River Sharing: Implementation of Efficient Water Consumption*

Jens Gudmundsson† Jens Leth Hougaard‡ Chiu Yu Ko§

February 8, 2018

Abstract

We examine the river sharing problem with a novel focus on implementing efficient outcomes. We introduce a class of two-stage games in which agents first announce desired levels of consumption and then bargain over monetary compensations. An impartial outside observer with partial information connects the two stages by conditioning agents' recognition probabilities in the second stage on the allocation realized in the first stage. We identify ways of setting recognition probabilities that give agents incentives to choose the efficient allocation of water all the while ensuring that the surplus obtained from efficient water usage ends up being shared in a fair way. In particular, we take an axiomatic approach to determine three new and implementable solutions.

JEL Classification: C7, D4, Q34
Keywords: River sharing · Implementation · Bargaining · Consistency

1 Introduction

We revisit the river sharing problem as introduced by Ambec and Sprumont (2002). A river flows through several countries. The flow of the river causes consumption externalities: the more water that is diverted by upstream countries, the less is available for downstream countries to consume. At the same time, a sovereign country may feel entitled to use the resources within its borders as it sees fit. In this way, the absence of well-defined property rights to the water makes it difficult for the involved parties to coordinate on a socially desirable water allocation. As a remedy, we assume that an outside observer (a "planner") assists in recommending a solution that is agreeable to all parties. However, whereas the countries ("agents") have complete information about the river, the planner cannot discern the level of utility that the agents derive from consuming water.[1] This is resolved by letting the agents interact through a game form. When designed and played optimally, the game reveals the efficient allocation of water and matches it with bilateral monetary compensations in such a way that all agents improve upon the status quo (when there are no water

*We thank Frank Jensen and participants at the Workshop on Network Economics (Seville, 2017) for valuable comments.

†Department of Economics, University of Copenhagen, Denmark. Email: jens.gudmundsson@econ.ku.dk

‡Department of Food and Resource Economics, University of Copenhagen, Denmark. Email: jlh@ifro.ku.dk

§Department of Economics, National University Singapore, Singapore. Email: kochiuyu@nus.edu.sg

[1]Alternatively, the planner may represent a protocol or a convention that has been established over time. The lack of complete information may then capture that the protocol only depends on publicly verifiable information.

transfers between the agents). We investigate how such a game form can be designed in order to always implement efficient outcomes. We further pinpoint implementable game forms that divide the gains obtained from efficient water usage in normatively compelling ways.

More specifically, a problem is defined as follows. Each agent has an inflow to the river and a benefit function that captures the agent's utility of water consumption. Benefit functions are strictly increasing, so for an agent to agree to a downstream transfer of water (that is, to abstain from consuming some water), the downstream agent has to compensate the upstream agent. A practical example of such a compensation is hydropower investments, which have the added benefit of not diverting any water from the river (in contrast to irrigation of fields) and therefore not harming downstream agents. Throughout, we exclude the more costly option of transferring water upstream. When done correctly, the downstream transfers increase aggregate benefits compared to the situation without transfers. The monetary compensations govern how this "surplus" is shared. Preferences are quasi-linear in money, so an allocation is (Pareto) efficient whenever it maximizes aggregate benefits. Furthermore, benefit functions are strictly concave, so there is a unique allocation that maximizes aggregate benefits. This is the socially optimal consumption plan that the planner intends to implement.

The planner, unable to observe the benefit functions, must consult the agents in order to determine the efficient allocation. This is modeled as a two-stage game. In the first stage, the agents sequentially submit their desired consumption level. The most upstream agent (the "source") chooses first and we then proceed downstream. Ideally, the agents, who have complete information, choose to consume according to the efficient allocation in equilibrium. However, depending on how the game proceeds, they may have incentives to do otherwise. In the second stage, agents bargain over the monetary compensations. In each round of negotiation, one agent is selected at random to propose how the agents should compensate one another. If there is unanimous agreement on the proposal, then it is implemented. On the other hand, if some agent rejects the proposal, we proceed to a new round with a possibly different proposer. The longer the negotiations take, the more the agents discount the outcome implemented if, and when, agreement is reached. (However, we show in Theorem 1 that there is no delay in equilibrium.) The planner ties the two stages together by making the probability distribution over who gets to propose in Stage 2 a function of the allocation realized in Stage 1. Formally, an agent is "recognized" as the proposer with a certain probability. We examine how the recognition probabilities can be set in order to incentivize agents to choose the efficient allocation (to maximize surplus) in Stage 1 all the while resulting in a "fair" division of the surplus is Stage 2. To do so, we analyze the subgame-perfect (Nash) equilibrium of the two-stage game through backward induction.

Our first result, Theorem 1, shows that the bargaining game in Stage 2 generally has a unique subgame-perfect equilibrium in stationary strategies.In equilibrium, the surplus is shared in proportion to the recognition probabilities. This shows a potential trade-off: an agent prefers to consume an inefficient amount if this increases the agent's recognition probability by relatively more than it decreases the surplus. To ensure that the game results in an efficient outcome, these trade-offs must be avoided when designing the recognition function. Theorem 2 shows that, for the game to always implement efficient outcomes, an agent's recognition probability can only depend on the consumption choices made by those upstream from the agent. Equivalently, an agent's consumption choice can only affect the recognition probabilities of those downstream from the agent. Theorem 2 sets the overall structure of implementable recognition functions, but it still leaves open many ways to share the surplus. Among these, we wish to identify solutions that, in some regard, award the surplus to the agents most responsible for generating it. The most direct way of doing so is to determine how much benefit each agent derives at the efficient allocation and then share the surplus accordingly, but this is not feasible without knowledge of the benefit functions. An alternative is to emphasize that the surplus always is created by some agent transferring water for another agent to use more efficiently. However, Theorem 2 shows that conditioning the recognition probability on how much the agent consumes or abstains from consuming is incompatible with efficient implementation. To further emphasize this point, our third result turns the argument around. If no water is transferred to or from an agent, then the agent certainly does *not* contribute to the surplus. Theorem 3 shows that there is no implementable recognition function that always excludes these "dummy" agents from getting a share of the surplus. To narrow down the number of possible solutions, we turn to less demanding conditions that still reflect normatively compelling properties in the present context. Through this axiomatic approach, we give normative foundation to three new solutions to the river sharing problem.

The first axiom pertains to two-agent problems. When there are only two agents, any surplus is created by the upstream agent transferring water for the downstream agent to consume more efficiently. As this is common to all two-agent problems, we may as well recommend the same *solution* to all such problems. This is captured by *two-agent independence*. The second axiom, *consistency*, concerns scalability: given that we know how to solve two-agent problems in a desirable way, consistent extension of such a solution allows scaling to any number of agents. Theorem 4 shows that *two-agent independence* and *consistency* together characterize recognition functions for which the probabilities form geometric series. Next, we note that while we cannot award agents for consuming or transferring more water, we can award them for having more water *available*. This is a reasonable proxy for what we want to capture: if an agent has more water available, then the agent has to consume and/or transfer more water. There is no conflict with efficient implementation as the water available only depends on the inflows and the consumption choices of the upstream agents. We obtain this solution by first settling the recognition probability for the source as captured by the axiom *source-solution*. We then, again, want a consistent extension of this principle (now slightly different) in order to determine the recognition probabilities for the other agents. Theorem 5 shows that *source-solution* and *source-consistency* characterize the recognition function that sets probabilities in proportion to the water available to the agents. For our final result, we turn to an invariance property. *Source merging neutrality* implies that a country's bargaining power is independent of whether we treat the country as one agent or divide it into separate regions which we treat as separate agents. Theorem 6 shows that *source-solution* and *source merging neutrality* together characterize the simple recognition function that sets probabilities in proportion to inflows.

