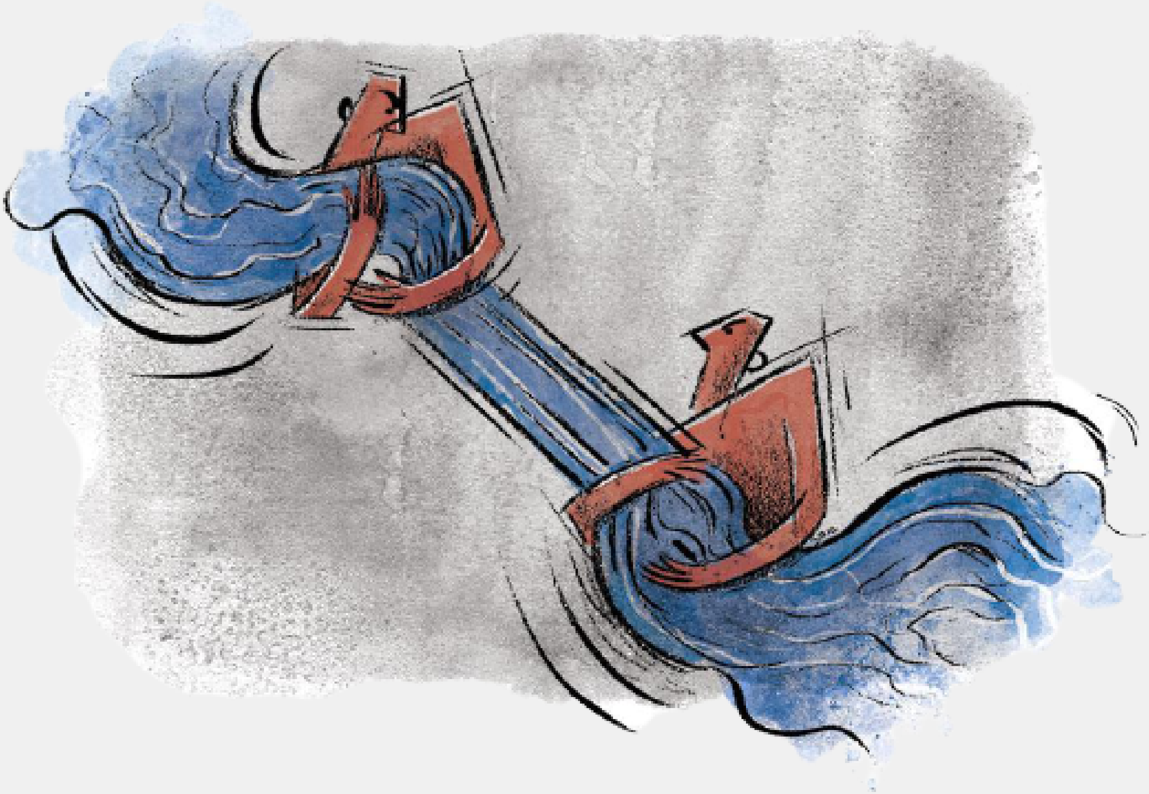
Operationalising stability and fairness in transboundary water resource allocations

by

Sahiti Sarva



Artwork by Tarique Aziz

Student number:	5137713	
Thesis committee:	Dr. ir. J. H. Kwakkel,	TU Delft, Chairperson
	Dr. J. Salazar,	TU Delft, First supervisor
	Prof. dr. mr. ir. N. Doorn,	TU Delft, Second supervisor

Operationalising stability and fairness in transboundary water resource allocations

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Science in Engineering and Policy Analysis

To be publicly defended on 24^{th} August, 2021

by

Sahiti Sarva

An electronic version of this thesis will be available at
http://repository.tudelft.nl/.

Associated code and models are available at
https://github.com/SahitiSarva/Susquehanna.git.

