $H(2^{196}) = 4e350\ldots$, respectively. In Step 1b, they submit c_i' to the "actual commitment" function together with deposits of at least c_i. The minimum deposits are 4 and 0; anything in excess of this is refunded at the end of the mechanism. In the bargaining stage, Step 2, say the common discount factor is $\delta = 1/2$. Then, if region 1 is selected as proposer, the region submits $(7, -7)$ to the "proposal" function. For region 2 to accept the proposal, it must deposit at least 7 as well. If instead region 2 is selected as proposer, then it submits $(5, -5)$ to the "proposal" function together with a deposit of at least 5. As region 1 is set to receive funds here, it can accept the proposal without making a deposit. Once agreement is reached, the 4 units deposited by region 1 in Step 1b are used to purchase abatement licenses as discussed in Subsection 5.2. Once the licenses are activated, the 5 or 7 units deposited in Step 2 are paid out to region 1.

Appendix B. Proofs

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

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

Proof. Let $\zeta_k \geq 0$ denote the amount of clean water available to region k, so $q_k = \zeta_k / t_k$. 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

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

Using $\zeta_1 = e_1$ and defining $x_0 \equiv 0$,

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

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

Theorem 1. *There is a unique $x^* \in \mathcal{X}$ such that, for each $i \in N$ and $x = (x_i, x_{-i}) \in \mathcal{X}$ with $x_i \neq x_i^*$,*

$$\Delta(x_i^*, x_{-i}) > \Delta(x_i, x_{-i}).$$

Proof. Differentiate $W(x) = \sum_k (q_k(x) b_k - C_k(x_k))$ with respect to x_i and use Proposition 1:

$$\frac{\partial W(x)}{\partial x_i} = \sum_k \frac{\partial q_k(x)}{\partial x_i} \cdot b_k - \frac{\partial C_i(x_i)}{\partial x_i} = \sum_{k>i} \prod_{i<j<k} \left(1 - \frac{y_j}{t_j}\right) \cdot \frac{b_k}{t_k} - \frac{\partial C_i(x_i)}{\partial x_i}.$$

Here, the first term is non-negative and independent of x. The second term, $C_i'(x_i)$, is continuous and increasing in x_i and zero at $x_i = 0$. Define $\tilde{x}_i \geq 0$ such that $W'_{x_i}(\tilde{x}_i, x_{-i}) = 0$. Then $x_i^* = \min\{\tilde{x}_i, y_i\}$. □

Proposition 2. *In all stationary subgame-perfect equilibria of the bargaining stage, every $i \in N$ has expected payoff*

$$\mathbb{E}\pi_i = q_i(x^0) b_i + \max\{\Delta(x), 0\}/n.$$

Proof. Take as given commitment c to abatement profile x from the first stage, so $c_i = C_i(x_i)$ for each region i. Consider a stationary subgame-perfect equilibrium in the bargaining stage.

As i is guaranteed payoff $q_i(x^0) b_i$ by rejecting all proposals and only making proposals $\tau \in \mathcal{T}$

with sufficiently high τ_i, $\mathbb{E}\pi_i \geq q_i(x^0)b_i$. By stationarity, the continuation payoffs are constant: in any round, if the proposal is rejected, then i's continuation payoff is $v_i \equiv (1-\delta)q_i(x^0)b_i + \delta\mathbb{E}\pi_i$. By subgame perfection, i accepts precisely the proposals that yield payoff at least v_i. That is, $\tau \in \mathcal{T}$ is accepted whenever, for every region i,

$$q_i(x)b_i - c_i + \tau_i \geq v_i \geq q_i(x^0)b_i.$$

Summing over all regions and using that $\sum_i \tau_i = 0$,

$$\textstyle\sum_i(q_i(x)b_i - c_i + \tau_i) \geq \sum_i q_i(x^0)b_i \iff \sum_i(q_i(x)b_i - c_i) - \sum_i q_i(x^0)b_i \geq 0 \iff \Delta(x) \geq 0.$$

Hence, if $\Delta(x) < 0$, then agreement will not be reached and $\mathbb{E}\pi_i = q_i(x^0)b_i$.

Consider instead $\Delta(x) > 0$. By contradiction, suppose that agreement is not reached, so $\mathbb{E}\pi_i = v_i = q_i(x^0)b_i$. Then i is better off proposing $\tau \in \mathcal{T}$ such that $q_i(x)b_i - c_i + \tau_i = v_i + \Delta(x)/n$, which is accepted by all regions and yields payoff $v_j + \Delta(x)/n > v_j$. By contradiction, suppose that agreement is not reached in the first round. Then the region j making the first-round rejected proposal is better of proposing the proposal that eventually gets accepted.

Define $a_j \equiv v_j - q_j(x)b_j + c_j$, so j accepts $\tau \in \mathcal{T}$ whenever $\tau_j \geq a_j$. In equilibrium, region i awards j precisely $\tau_j^i = a_j$. By balance, $\tau_j^j = -\sum_{i \neq j} \tau_i^j = -\sum_{i \neq j} a_i = a_j - \sum_i a_i$. Therefore,

$$\mathbb{E}\pi_j = q_j(x)b_j - c_j + \frac{1}{n}\textstyle\sum_i \tau_j^i = v_j - a_j + a_j - \frac{1}{n}\sum_i a_i = v_j - \frac{1}{n}\sum_i a_i.$$

Summing over all regions,

$$\textstyle\sum_j \mathbb{E}\pi_j = \sum_j(v_j - a_j) = \sum_j(q_j(x)b_j - c_j) = W(x).$$

Therefore,

$$\textstyle\sum_j a_j = (1-\delta)\underbrace{\textstyle\sum_j q_j(x^0)b_j}_{W(x^0)} + \delta\underbrace{\textstyle\sum_j \mathbb{E}\pi_j}_{W(x)} - \underbrace{\textstyle\sum_j \left(q_j(x)b_j - c_j\right)}_{W(x)} = -(1-\delta)\Delta(x).$$

Hence,

$$v_j = (1-\delta)q_j(x^0)b_j + \delta\mathbb{E}\pi_j = \mathbb{E}\pi_j + \frac{1}{n}\textstyle\sum_i a_i \iff \mathbb{E}\pi_j = q_j(x^0)b_j + \frac{1}{n}\Delta(x). \qquad \square$$

Theorem 2. *There is a stationary subgame-perfect equilibrium of the game induced by Commit and Bargain in which every $i \in N$ commits $c_i = C_i(x_i^*)$.*

Proof. In the bargaining stage, by Proposition 2, region i expects

$$\mathbb{E}\pi_i = q_i(x^0)b_i + \max\{\Delta(x), 0\}/n.$$

By Theorem 1, x_i^* always maximizes $\Delta(x)$. That is, x_i^* is a best response to all choices x_{-i} by the other regions. This holds in particular against x_{-i}^*, so x^* is an equilibrium. As $\Delta(x^*) > 0$, the equilibrium is strict: x_i^* is the unique best response to x_{-i}^*. □