The solution introduced by Ambec and Sprumont (2002), *Downstream incremental* (DI), is intrinsically different from those that we study. Most importantly, DI generically recommends different solutions to two problems that differ only in benefit functions. In contrast, each solution on how to share the welfare surplus that we consider is set without knowledge of the benefit functions and thus selects the same outcome in such pairs of problems. Moreover, DI never awards a "dummy" agent a share of the surplus. There are such solutions in our class as well, but Theorem 3 shows that none of them implements efficient outcomes, so they fall outside the set of solutions that we primarily focus on. Furthermore, whereas DI never awards the source a share of the surplus, the three solutions that we characterize generally do. Finally, we use simulations to show that the solutions present very different ways of awarding the surplus. The solutions that we introduce divide the surplus in a more egalitarian way than DI which (in the simulations) typically awards more than half of the surplus to a single agent. For the intuition behind this, consider a problem in which all but agent i have the same inflow whereas agent i is without inflow. All agents have identical benefit functions. Then DI awards the entire surplus to agent i, the agent without inflow. In contrast, the solutions that we introduce share the surplus more or less equally among all agents but i. We find that *Upstream incremental* (van den Brink et al., 2007; Herings et al., 2007) behaves in a way similar to DI in these respects. We remark finally that, whether a more egalitarian solution is "better" or "more fair" depends entirely on the circumstances surrounding the river. If there only is transfer of water between two agents, then sharing the surplus between these two agents (which is not particularly egalitarian for the river system as a whole) may be viewed as fair. On the other hand, if there is an even flow of water along the entire river, then a more egalitarian solution may be preferable.

We acknowledge that there are other solutions established in the literature but restrict the comparisons to the incremental solutions. By now, there is a large literature on the river sharing problem with extensions to satiated preferences (Ambec and Ehlers, 2008b), multiple springs (Khmelnitskaya, 2010; van den Brink et al., 2012; Béal et al., 2015) and uncertain inflows (Ambec et al., 2013). In Section 7, we introduce a generalized model which covers many of these extensions. Other solutions have been derived for instance through a market-based approach (Wang, 2011) or through strategic negotiations (Adams et al., 1996). For surveys of the literature, we refer to Ambec and Ehlers (2008a), Beal et al. (2013), and Ansink and Houba (2015). There are also closely related problems that take a slightly different approach: sharing the cost of cleaning a polluted river (Ni and Wang, 2007; Dong et al., 2012; Alcalde-Unzu et al., 2015; van den Brink et al., 2017), sharing irrigation costs (Aadland and Kolpin, 1998, 2004; Moulin, 2004), and river claims problems (Ansink and Weikard, 2015; Ansink et al., 2017) to which an axiomatic approach also is used. In contrast, the topic that we focus mainly on, implementation, has been less studied. The single exception is the working paper by Ambec and Sprumont (2000) in which DI is shown to be implemented through a particular extensive-form game.

Since we use noncooperative bargaining as part of our implementation our study is also connected to the literature on cooperative bargaining solution to the water allocation problem (Kilgour and Dinar, 2001; Dinar, 2001; Houba, 2008; Ansink and Weikard, 2009; Houba et al., 2014; Degefu et al., 2016). In particular, Houba et al. (2014) suggest to use asymmetric bargaining utilizing, as we do, that quasi-linear preferences allows the problem to be decomposed into two problems: finding a welfare maximizing allocation, and subsequently finding monetary compensations.

Our second stage is adapted from the literature on noncooperative bargaining with random proposer and draws from the work of 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). In connection with contests (Tullock, 1980), several papers analyze ways to determine agents' bargaining power through contest success functions (see, for instance, Skaperdas, 1996; Rai and Sarin, 2009; Yildirim, 2010). However, in this literature, the success an agent has typically depends on the agent's effort, whereas our recognition functions that implement efficient outcomes are not permitted to.

The paper is structured as follows. We present the model in Section 2. In Section 3, we introduce the class of mechanisms. In Section 4, we present our results on implementation. In Section 5, we take an axiomatic approach to select from the implementable solutions. In Section 6, we contrast the new solutions with the incremental solutions through simulations. Finally, we discuss an extension in Section 7.

2 Model

We follow the model introduced by Ambec and Sprumont (2002). There is a finite set of locations along a river, or **agents**, $N = \{1, \ldots, n\}$. Lower numbers of N represent more upstream locations, so agent 1 is the most upstream agent (the "source"). For each agent i, there is an **inflow** to the river $e_i \geq 0$ and a **benefit function** $b_i \colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ that captures i's utility of water consumption. Let $e = (e_1, \ldots, e_n) \in \mathbb{R}^n_{\geq 0}$. Benefit functions are differentiable for positive consumption, strictly increasing, strictly concave, and slope infinitely at the origin.

Although agents are entitled to consume their own inflow, welfare gains for the river system as a whole may be obtained by passing water downstream. The more costly option of sending water upstream is precluded throughout. Thus, a water **allocation** $x = (x_1, \ldots, x_n) \in \mathbb{R}^n_{\geq 0}$ is such that $x_1 \leq e_1$, $x_1 + x_2 \leq e_1 + e_2$, and so on, until $x_1 + \cdots + x_n = e_1 + \cdots + e_n$. Let $B(x) = \sum_i b_i(x_i)$ be the aggregate welfare at allocation x and $\Delta(x) = B(x) - B(e)$ be the change in welfare due to the transfer of water. The set of allocations is $X \subseteq \mathbb{R}^n_{\geq 0}$.

To reward agents for transferring water and thereby creating a surplus, we use **monetary compensations** $c = (c_1, \ldots, c_n) \in \mathbb{R}^n$ such that $\sum_i c_i = 0$. Preferences are quasi-linear in money: given allocation x and compensations c, agent i's utility is $u_i(x, c) = b_i(x_i) + c_i$. The set of utility vectors ("outcomes") attainable under x is $U(x) \subseteq \mathbb{R}^n$:

$$U(x) = \{(b_i(x_i) + c_i)_{i \in N} \in \mathbb{R}^n \mid \textstyle\sum_i c_i = 0\} = \{u \in \mathbb{R}^n \mid \textstyle\sum_i u_i = B(x)\}.$$

The outcome u Pareto-improves upon another outcome v if, for each agent i and some agent j, $u_i \geq v_i$ and $u_j > v_j$. An outcome is **efficient** if it cannot be Pareto-improved. As preferences are quasi-linear, the outcome $u \in U(x)$ is efficient if and only if its associated allocation x maximizes the sum of benefits, $B(x)$. Furthermore, as each benefit function b_i is strictly concave, there is a unique allocation x^* that solves $\max_{x \in X} B(x)$. Notice that x^* also maximizes the welfare change $\Delta(x^*) \geq 0$. Hence, the set of efficient outcomes is $U(x^*)$.

3 Class of mechanisms

We imagine that a benevolent social planner assists in recommending a solution that is agreeable to all agents. The planner has only partial information: the agents' locations and inflows are naturally observable, but the individual benefit functions are not. Therefore, the efficient allocation cannot be determined without consulting the agents. For this purpose, the planner designs a game ("mechanism") in such a way that equilibrium play reveals the efficient allocation. Throughout, agents are assumed to have complete information (in particular about the other agents' benefit functions).

We restrict attention to a specific class of mechanisms. These mimic the way conflicts on international waterways are resolved in practice, that is, through multilateral bargaining. The mechanisms consist of two stages: the agents first sequentially choose consumption and then bargain over the associated monetary compensations. Moreover, the consumption choices in Stage 1 affect the distribution of power in Stage 2. Specifically, to each allocation $x \in X$, the planner associates a point $p(x) \in [0, 1]^n$ in the n-simplex, where $p_i(x)$ is agent i's "recognition probability". Formally, $p_i(x)$ is the probability with which agent i gets to make a proposal in Stage 2 given that allocation x is realized in Stage 1: a high probability means that agent i is likely to be recognized as the proposer. In this way, $p_i(x)$ is a measure of i's bargaining power (we use the terms interchangeably).