Acknowledgements

There is a bittersweet moment when a long, arduous journey is going to end and you wish, just for a few seconds, that it is not over yet. Being in that moment and writing this section of my thesis almost feels surreal. There is disbelief that I actually finished writing, joy that I thoroughly enjoyed the process and surprise that the thoughts of worry have finally been replaced with a deep sense of gratitude. I remember having had 'Master thesis' as a criterion of selection when I decided to pursue a masters degree. I did not know then what it would entail but I'm glad I wanted it.

Now I know that it entails starting from a blank sheet of paper and there is no other feeling in the world that is as empowering or frightening. I realized that we are the choices we make and the same applies to research. I saw that as solitary as the process may seem, you will always find help if you learn to ask the right questions. Finally, I noticed that the skill I was learning was a crucial one, the patience to read things I could not yet understand. Fortunately for me, the process of thesis writing ends with having a few moments of pure enlightenment amidst multiple moments of utter confusion and I want to use one of those moments to thank the people who made this process as rewarding as it was.

First, I thank Dr.Jan Kwakkel, for always encouraging me to think on my feet, being curious about what I had to say and noticing the little things in my thesis I would have spent hours trying to write. Thank you Prof. Neelke Doorn for being the kindest supervisor one can ever have, for taking the time to give detailed feedback and celebrating my small victories for me and finally, thank you, Dr. Jazmin Zatarain Salazar, for always being in my corner, motivating me when I needed it the most and being as excited for my results as I was. This thesis began with respect for all your work and ended with admiration for your work ethic.

Second, I thank Jin Rui Yap, my espresso shot of inspiration, for everything you taught me. Thank you, Mikhail Sirenko. These last two years would not have been the same without you. Thank you, Ignasi Cortés Arbués for proofreading my thesis and many other reports along the way. Thank you Nafisa Poonawala for being my rock and Siddhanth Janakiraman for being a breath of fresh air on stressful days. And finally, thank you Sowmya Kumar and Aarthi Sundaram for making this place feel a little more like home.

Last, I thank my family for letting go of all the phone calls I never returned and being the people who always cheer for me. You may not know what I do or why I do it but it means a lot that you continue to support your crazy sister, daughter and niece.

Executive Summary

"Water cuts across so many aspects of human existence, that it would almost be a surprise if it never led to conflict, or sometimes even violence" (Wegerich & Warner, 2010). While there is significant truth to this idea, in the last few decades, countries that are otherwise hostile towards each other have joined hands to cooperate over water sharing. This resource, useful in most of its forms and harmful in some, has left people thirsty not because of its absence but because of its poor allocation. Resource allocation has been a policy problem for many decades now. While economists argue that water must be distributed efficiently, ethicists argue for its fairness, and politicians argue for the stability of the signed treaty. This research looks to explore if an existing efficient allocation of water can also meet both the requirements of fairness and stability.

Under the premise that utility from the benefits of water is distributed among equally powerful actors in a shared river basin, this research uses the Lower Susquehanna River Basin case study to generate a set of Pareto optimal allocations and goes on to evaluate their fairness and stability. Fairness is ascertained using three ethical principles: utilitarianism, operationalized using a utilitarian social welfare function, egalitarianism, operationalized using the Gini index, and prioritarianism, using the prioritarian social welfare function. Stability is ascertained using fallback bargaining in a cooperative setting. In a non-cooperative setting, four equilibrium conditions are used. They all assume different levels of foresight, knowledge of other actors' preferences, and actors' interest to cause disimprovement to other actors.

The Pareto optimal outcomes obtained from the Susquehanna River Basin model are evaluated for stability and fairness against these definitions. These definitions allow for the outcomes to be evaluated along a spectrum of aggregate utility (utilitarian fairness metric), equality, and prioritarian utility (prioritarian fairness metric). We found that similar values of aggregate utility can be distributed in multiple ways resulting in different levels of equality and prioritarian utility for policies. Policies that balance aggregate utility, Gini index, and prioritarian utility were identified using visual analysis. These policies can be categorized as 'compromise' policies that reduced the tradeoffs between the actor utilities and increased the overall utility distributed. This is in line with the idea that cooperative benefit-sharing can result in making the 'pie' bigger so there is more utility available to distribute, creating win-win situations for everyone. The fairness metrics and tradeoffs between metrics are found to be highly dependent on changes in the value of marginal utility, priority to the worse-off, and bare minimum utility required for each actor.