By using a well-designed recognition function, the planner can incentivize agents to choose the efficient allocation. The only restriction that we impose on the recognition probabilities is that $p_i(x) = p_i(x_1, \ldots, x_n)$ is continuous in each x_j. The 'class of mechanisms' covers all possible recognition functions. Next, we describe the two stages in further detail.

Stage 1: Sequential consumption choice The desired consumption levels are submitted in sequence, starting from the source and proceeding downstream. The feasibility constraints must be met: agent 1 chooses $x_1 \geq 0$ such that $x_1 \leq e_1$, agent 2 chooses $x_2 \geq 0$ such that $x_1 + x_2 \leq e_1 + e_2$, and so on, until agent n is left with $x_n \geq 0$ such that $x_1 + \cdots + x_n = e_1 + \cdots + e_n$. Thus, at the end of Stage 1, the allocation x has been determined and thereby also the set of outcomes $U(x)$ and the recognition probabilities $p_i(x)$.

Stage 2: Infinitely repeated, random proposer, unanimity bargaining Agents bargain, possibly indefinitely, over which outcome in $U(x)$ to implement (implicitly, they determine the monetary compensations c). Until unanimous agreement has been reached, each round of negotiations starts with a proposer being randomly selected. In particular, agent i is recognized as the proposer with probability $p_i(x)$ as determined in Stage 1. The proposer chooses a utility vector $u \in U(x)$, which the agents in random order either accept or reject. If the proposal is unanimously accepted, then it is implemented and the game ends. Otherwise, we proceed to the next round with a possibly different proposer. Agents have common discount factor $\delta \in (0, 1)$. We restrict attention to *stationary* strategies, which can be represented by a pair $(u^i, a_i) \in U(x) \times \mathbb{R}$: agent i always proposes u^i and always accepts a proposal if awarded at least a_i.

If agreement is reached on utility vector u in the first round of negotiations, so $u_j \geq a_j$ for each agent j, then agent i's payoff is u_i. If agreement is never reached, then agent i's payoff is $b_i(e_i)$ as i consumes e_i throughout the endless negotiations. This is consistent with exponential discounting whenever agent i derives utility $(1 - \delta)b_i(e_i)$ from each round of failed negotiations:

$$b_i(e_i) = \lim_{r\to\infty}(1 + \delta + \cdots + \delta^r)(1 - \delta)b_i(e_i).$$

Thus, if agreement is reached on proposal u after r rounds of failed negotiations, then i's payoff is

$$(1 + \delta + \cdots + \delta^{r-1})(1 - \delta)b_i(e_i) + \delta^r u_i.$$

4 Implementation of efficient outcomes

We solve the two-stage game by backward induction and first determine expected payoffs in Stage 2. This is by itself a "game within the two-stage game" onto which we apply subgame-perfection. As Stage 2 is an infinite-horizon game with discount factor δ less than one, we can apply the *one-stage-deviation principle*: a strategy profile is subgame-perfect if and only if no agent can benefit by deviating in a single round and reverting to the original strategy thereafter (see Fudenberg and Tirole, 1991, Theorem 4.2).

Theorem 1 shows that the surplus, in expectation, is shared in proportion to the recognition probabilities when agents bargain optimally. The result does not rely on particularly patient agents: it holds regardless of the discount factor $\delta \in (0, 1)$. We refer to Eraslan and McLennan (2013) and references therein for a general discussion on the uniqueness of equilibrium payoffs in models of coalitional bargaining.

Theorem 1. *An agent's expected payoff is the same in all stationary, subgame-perfect equilibria of Stage 2. In particular,*

A. *If $\Delta(x) > 0$, then the expected payoff is $b_i(e_i) + p_i(x)\Delta(x)$ and the equilibrium is unique;*
B. *If $\Delta(x) \leq 0$, then the expected payoff is $b_i(e_i)$ and there are several payoff-equivalent equilibria.*

Proof. As each agent j is guaranteed payoff $b_j(e_j)$ by rejecting all proposals through a high-enough acceptance threshold, equilibrium acceptance thresholds satisfy $a_j \geq b_j(e_j)$. Therefore, if $\Delta(x) = B(x) - B(e) < 0$, then there is no proposal $u \in U(x)$ such that $u_j \geq b_j(e_j)$ for each agent j, so agreement is never reached. Agent i's expected payoff is then $\mathbb{E}\pi_i = b_i(e_i)$. If $\Delta(x) = 0$, then the only acceptable proposal is u such that $u_i = b_i(e_i)$, so again $\mathbb{E}\pi_i = b_i(e_i)$.

To derive a contradiction, suppose that $a_i > b_i(e_i)$ in an equilibrium in which agreement is never reached. Then there exists a subgame away from the equilibrium path in which a proposal is accepted by all the other agents and i is offered u_i such that $\mathbb{E}\pi_i = b_i(e_i) < u_i < a_i$. Agent i is then better off accepting the proposal, contradicting that the equilibrium is subgame-perfect. Hence, $a_i = b_i(e_i)$ in such an equilibrium.

If $\Delta(x) > 0$, then agreement is reached in equilibrium. If not, then each proposer benefits from proposing, say, u such that $u_i = b_i(e_i) + p_i(x)\Delta(x)$. Moreover, agreement is reached without delay. If not, say agent i's first-round proposal is rejected, then i should look at the other agents' proposals and switch to the one that is accepted and awards i the most. Hence, i's expected payoff is $\mathbb{E}\pi_i = \sum_j p_j(x)u_i^j$. Furthermore, each proposal is in $U(x)$ so it adds to $B(x)$. Therefore,

$$\textstyle\sum_i \mathbb{E}\pi_i = \sum_i \sum_j p_j(x)u_i^j = \sum_j p_j(x) \sum_i u_i^j = \sum_j p_j(x)B(x) = B(x).$$

By the stationarity of the game and the strategies, i's expected payoff from rejecting a proposal is

$$\delta\mathbb{E}\pi_i + (1-\delta)b_i(e_i).$$

In equilibrium, i accepts all proposals that exceed this. Hence, this is the acceptance threshold a_i:

$$a_i = \delta\mathbb{E}\pi_i + (1-\delta)b_i(e_i) \implies \textstyle\sum_i a_i = \delta \sum_i \mathbb{E}\pi_i + (1-\delta)\sum_i b_i(e_i) = \delta B(x) + (1-\delta)B(e).$$

Each agent j optimally keeps as much as possible of the surplus by awarding i precisely $u_i^j = a_i$. Thus:

$$\begin{aligned}\mathbb{E}\pi_i &= \textstyle\sum_j p_j(x)u_i^j \\ &= a_i + p_i(x)(B(x) - \textstyle\sum_j a_j) \\ &= \delta\mathbb{E}\pi_i + (1-\delta)b_i(e_i) + p_i(x)(B(x) - \delta B(x) - (1-\delta)B(e)) \\ &= \delta\mathbb{E}\pi_i + (1-\delta)(b_i(e_i) + p_i(x)\Delta(x)).\end{aligned}$$

It follows that $\mathbb{E}\pi_i = b_i(e_i) + p_i(x)\Delta(x)$. □

Whereas the surplus Δ is maximized at the efficient allocation x^*, an agent's recognition probability p_i generally is not. Therefore, if agent i can enforce an inefficient allocation that increases p_i by relatively more than it decreases Δ, then i's expected payoff increases and the implemented outcome is inefficient (note here that it is never beneficial for an agent to make Δ negative). In this way, only some of the mechanisms implement efficient outcomes. This is made precise in Theorem 2 below: in terms of recognition probabilities, only those of agents downstream of i can be affected by i's choice x_i.

Theorem 2. *The following statements are equivalent:*

A. There exists an efficient subgame-perfect equilibrium outcome of the two-stage game;
B. For each agent i and allocations x and y, $x_1, \dots, x_{i-1} = y_1, \dots, y_{i-1} \implies p_i(x) = p_i(y)$.

In addition, if $p_i(x) > 0$ for each agent i and allocation x, then the subgame-perfect equilibrium outcome is unique.

Proof. By Theorem 1, expected payoffs are $\mathbb{E}\pi_i = b_i(e_i) + p_i(x)\Delta(x)$ in Stage 2. We solve Stage 1 through backward induction.

Part $A \impliedby$ Part B: By assumption, $p_i(x)$ is independent of agent i's choice x_i and the choices of those downstream of i. Hence, i can only affect $\Delta(x)$. If $p_i(x) > 0$, then i chooses to maximize Δ, knowing that the agents downstream will do the same. If $p_i(x) = 0$, then i is indifferent between all alternatives, so maximizing Δ is one out of several optimal choices. As agents have complete information, they can determine the efficient $x = x^*$ and choose accordingly. This is an equilibrium, and it is unique if $p_i(x) > 0$ for all agents.

Part $A \implies$ Part B: By contradiction, suppose that Part A is true but Part B is not.

Step 1. Agent n chooses last and has to select x_n such that $\sum_i x_i = \sum_i e_i$. Hence, there cannot be x_n and $y_n \neq x_n$ for n to choose between, so Part B is always true for agent n.

Step 2. Consider agent $n-1$. Conditional on the choices by agent 1 through $n-2$, $n-1$ chooses x_{n-1} and indirectly x_n. Label two such choices y_{n-1} and z_{n-1}. Let $y \equiv (x_1, \dots, x_{n-2}, y_{n-1}, y_n) \in X$ and $z \equiv (x_1, \dots, x_{n-2}, z_{n-1}, z_n) \in X$. Suppose that $p_{n-1}(y) \neq p_{n-1}(z)$, in particular, $p_{n-1}(y) < p_{n-1}(z)$. By continuity, we may assume that $y_{n-1}, z_{n-1} > 0$ and that y and z are such that there is some (arbitrarily small) transfer of water between all agents. We will identify benefit functions for which y is efficient but agent $n-1$ nevertheless prefers z. That is, $p_{n-1}(y)\Delta(y) < p_{n-1}(z)\Delta(z)$.

Define $\ell_i(w_i) = \min\{w_i, y_i, z_i\}$. Let $\varepsilon > 0$ and assume that each benefit function b_i is such that $\ell_i(w_i) \leq b_i(w_i) \leq \ell_i(w_i) + \varepsilon$ everywhere; see Figure 1. As $\varepsilon > 0$, there is sufficient room between the two piece-wise linear functions to "twist" the continuous benefit functions as we like. In particular, we can equate marginal benefits at y: for each $\{i, j\} \subseteq N$, $b_i'(y_i) = b_j'(y_j)$. This is necessary and sufficient to ensure that y is efficient when there is water transferred between all agents.

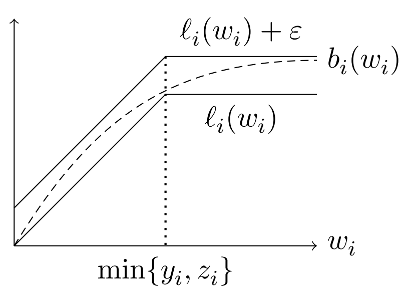

Figure 1: The benefit function b_i is within the bounds set by the piece-wise linear functions.

By construction, $b_i(y_i) \leq \ell_i(y_i) + \varepsilon = \ell_i(z_i) + \varepsilon \leq b_i(z_i) + \varepsilon$. Thus, $B(y) \leq B(z) + n\varepsilon$, so $\Delta(y) \leq \Delta(z) + n\varepsilon$. As $y \neq e$ and y is efficient, $\Delta(y) > 0$, so, for a small-enough ε, $\Delta(z) > 0$. Therefore, as $p_{n-1}(z) > p_{n-1}(y)$, there is a small-enough $\varepsilon > 0$ such that

$$\frac{\Delta(y)}{\Delta(z)} \leq \frac{\Delta(z) + n\varepsilon}{\Delta(z)} < \frac{p_{n-1}(z)}{p_{n-1}(y)}.$$

Then $n-1$ prefers the inefficient z over y and we obtain the desired contradiction. Hence, for all choices y_{n-1} and z_{n-1} that $n-1$ can make, $p_{n-1}(y) = p_{n-1}(z)$. But then, as in Part $A \Longleftarrow$ Part B, agent $n-1$ always maximizes Δ.

Step 3. Next, consider agent $i = n-2, n-3, \ldots, 1$ who chooses x_i foreseeing that agents $i+1$ through n will consume to maximize Δ. We proceed as in Step 2 and identify benefit functions such that y is efficient and agent i prefers z over y. However, we also ensure that the choices by agents $i+1$ through n at z maximize Δ given that i chooses z_i. (In Step 2, only agent n follows $n-1$, and n must maximize Δ.) By continuity, we may assume that $y_i, z_i > 0$ and that y and z are such that there is water transferred between all agents. Then z is efficient for the agents downstream of i whenever their marginal benefits are equal. Thus, for each $\{j,k\} \subseteq \{i+1,\ldots,n\}$, we set $b_j'(z_j) = b_k'(z_k)$. Hence, if $y_i > z_i$, so there is more to divide for $i+1,\ldots,n$ at z, then all consume more at z than at y. Figure 2 sketches such a profile of benefit functions.

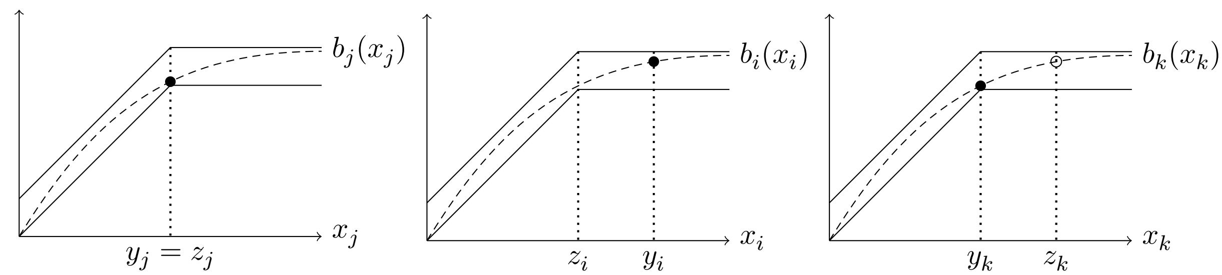

Figure 2: (Illustration for $y_i > z_i$.) Left is for $j < i$ and right for $k > i$. The benefit functions should equate marginal benefits at the filled dots and likewise at the non-filled dots.

The proof now proceeds as in Step 2 and we derive the same contradiction that implies $p_i(y) = p_i(z)$. Once we have shown this for agent $i = 1$, the proof is complete. □

A stronger solution concept that allows collusion between agents would further limit the set of viable mechanisms. We conjecture that, if coordination is possible in Stage 1 and the planner wants to implement an efficient outcome, then recognition probabilities cannot depend on the allocation at all. That is, for each agent i and allocations x and y, $p_i(x) = p_i(y)$.