In the stability evaluation, the disagreement point, or the outcome of no agreement, is varied between the current operating policy in the Susquehanna River basin and a 'get nothing' policy. In the 'get nothing' case, all Pareto optimal policies are found to be stable in a non-cooperative setting. When the Susquehanna operating policy is used as a disagreement point, no stable policy is found when all 7 players are in the game. This is attributed to a *status quo bias* among actors where they tend to prefer the existing policy over a new one. As the number of players in the game reduced, the status quo bias went down. This goes to say that unanimous agreements with as many as 7 actors can be expensive. The procedure of fallback bargaining identifies a stable policy that does not allocate the worst-case utility to any actor. This outcome does not exhibit status quo bias, is in the 95th percentile of aggregate utility, and the 75th percentile of the Gini index. Given that cooperative games assume that actors will abide by a social contract, the outcome from cooperative solution concepts is considered an ideal scenario while the outcome from non-cooperative stability is seen as the worst-case scenario.

Overall, the proposed method can be used to identify stable and fair solutions in a transboundary river basin with non-transferable utilities. The operationalization of fairness and stability in this method is limited by the choice of using the same utility function for all seven players and considering the players to be equal. It would be interesting to explore different utility functions and varying levels of influence between actors and using allocation principles other than Pareto optimality as part of future research.

Contents

List of Figures

1

Introduction

On some days we are a world of 194 independent countries and on some days we are a world of 7 billion dependent people. 261 international rivers cover almost half the globe and 109 shared river basins providing water to 40% of the global population (Wolf, 1999). Water resources are, simultaneously, borders, means of transportation, tools of trade, a necessity for survival, producers of electrical power, underpinnings of food production, and sources of risk (McCaffrey, 2001; Doorn, 2013). This omnipresence of water in human existence has historically made it a ground for conflict and cooperation (Wegerich & Warner, 2010).

The lack of access to water has seen more than 400 conflicts across the world in the last decade alone (Pacific Institute, 2019). Concurrently, there have been 50% more flood events across the world. While developments such as population growth, urbanization, and increased affluence are projected to increase global water demand, phenomena such as climate change will result in increased flood and drought conditions in the future. The discussions on water scarcity point out that the main issue is not availability but access, allocation and poor management decisions (Anand, 2007; Kijne, 2003).

The disproportionate use of water in the Global North compared to the Global South, the overutilization of water by upstream countries leaving little to downstream areas, and the unequal impacts of water-related disasters on some parts of the world over others all underscore the importance of justice in water allocation and governance (Doorn, 2019; S. G. Yalew et al., 2021). While the necessity of water justice is generally agreed upon, the mechanisms to ensure distributive justice remain complex, debated, and often difficult to operationalize.

Justice has been conceptualized as fairness by John Rawls (Rawls, 1958). He argues that firstly, all parties need to agree to a system of justice. Following this, they should check if this method of justice results in a stable society over time. In his seminal work, A Theory of Justice, he argued that "A sense of justice is more stable than another if the sense of justice that it tends to generate is stronger and more likely to override disruptive inclinations" (Rawls, 1971). The United Nations proposed that the division of water between states sharing resources must be 'equitable and reasonable' to ensure water cooperation between riparian states (SDG 6.5.2). These ideas suggest that fairness and stability are important criteria for justice in water resource allocations.

While these criteria remain theoretical in nature, operationalizing resource allocation has often been a result of market forces at the helm of which is the principle of efficiency. Welfare economists have shown that 'all other things being equal', an efficient allocation, identified through Pareto optimality, is a perfectly competitive equilibrium and is morally desirable (Hausman, McPherson, & Satz, 2016). However, Pareto optimality, albeit argued to be a necessary condition for justice, has been critiqued to be too ideal and not ensuring fairness or stability (Grasso, 2007; Cudd, 1996).

1.1. Research Gap

In literature, efficiency, stability, and fairness are three criteria operationalized to inform water allocation policies. Multi-objective optimization techniques have been used extensively to identify efficient means of resource allocation (Reed et al., 2013; Kasprzyk et al., 2016; Giuliani et al., 2014).

Game-theoretic approaches have been extensively used to determine stable allocation solutions (Madani & Dinar, 2012; Madani, 2010, 2011; Madani & Hipel, 2011; Sheikhmohammady & Madani, 2008; Brams & Kilgour, 2001). Stability is also referred to as 'accepted' solutions or implementable solutions within the literature. These studies acknowledge that stability does not warrant optimality or the other way around.

Implicitly, fairness principles are embedded within cooperative game-theoretic solutions such as the core, Shapley value or Nucleolus (De Clippel & Rozen, 2019). However, there are fewer studies that explicitly discuss the fairness of allocations. From the existing studies, utilitarianism, egalitarianism, prioritarianism, and envy-free theories of fairness have been operationalized in cases where the water resource being distributed is comparable between players (Ciullo et al., 2020; Tian et al., 2019a).

Additionally, most of the studies consider the allocation of comparable resources such as water or money among players. In shared river basins, purely due to geographic distribution, some risks are localized to some actors such as flood risk which often affects areas downstream. Similarly, the rewards also may not be transferable across borders such as hydropower revenue. In such situations, limited studies analyze the stability of outcomes (Gold, Reed, Trindade, & Characklis, 2019) and no study, to our knowledge, has evaluated fairness.

Thus, two important gaps in research were identified. Firstly, no one study accounts for stability, fairness, and efficiency considerations when identifying water resource allocation policies. Secondly, there is no single method to evaluate stability and fairness when the outcomes cannot be transferred between players.

1.2. Research Question

To study the identified research gap, the following research question is formulated:

Given a Pareto optimal solution set of non-transferable water resource allocations in a shared river basin, how can stability and fairness be evaluated?

To evaluate something, a method must be identified, and then it must be operationalized. To enable this, the research question is broken down into 4 sub-questions.

- What are the ethical principles suitable to evaluate allocation fairness?
- What are the stability concepts suitable to evaluate allocation stability?
- How can fairness principles be applied to evaluate a given water resource allocation?
- How can stability concepts be applied to evaluate a given water resource allocation?

Attention is drawn here to the difference between evaluating and deciding an allocation. The process of evaluation ascertains the fairness or stability of an existing allocation while deciding on the allocation implies identifying how the resource can be divided.

This research aims to propose a method to identify efficient, stable, and fair water resource allocation policies. The proposed method will be applied to a case study of the Susquehanna river basin. The Pareto optimal policies which will be evaluated will be identified by using a model of the Susquehanna river basin developed by Salazar, Reed, Herman, Giuliani, and Castelletti (2016).

1.3. Story line ahead

The structure of the report is as follows

- Chapter 2 consists of a brief overview of related work of authors that made novel contributions to this area of research
- Chapter 3 introduces the context within which this research operates
- Chapter 4 introduces the fairness principles and stability criteria that are considered within this study and answers the first two research questions.
- Chapter 5 introduces the Susquehanna river basin case study
- Chapter 6 elucidates on the methods used to operationalize stability and fairness criteria.
- Chapter 7 presents the results obtained from applying these methods to the case study
- Chapter 8 has a detailed discussion of the implications of the results to the research question
- Chapter 9 concludes the research and identifies possible future lines of research and addresses limitations

2

Related Work

To answer the research question, stability definitions were identified from game theory and fairness principles were identified from welfare economics. The related work within these fields and other relevant areas is briefly discussed within this chapter. The main ideas upon which this research is built are extensively discussed in Chapter 3. Therefore, this chapter is limited to the novel contributions made by authors to this area of research and studies that were found interesting.