A plausible objective for the planner is to award the surplus to those "most deserving" of it. However, this is somewhat ambiguous: the planner must set the recognition probabilities without knowing the benefit functions, and, furthermore, agents can contribute in different ways. For instance, with inflows $e = (2,1,0)$ and identical benefit functions, all agents contribute to the efficient allocation $x^* = (1,1,1)$: agent 1 has excess inflow, agent 2 passes it on, and agent 3 extracts more benefit from it. Hence, there is generally not a unique "right" way to divide the surplus. However, the planner may justify *excluding* an agent from sharing the surplus. In particular, say agent i is a **dummy agent** at allocation x if no water is passed on to or from i: $\sum_{j<i} x_j = \sum_{j<i} e_j$ and $x_i = e_i$. With inflows $e = (2,0,1)$ and identical benefit functions, agent 3 is a dummy: water is only transferred from 1 to 2 at the efficient $x^* = (1,1,1)$. As it suffices to know the inflows and the allocation to determine whether an agent is a dummy, the planner can easily exclude dummy agents from sharing the surplus. However, this comes at a cost: Theorem 3 shows that no such mechanism always implements an efficient outcome.

Theorem 3. *Let $n \geq 4$. The following statements are mutually exclusive:*

A. There exists an efficient subgame-perfect equilibrium outcome of the three-stage mechanism;
B. If agent i is a dummy agent at the efficient allocation x^*, then i's equilibrium payoff is $b_i(e_i)$.

Proof. By Theorem 2, Part A is equivalent to that only the probabilities of agents downstream of i can depend on i's choice. In particular:

Step 1. Only $p_n(x)$ can depend on x_{n-1}. However, as probabilities add to one, if $p_n(x)$ changes with x_{n-1}, then so must also someone else's probability. This is a contradiction, so even p_n is independent of x_{n-1}. Hence, the planner cannot condition p_n on whether $n-1$ passes on water or not. Thus, to guarantee that agent n gets no share of the surplus when a dummy, the planner sets $p_n(x) = 0$ for each allocation x, including the efficient x^*.

Step 2. Only $p_n(x)$ and $p_{n-1}(x)$ can depend on x_{n-2}. However, Step 1 implies that $p_n(x) = 0$. The argument proceeds as in Step 1 with the conclusion that the planner sets $p_{n-1}(x) = 0$ for each allocation x.

Step 3. We repeat the argument to find that $p(x) = (0, \dots, 0)$, contradicting $\sum_i p_i(x) = 1$. □

Theorem 3 does not extend to $n \leq 3$. For instance, for $n = 3$, if $\Delta(x) \neq 0$, then agent 2 cannot be a dummy agent. Therefore, $p(x) = (0, 1, 0)$ satisfies both statements of Theorem 3.

To conclude, Theorem 2 guides the planner on how to set the recognition probabilities such that the resulting water allocation is efficient, and Theorem 3 shows that implementation is incompatible with excluding dummy agents from sharing the surplus. An agent's bargaining power cannot depend on the agent's consumption or transfer of water, and consumption is only allowed to affect the bargaining power of upstream agents. However, on the positive side, Theorem 2 still offers a wide range of mechanisms that implement efficient outcomes and condition bargaining power on consumption choice.

5 Choosing recognition probabilities

In order to narrow down the set of desirable recognition functions, we take an axiomatic approach. Each proposed axiom reflects a normatively appealing property in the present context. Taken in conjunction, the axioms lead uniquely to new solutions which all satisfy Part B of Theorem 2. That is, if the planner uses the probabilities specified in equations (1), (2), or (3), then the associated two-stage mechanism implements an efficient outcome.

As a first step, we extend the model in order to relate problems that differ in population size. We assume now an infinite set of potential agents $\mathbb{N} = \{1, 2, \dots\}$. A **population** is a finite subset of $\mathbb{N}$ and represents a set of agents as in Section 2; with some abuse, let n denote the number of agents. Given inflows $e \in \mathbb{R}^n_{\geq 0}$, the set of allocations is $X(e) \subseteq \mathbb{R}^n_{\geq 0}$. A **recognition function** p selects, for each population $N \subset \mathbb{N}$, inflows $e \in \mathbb{R}^n_{\geq 0}$, and allocation $x \in X(e)$, a point $p(N, e, x) \in [0, 1]^n$ in the n-simplex. We assume that p is *anonymous* in the sense that only agents' relative locations matter, not their identities. Hence, it is without loss to consider the generic population $N = \{1, \dots, n\}$.

In a two-agent problem, any surplus is created by agent 1 passing on water downstream for agent 2 to use more efficiently – this is true no matter the benefit functions, inflows, or allocation. *Two-agent independence* asserts that we should share the surplus in the same way in all two-agent problems. That is, if we value excess inflow as twice as important as being able to use the water more efficiently in one problem, then we should keep this trade-off for all two-agent problems. Moreover, both agents play a part in creating the gain, so both should get a share of it.

Axiom 1 (Two-agent independence)**.** For each $\{e, e'\} \subseteq \mathbb{R}^2_{\geq 0}$, $x \in X(e)$, and $x' \in X(e')$,

$$p(\{1,2\}, e, x) = p(\{1,2\}, e', x') \in (0,1)^2.$$

Next, we turn to the consistency principle (for normative underpinnings, see Thomson, 2011). Loosely speaking, consistency concerns scalability: given that we know how to solve simple two-agent situations in a desirable way, consistent extension of such a solution allows scaling to any number of agents. Here, we focus on a minimal version that refers only to the case when either the most upstream or most downstream agent is removed. *Consistency* requires that the recognition probabilities of the agents in the reduced problem are unchanged when adjusting the scale of the probabilities accordingly.

Axiom 2 (Consistency)**.** For each $N \subset \mathbb{N}$, $e \in \mathbb{R}^n_{\geq 0}$, $x \in X(e)$, $j \in \{1, n\}$, and $i \neq j$,

$$p_i(N, e, x) = p_i(N \setminus \{j\}, e_{-j}, x_{-j}) \cdot (1 - p_j(N, e, x)).$$

Theorem 4 shows that a mechanism satisfies *two-agent independence* and *consistency* whenever its recognition probabilities form a geometric series, where the parameter λ is the ratio between the probabilities of two consecutive agents. A special case is *Equal division* which always shares the surplus equally through $p_i(N, e, x) = 1/n$. Note that equation (1) tends to $1/n$ as $\lambda \to 1$.

Theorem 4. *The recognition function p satisfies two-agent independence and consistency if and only if it is Equal division or there is $\lambda > 0$ such that, for each $N \subset \mathbb{N}$, $e \in \mathbb{R}^n_{\geq 0}$, $x \in X(e)$, and $i \in N$,*

$$p_i(N, e, x) = p_i^\lambda(N, e, x) = \frac{1-\lambda}{1-\lambda^n}\lambda^{i-1}. \tag{1}$$

Proof. $\Longleftarrow$ Fix $\lambda > 0$, $\lambda \neq 1$. As p^λ is independent of e and x, p^λ satisfies *two-agent independence.* Let (N, e, x) be an arbitrary problem and create the reduced problem by removing agent 1. First,

$$p_1^\lambda(N, e, x) = \frac{1-\lambda}{1-\lambda^n} \implies 1 - p_1^\lambda(N, e, x) = \frac{1-\lambda^n - 1 + \lambda}{1-\lambda^n} = \frac{\lambda - \lambda^n}{1-\lambda^n} = \frac{1-\lambda^{n-1}}{1-\lambda^n}\lambda.$$

Therefore,

$$\begin{aligned} p_i^\lambda(N \setminus \{1\}, e_{-1}, x_{-1}) \cdot (1 - p_1^\lambda(N, e, x)) &= \left(\frac{1-\lambda}{1-\lambda^{n-1}}\lambda^{i-2}\right)\left(\frac{1-\lambda^{n-1}}{1-\lambda^n}\lambda\right) \\ &= \frac{1-\lambda}{1-\lambda^n}\lambda^{i-1} = p_i^\lambda(N, e, x). \end{aligned}$$

The approach is analogous for the reduced problem that is obtained by removing agent n. Hence, p^λ satisfies *consistency.* Furthermore, it is easy to verify that *Equal division* satisfies *two-agent independence* and *consistency.*

$\Longrightarrow$ By *two-agent independence*, there is $\lambda > 0$ such that, for all two-agent problems,

$$p(\{1,2\}, e, x) = \left(\frac{1}{1+\lambda}, \frac{\lambda}{1+\lambda}\right) = p^\lambda(\{1,2\}, e, x).$$

That is, λ denotes the ratio of the two agents' probabilities.