2.1. Stability of water resource allocations

Kaveh Madani and Ariel Dinar are two authors whose work is extremely related to this research from the aspect of using game theory in water resource allocations. Madani (2010) reviews multiple transboundary water resource allocation studies that use game theory to determine resource allocation. In the 12 studies documented within his paper, solution concepts from both cooperative and non-cooperative game theory are used. Hipel et al. (2020) and Madani and Hipel (2011) define solution concepts that can be used under different assumptions of foresight, knowledge of other's preferences and disimprovement. Madani and Dinar (2012) made a novel contribution of evaluating the stability using cooperative game theory stability concepts under non-cooperative management institutions. Madani et al. (2015) used fallback bargaining to predict the outcomes of conflict in the Sacramento-San Joaquin Delta case study under uncertain conditions. Madani (2011) used Monte-Carlo game-theoretic approach to multi-criteria decision making under uncertainty, an idea similar to decision making under deep uncertainty (Marchau et al., 2019; Kwakkel et al., 2016).

Gold et al. (2019) used fallback bargaining to identify cooperative outcomes while using robustness of a policy as a measure of utility over which policymakers make their preferences. This can prove to be a novel alternative to the conventional utility metrics that are used in literature. Fu et al. (2018) used bankruptcy theory to determine disagreement points and bargaining weights for a transboundary water resource allocation problem. They go on to use principles of equity and efficiency to determine the bargaining weights. Slightly on a tangent but still relevant, the work of Chevaleyre, Endriss, Lang, and Maudet (2007) discusses the computational problems that exist in social science problems, many of which are related to resource allocation of game theory.

2.2. Fairness of water resource allocations

Tian et al. (2019a) incorporated the envy-free allocation mechanism without consideration of the stakeholder preferences. They ration out a minimum amount of water to each region as it is an essential resource and divide the rest in a way that no party can be envious of another. The study by Ciullo et al. (2020) incorporated fairness into the optimization problem itself. They used cost-benefit analysis, egalitarianism, and prioritarianism to create problem formulations and applied those constraints to obtain the allocation of risk and reward. Jafino and Kwakkel (2021) and S. Yalew, Kwakkel, Zatarain Salazar, and Doorn (2021) identified fairness criteria that can be used to operationalize fairness in the context of water resource allocation.

He et al. (2018) used non-linear weights of water supply i.e. different utilities of water at different times, converted it into a piece-wise linear weight based on the law of diminishing marginal utility and optimized water resource allocation using this utility. They attempt to share the risk of a concentrated water shortage across players. This is one of the few recent studies to incorporate non-linear utility into the problem formulation. Karamouz et al. (2011) created a system dynamics bargaining model to determine the non-linear utility of water quality and used it to determine allocations of water quality to a reservoir-river system. UNESCO and World Water Assessment Programme (2021) compiled an extensive report outlining the different methods for the valuation of water based on its different uses. This is one of the larger compilations of research on quantifying the utility from water.

Although not within the domain of water, Matthew Adler conducted a significant amount of research in applying fairness principles to climate change policies which are closely related to the work undertaken in this study (M. Adler et al., 2017; M. Adler, 2012; M. D. Adler, 2016, 2019). Outside of these, the work of Hausman et al. (2016) was found to be a good compilation of the key concepts from welfare economics and moral philosophy that informed a lot of the general definitions used within this research. Lastly, Craswell (1998) published a well-written piece outlining the fundamental ideas of welfare economics that can prove useful to clarify the vocabulary that is often used in these fields.