Generally, consider any consecutive agents i and $i+1$. By repeated application of *consistency*, we can remove all agents upstream from i and downstream from $i+1$. This leaves a two-agent problem in which the probabilities of agents i and $i+1$ have been rescaled compare to the original problem. However, the scale does not affect the probability ratio. Hence,

$$\frac{p_{i+1}(N,e,x)}{p_i(N,e,x)} = \frac{p_{i+1}(\{i,i+1\},e,x)}{p_i(\{i,i+1\},e,x)} = \lambda.$$

This pins down the functional form: there is $C > 0$ such that $p_i(N,e,x) = C\lambda^{i-1}$. We determine C through $\sum_i p_i(N,e,x) = C\sum_i \lambda^{i-1} = 1$. If $\lambda = 1$, then $C = 1/n$, and we obtain *Equal division.* If $\lambda \neq 1$, then we derive the desired expression:

$$\begin{array}{rl} C(1+\lambda+\cdots+\lambda^{n-1}) & = 1 \\ C(\lambda+\cdots+\lambda^{n-1}+\lambda^n) & = \lambda \end{array} \implies C(1-\lambda^n) = 1-\lambda \iff C = \frac{1-\lambda}{1-\lambda^n}. \qquad \square$$

By Theorem 2, implementability implies that the source's recognition probability is constant. An alternative way to characterize a desirable recognition function is to settle the source's bargaining power for any problem. In particular, *source-solution* asserts that proportionality of inflows constitutes a desirable recognition probability for the source.

Axiom 3 (Source-solution)**.** For each $N \subset \mathbb{N}$, $e \in \mathbb{R}^n_{\geq 0}$, and $x \in X(e)$,

$$p_1(N,e,x) = \frac{e_1}{\sum_i e_i}.$$

Source-solution may appear unfair when the source is a dummy agent with positive inflow, but recall that implementability is incompatible with excluding dummy agents from getting positive shares of the surplus (Theorem 3). Introducing *source-solution* allows us weaken consistency:

Axiom 4 (Source-consistency)**.** For each $N \subset \mathbb{N}$, $e \in \mathbb{R}^n_{\geq 0}$, $x \in X(e)$, and $i \geq 2$,

$$p_i(N,e,x) = p_i(N \setminus \{1\}, e', x_{-1}) \cdot (1 - p_1(N,e,x)),$$

where $e' = (e_1 + e_2 - x_1, e_3, \ldots, e_n)$.

Source-solution requires that the source's bargaining power should reflect the water available to the source (e_1) in proportion to the total amount of water available in the river ($e_1 + \cdots + e_n$). *Source-consistency* then ensures that we apply the same reasoning as we proceed downstream. For the second agent, the water available is both the local inflow and the water passed on from the source ($e_1 + e_2 - x_1$). This is set in proportion to the water available from agent 2 and downstream ($e_1 + \cdots + e_n - x_1$). Thus, *source-solution* and *source-consistency* characterize a solution that equates an agent's recognition probability with the power as implied by the water made available to the agent.

Theorem 5. *The recognition function p satisfies source-solution and source-consistency if and only if, for each $N \subset \mathbb{N}$, $e \in \mathbb{R}^n_{\geq 0}$, $x \in X(e)$, and $i \in N$,*

$$p_i(N,e,x) = p_i^*(N,e,x), \tag{2}$$

where

$$\begin{aligned}
p_1^*(N,e,x) &= 1 - f(1),\\
p_2^*(N,e,x) &= f(1)(1-f(2)),\\
&\vdots\\
p_n^*(N,e,x) &= f(1)\cdots f(n-1)(1-f(n))
\end{aligned}$$

with $f(k) = \frac{\sum_{j>k} e_j}{\sum_j e_j - \sum_{j<k} x_j}$ *and* $f(n) = 0$.

Proof. $\Longleftarrow$ We first show that p^* satisfies *source-solution*:

$$p_1^*(N,e,x) = 1 - f(1) = 1 - \frac{\sum_{j>1} e_j}{\sum_j e_j} = \frac{e_1}{\sum_j e_j}.$$

We turn to *source-consistency*. As in its definition, remove agent 1 to obtain the reduced problem $(N \setminus \{1\}, e', x_{-1})$. The total inflow is reduced by 1's consumption:

$$\textstyle\sum_{j>1} e'_j = e_1 + e_2 - x_1 + e_3 + \cdots + e_n = \sum_j e_j - x_1.$$

Moreover, $\sum_{1<j<k} x_j = \sum_{j<k} x_j - x_1$. Therefore, we obtain the same value for $f(k)$ if we compute it on the original problem (N,e,x) as if we do so on the reduced problem $(N \setminus \{1\}, e', x_{-1})$:

$$\frac{\sum_{j>k} e'_j}{\sum_{j>1} e'_j - \sum_{1<j<k} x_j} = \frac{\sum_{j>k} e_j}{\sum_j e_j - x_1 - \sum_{j<k} x_j + x_1} = \frac{\sum_{j>k} e_j}{\sum_j e_j - \sum_{j<k} x_j} = f(k).$$

It follows that $p_i^*(N \setminus \{1\}, e', x_{-1}) = f(2)\cdots f(i-1)(1-f(i))$. Note also that the above step can be repeated: if we remove the new source in the reduced problem, then we again derive the same value $f(k)$ for the remaining agents k. As $f(1) = 1 - p_1^*(N,e,x)$, we derive the desired conclusion:

$$p_i^*(N,e,x) = f(1)\cdots f(i-1)(1-f(i)) = p_i^*(N \setminus \{1\}, e', x_{-1}) \cdot (1 - p_1^*(N,e,x)).$$

$\Longrightarrow$ Consider an arbitrary problem (N,e,x). For $j \in N$, define the reduced problem $(\{j,\ldots,n\}, e^j, x^j)$ with $e^j = (\sum_{i\leq j} e_i - \sum_{i<j} x_i, e_{j+1}, \ldots, e_n)$ and $x^j = (x_j, \ldots x_n)$. By *source-solution*, agent j is awarded $1 - f(j)$ in her associated reduced problem:

$$p_j(\{j,\ldots,n\}, e^j, x^j) = \frac{e^j_j}{\sum_i e^j_i} = \frac{\sum_{i\leq j} e_i - \sum_{i<j} x_j}{\sum_i e_i - \sum_{i<j} x_j} = 1 - f(j).$$

Therefore, $1 - p_j(\{j,\ldots,n\}, e^j, x^j) = f(j)$. Applying repeatedly to *source-consistency*, we derive desired conclusion:

$$\begin{aligned}
p_j(N,e,x) &= p_j(\{2,\ldots,n\}, e^2, x^2) \cdot (1 - p_1(N,e,x))\\
&= p_j(\{3,\ldots,n\}, e^3, x^3) \cdot (1 - p_2(\{2,\ldots,n\}, e^2, x^2)) \cdot (1 - p_1(N,e,x))\\
&= \ldots\\
&= p_j(\{j,\ldots,n\}, e^j, x^j) \cdot (1 - p_{j-1}(\{j-1,\ldots,n\}, e^{j-1}, x^{j-1})) \cdots (1 - p_1(N,e,x))\\
&= f(1)\cdots f(j-1)(1-f(j)) = p_j^*(N,e,x).
\end{aligned}$$

□

Lastly, we note that the term 'location' can be ambiguous. For instance, we may consider regional inflows and let agents be represented by regions, or we add up to country inflows and let agents represent countries. Ideally, an agent's bargaining power should be independent of the way in which the agent is represented. *Source merging neutrality* captures this but only in a minimal way: it applies only when the two most upstream agents merge into a new source. Moreover, the axiom is silent on how the other agents may be affected from this change.