2.3. Efficiency of water resource allocations

While stability and fairness are the two lenses according to which this research will be conducted, the input to the method is a set of Pareto optimal alternatives. Over the years, the field of multi-objective optimization has become considerably more sophisticated and well communicated to underscore its relevance to evidence-based policymaking. Given this modeling approach forms the case study for this research, novel ideas in this field are also discussed.

The work of Gold et al. (2019), already mentioned before, also uses multi-objective optimization. The work of Salazar et al. (2016) forms the case study used in this thesis. Using the same case of the Susquehanna River basin, Giuliani et al. (2014) performs multi-objective optimization to generate a decision-analytic framework to overcome policy inertia and myopia in river basin management. Kasprzyk et al. (2016) shows how the use of multi-objective optimization battles Arrow's paradox. Doering et al. (2021) conducts interesting research on how the value of the information about inflow into a reservoir system changes concerning stakeholder preferences and evolving hydrologic conditions. They saw that the value of information depends on the objectives emphasized and can decline when physical constraints of reservoirs inhibit the use of forecasts in decision making.

2.4. Water Justice

The last area of work that also informs parts of the research and is related to it is that from water justice. Doorn (2019) put together the theory of water ethics for those just entering the field and introduces the key concepts in water ethics. Doorn (2013) talks about the key characteristics of water that pose interesting issues in its equitable allocation. S. G. Yalew et al. (2021) proposes ideas for the operationalization of distributive justice in transboundary water resource allocations.

The work of Wolf (1999) clarified the fact that there have been more instances of cooperation over water than there has been conflict, although much of the media and general narrative around water is talks about the contrary. Based on his work, the International Freshwater Treaties Database was created which continues to update information about treaties signed over water cooperation (McCracken & Wolf, 2019). Wegerich and Warner (2010) compiled a series of essays around the politics of water which serve as a fresh perspective to the mathematical world of optimization and game theory.

3
Setting the stage

The substantial, growing body of scholarship in the area of resource allocation demands any new research question to be placed in the larger context. When discussing the distributive justice of a distribution, the context is set by three main questions:

- **What** is being distributed?
- **Who** is it being distributed to?
- **How** is it being distributed?

Within this study, just allocations are considered those that are fair and stable (Rawls, 1971). Given the main research question about evaluating the stability and fairness of an allocation, it becomes imperative to deliberate, albeit briefly, upon these contextual questions. They not only set the stage for the research but also inform some of the key choices and assumptions made in the process. The main research question of this study, primarily intended to create methodological research, is nestled under the 'how' of just distribution. Answering the other two contextual questions sets the tone for the research. This chapter discusses the premise of the study.

Determining what is being distributed and to whom would take one into different worlds of literature. Figure 3.1 identifies the broad choices that can be made to narrow down the context.

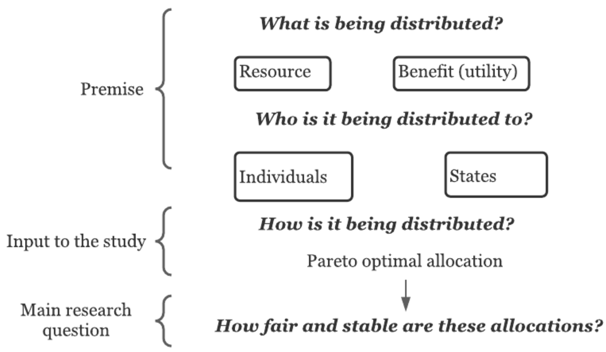


Figure 3.1: Research question in the context of distributive justice

3.1. What is being distributed?

The first aspect of resource allocation is discussing what is being distributed. A resource can be distributed directly where everyone gets a share of water or, the value from the resource can be distributed.