Axiom 5 (Source merging neutrality)**.** For each $N \subset \mathbb{N}$, $e \in \mathbb{R}^n_{\geq 0}$, and $x \in X(e)$,

$$p_1(N, e, x) + p_2(N, e, x) = p_2(N \setminus \{1\}, e', x'),$$

where $e' = (e_1 + e_2, e_3, \ldots, e_n)$ and $x' = (x_1 + x_2, x_3, \ldots, x_n)$.

Together with *source-solution*, *source merging neutrality* characterizes the simple recognition function that sets recognition probabilities in proportion to inflows.

Theorem 6. *The recognition function p satisfies source-solution and source merging neutrality if and only if, for each $N \subset \mathbb{N}$, $e \in \mathbb{R}^n_{\geq 0}$, $x \in X(e)$, and $j \in N$,*

$$p_j(N, e, x) = p^\circ_j(N, e, x) = \frac{e_j}{\sum_i e_i}. \tag{3}$$

Proof. $\Longleftarrow$ It is immediate that p° satisfies *source-solution*. For *source merging neutrality*,

$$p^\circ_1(N, e, x) + p^\circ_2(N, e, x) = \frac{e_1}{\sum_i e_i} + \frac{e_2}{\sum_i e_i} = \frac{e_1 + e_2}{\sum_i e_i} = \frac{e'_2}{\sum_i e'_i} = p^\circ_2(N \setminus \{1\}, e', x').$$

$\Longrightarrow$ Consider an arbitrary problem (N, e, x). By *source-solution*,

$$p_1(N, e, x) = \frac{e_1}{\sum_i e_i} = p^\circ_1(N, e, x).$$

Let $j \in N$ and assume, for each $i = 1, \ldots, j - 1$, $p_i(N, e, x) = p^\circ_i(N, e, x)$. Define the reduced problem $(\{j, \ldots, n\}, e^j, x^j)$ with $e^j = (e_1 + \cdots + e_j, e_{j+1}, \ldots, e_n)$ and $x^j = (x_j, \ldots x_n)$. By *source-solution*,

$$p_j(\{j, \ldots, n\}, e^j, x^j) = \frac{e^j_j}{\sum_i e^j_i} = p^\circ_j(\{j, \ldots, n\}, e^j, x^j).$$

Applying to *source merging neutrality* repeatedly, we get that

$$\begin{aligned} p^\circ_j(\{j, \ldots, n\}, e^j, x^j) &= p_j(\{j, \ldots, n\}, e^j, x^j) \\ &= p_1(N, e, x) + \cdots + p_{j-1}(N, e, x) + p_j(N, e, x) \\ &= p^\circ_1(N, e, x) + \cdots + p^\circ_{j-1}(N, e, x) + p_j(N, e, x). \end{aligned}$$

Therefore,

$$\begin{aligned} p_j(N, e, x) &= p^\circ_j(\{j, \ldots, n\}, e^j, x^j) - \big(p^\circ_1(N, e, x) + \cdots + p^\circ_{j-1}(N, e, x)\big) \\ &= \frac{e_1 + \cdots + e_j}{\sum_i e_i} - \left(\frac{e_1}{\sum_i e_i} + \cdots + \frac{e_{j-1}}{\sum_i e_i}\right) = \frac{e_j}{\sum_i e_i} = p^\circ_j(N, e, x). \end{aligned}$$

□

Example 1. To briefly illustrate how the three solutions, p^λ, p^*, p°, defined above, share the surplus $\Delta(x^*)$, consider the generic three agent case. Let $E = e_1 + e_2 + e_3$ be the total inflow.

First, consider p^λ, that is independent of inflows and consumption. Say, for $\lambda = 0.5$ (that is, for two consecutive agents i and $i+1$: $p_{i+1}/p_i = \lambda = 0.5$) we get:

$$p_1^{0.5} = 0.57, \quad p_2^{0.5} = 0.29, \quad p_3^{0.5} = 0.14.$$

Next, consider p^*, that depends on both the inflows and the consumption of upstream agents. Here we get $f(1) = \frac{e_2+e_3}{E}$ and $f(2) = \frac{e_3}{E-x_1}$. Thus,

$$p_1^* = \frac{e_1}{E}, \quad p_2^* = \frac{e_2+e_3}{E}(1 - \frac{e_3}{E-x_1}), \quad p_3^* = \frac{e_2+e_3}{E}\frac{e_3}{E-x_1}.$$

Finally, for p° we get:

$$p_1^\circ = \frac{e_1}{E}, \quad p_2^\circ = \frac{e_2}{E}, \quad p_3^\circ = \frac{e_3}{E}.$$

Resulting payoffs can be computed as $u_i = b_i(e_i) + p_i\Delta(x^*)$ for every agent i and solution p.

6 Simulations

We conduct simulations to contrast the solutions introduced in Section 5 with the well-known *Downstream incremental* (Ambec and Sprumont, 2002) and *Upstream incremental* (van den Brink et al., 2007; Herings et al., 2007). We simulate 100 random 10-agent instances. Benefit functions are of the form $b_i(x_i) = x_i^{\alpha_i}$ with exponents α_i drawn uniformly from $[0, 1]$. Likewise, inflows are drawn uniformly from $[0, 1]$. For each instance, we determine the efficient allocation x^* and calculate the surplus $\Delta(x^*)$. The incremental solutions implicitly define recognition probabilities as follows. Assuming that the incremental solution selects outcome u, we solve for the recognition probability p_i:

$$u_i = b_i(e_i) + p_i\Delta(x^*) \iff p_i = \frac{u_i - b_i(e_i)}{\Delta(x^*)}.$$

We order the recognition probabilities from smallest to largest. Then, we average the ordered probabilities across the 100 instances. This lets us see how much, on average, the k agents who get the smallest shares of the surplus get in total. The simulation results are plotted in Figure 3.

Figure 3 shows that there is one agent who receives half of the surplus (on average) under the incremental solutions. (This is typically a different agent for *Downstream incremental* than for *Upstream incremental*.) Moreover, the incremental solutions essentially exclude half of the agents from the surplus. In contrast, our proposed solutions award $70 - 75\%$ of the surplus to the five agents who receive most and $25 - 30\%$ to the other agents. As such, our solutions represent a more egalitarian way of sharing the surplus compared to the conventional solutions. Finally, we note that all solutions are fitted surprisingly well by different geometric series.

7 Extension: Implementation in a general setting

Many intricacies of the model are not "needed" to prove that some mechanisms implement efficient outcomes. Therefore, we can generalize the model considerably and extend Theorem 1 and 2 in a

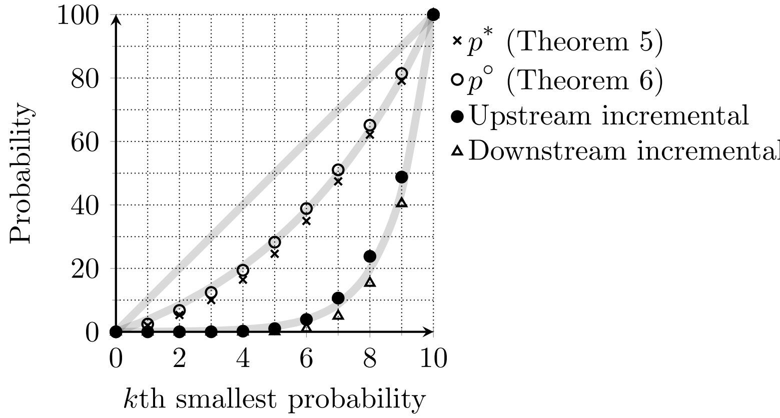

Figure 3: Cumulative distributions averaged across 100 random 10-agent instances. The three shaded curves represent *Equal division* and p^λ (Theorem 4) for $\lambda = 6/5$ (mid) and $\lambda = 9/4$ (low curve).

straightforward way. We do so next by introducing a more flexible model, repeating the relevant parts of Section 2 whenever necessary.

A problem still has a finite set of agents $N = \{1, \ldots, n\}$, but they need not be "on a line". Agent i is immediately upstream of agent j whenever $i \prec j$. The partial order $\leq$ on N is the reflexive, transitive closure of $\prec$. Intuitively, this specification allows parallel rivers with multiple sources. Identify a minimal element of $\leq$ and label this agent 1. Agent 1 chooses x_1 from a closed and bounded set E_1 that represents the consumption possibilities for agent 1. In Section 2, $E_1 = [0, e_1]$, but E_1 may now be in several dimensions, exclude the origin, be non-convex, and so on. In the absence of cooperation along the river, agent 1 has an outside option and derives some benefit therefrom; we label these e_1 and $b_1(e_1) \in \mathbb{R}$, respectively.

The resources not used by agent 1 are passed on to the immediate subordinates. Let agent $2 \succ 1$ be such a subordinate. The consumption space of agent 2 depends on the choice of agent 1. Specifically, agent 2 chooses x_2 from $E_2(x_1)$. An allocation $x = (x_1, \ldots, x_n)$ is such that, for each agent j, $x_j \in E_j(x_{i<j})$, where $x_{i<j} = (x_i)_{i<j}$. Collect in $e = (e_1, \ldots, e_n)$ the agents' outside options and denote each agent i's corresponding benefit $b_i(e_i)$. The allocation x creates aggregate benefit $B(x) \in \mathbb{R}$. We require only that B is continuous in x and that $B(e) = \sum_i b_i(e_i)$. That is, only in the specific point that reflects the agents' outside options does B have to be linear. At other points, B may feature complementarities between agents such as $B(x) = x_1 x_2$. In contrast to the simpler model in Section 2, there may now exist several allocations that maximize B. However, we conjecture that the results still generalize in a simple way: if agent j's choice cannot affect the recognition probability of agent $i \leq j$, then the mechanism implements efficient outcomes.

References

Aadland, D. and Kolpin, V. (1998). Shared irrigation costs: an empirical and axiomatic analysis. *Mathematical Social Sciences*, 35(2):203–218.

Aadland, D. and Kolpin, V. (2004). Environmental determinants of cost sharing. *Journal of*

Economic Behavior & Organization, 53(4):495–511.

Adams, G., Rausser, G., and Simon, L. (1996). Modelling multilateral negotiations: An application to California water policy. *Journal of Economic Behavior & Organization*, 30(1):97–111.

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

Ambec, S., Dinar, A., and McKinney, D. (2013). Water sharing agreements sustainable to reduced flows. *Journal of Environmental Economics and Management*, 66(3):639–655.

Ambec, S. and Ehlers, L. (2008a). *Cooperation and equity in the river sharing problem*. London: Routledge.

Ambec, S. and Ehlers, L. (2008b). Sharing a river among satiable agents. *Games and Economic Behavior*, 64(1):35–50.

Ambec, S. and Sprumont, Y. (2000). Sharing a River. Cahiers de recherche, Universite de Montreal, Departement de Sciences Economiques.

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

Ansink, E., Gengenbach, M., and Weikard, H.-P. (2017). River coalitions and water trade. *Oxford Economic Papers*, 69(2):453–469.

Ansink, E. and Houba, H. (2015). The economics of transboundary river management. In Dinar, A. and Schwabe, K., editors, *Handbook of Water Economics*, chapter 23, pages 434–466. Edward Elgar Publishing.

Ansink, E. and Weikard, H.-P. (2009). Contested water rights. *European Journal of Political Economy*, 25(2):247–260.

Ansink, E. and Weikard, H.-P. (2015). Composition properties in the river claims problem. *Social Choice and Welfare*, 44(4):807–831.

Baron, D. and Ferejohn, J. (1989). Bargaining in Legislatures. *American Political Science Review*, 83:1181–1206.

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

Béal, S., Ghintran, A., Rémila, E., and Solal, P. (2015). The sequential equal surplus division for rooted forest games and an application to sharing a river with bifurcations. *Theory and Decision*, 79(2):251–283.

Binmore, K. (1987). Perfect equilibria in bargaining models. In Binmore, K. and Dasgupta, P., editors, *The Economics of Bargaining*, chapter 5. Blackwell, Oxford.

Binmore, K., Rubinstein, A., and Wolinsky, A. (1986). The Nash bargaining solution in economic modelling. *The RAND Journal of Economics*, 17(2):176–188.

Degefu, D. M., He, W., Yuan, L., and Zhao, J. H. (2016). Water allocation in transboundary river basins under water scarcity: a cooperative bargaining approach. *Water resources management*, 30(12):4451–4466.

Dinar, A. (2001). Scale and equity in water resource development: a nash bargaining model. *Natural Resource Modeling*, 14(4):477–494.

Dong, B., Ni, D., and Wang, Y. (2012). Sharing a polluted river network. *Environmental and Resource Economics*, 53(3):367–387.

Eraslan, H. (2002). Uniqueness of stationary equilibrium payoffs in the baron-ferejohn model. *Journal of Economic Theory*, 103(1):11–30.

Eraslan, H. and McLennan, A. (2013). Uniqueness of stationary equilibrium payoffs in coalitional bargaining. *Journal of Economic Theory*, 148(6):2195–2222.

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

Herings, P. J.-J., van der Laan, G., and Talman, D. (2007). The socially stable core in structured transferable utility games. *Games and Economic Behavior*, 59(1):85–104.

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

Houba, H., van der Laan, G., and Zeng, Y. (2014). Asymmetric nash solutions in the river sharing problem. *Strategic Behavior and the Environment*, 4(4):321–360.

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

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

Moulin, H. (2004). *Fair division and collective welfare*. MIT press.

Ni, D. and Wang, Y. (2007). Sharing a polluted river. *Games and Economic Behavior*, 60(1):176–186.

Rai, B. and Sarin, R. (2009). Generalized contest success functions. *Economic Theory*, 40(1):139–149.

Skaperdas, S. (1996). Contest Success Functions. *Economic Theory*, 7(2):283–290.

Thomson, W. (2011). Consistency and its converse: an introduction. *Review of Economic Design*, 15(4):257–291.

Tullock, G. (1980). Efficient rent seeking. In J.M. Buchanan, R. T. and Tullock, G., editors, *Toward a theory of the rent-seeking society*. College Station: Texas A&M Press.

van den Brink, R., Heb, S., and Huang, J.-P. (2017). Polluted river problems and games with a permission structure. *Games and Economic Behavior*.

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

van den Brink, R., van der Laan, G., and Vasil'ev, V. (2007). Component efficient solutions in line-graph games with applications. *Economic Theory*, 33(2):349–364.

Wang, Y. (2011). Trading water along a river. *Mathematical Social Sciences*, 61(2):124–130.

Yildirim, H. (2010). Distribution of surplus in sequential bargaining with endogenous recognition. *Public Choice*, 142(1):41–57.