Water sharing agreements started with simple volume sharing dominating the treaties. These treaties evolved to have hydropower requirements, irrigation demands, and more recently, environmental regulations as factors in identifying equitable sharing agreements (Wolf, 1999; Dellapenna, 1994; Giordano et al., 2014). Here we see the operationalization of 'value' as a benefit to the consumer. The shift from direct water sharing to benefit sharing in cooperative treaties has also been backed by academic discourse as a move that creates a 'bigger pie to share', laying the foundation for equitable win-win situations among riparian states and eventually leading to stronger co-operations (Davidsen, 2010).

Creating equitable utilization of resources between riparian states is the current golden doctrine that governs international law concerning water allocations (UNWC, 2012b). Equitable allocation does not refer to equal portions of water but a fair balance of uses between all states (UNWC, 2012a). Water, serving the many purposes it does, can be a competitive resource adhering to the subtraction criteria (i.e. one's consumption subtracts from the total available to others) (Doorn, 2013). In such a situation, there are often trade-offs that must be made between different uses of water to create a fair balance of uses. Trade-offs in water allocations can come in many forms. For instance, between fisheries conservation and hydropower revenue, or environmental regulations and water supply (Villamayor-Tomas et al., 2016; Giuliani et al., 2014). Evaluating trade-offs, or identifying a 'fair' balance between uses, necessitates comparing the value of one user to another (intercomparison of utility). This requires that the value from each use of water is quantified.

Utility is a measure used to quantify the benefit or value received from consuming a resource. Given it is often difficult, if not impossible, to compare different purposes of the same resource, utility is typically a measure of the degree to which an outcome conforms to a person's preferences. If a person sees more value in one unit of water when it is consumed as hydropower compared to agricultural consumption, then the policy that generates hydropower has greater utility *because they see more value in it.* Welfare economists have historically used utility theory to justify, analyze and evaluate policy alternatives (Hausman et al., 2016). The outcome of each policy alternative will have a utility associated with it and these utility values are used to evaluate policy alternatives.

The utility from a resource can be a linear function of its use or a more complex, non-linear function. Harsanyi (1955) stated that if an individual has a set of preferences that satisfy the mathematical axioms of expected utility theory and the policies under evaluation are Pareto optimal, a linear function can be used to ascertain utility from each policy. The axioms of expected utility theory imply that the set of preferences made over a set of alternatives are that of a rational individual who does not change their preferences in whichever form they are presented, has a preference over every one of the policies presented to them, and does not change their preference order when new options are presented. In jargon-free terms, a linear utility function means that a person states that they get the same value by increasing hydropower revenue from 1 to 5 million dollars as they get if it is increased from 5 to 10 million dollars and they will not change their mind if the options are presented differently.

While water's worth is arguably infinite, recognizing, expressing, and measuring its worth in different uses and incorporating it into decision-making are important to identify equitable allocations (UNESCO & World Water Assessment Programme, 2021). Given the importance of its valuation, a linear function can arguably be too simplistic or simply not realistic for different uses of water. This can be seen in water use for agricultural practices. While too little water is dangerous, too much water supply is also counterproductive and reduces the utility obtained (Karamouz et al., 2011). For each use of water, a different utility function can be ascertained depending upon the use and the society using it.

To formalize the methodologies used for valuation of water, the literature was divided into five different perspectives by the United Nations Educational, Scientific and Cultural Organization (UNESCO) - valuing water sources, water infrastructure, water services, water as an input to production and socio-economic activity and other socio-cultural values of water. Within each perspective, quantitative valuation can be done using multiple methods such as using demand functions, linear programming (Renzetti & Dupont, 2003) and production functions (Kleinman, 1969) among others.

In this study, we use the concept of diminishing marginal utility to measure the utility that a user gets from consuming water in different forms. This allows considering utility as a linear function of outcomes or a non-linear function. The theoretical details of this are discussed in chapter 4. Unlike the comprehensive methods proposed by UNESCO and World Water Assessment Programme (2021), a common utility function is used across all use purposes of water. This sections funnels to underscore this key assumption made in the study. The assumption is placed in the context of the rest of the research in figure 3.2.

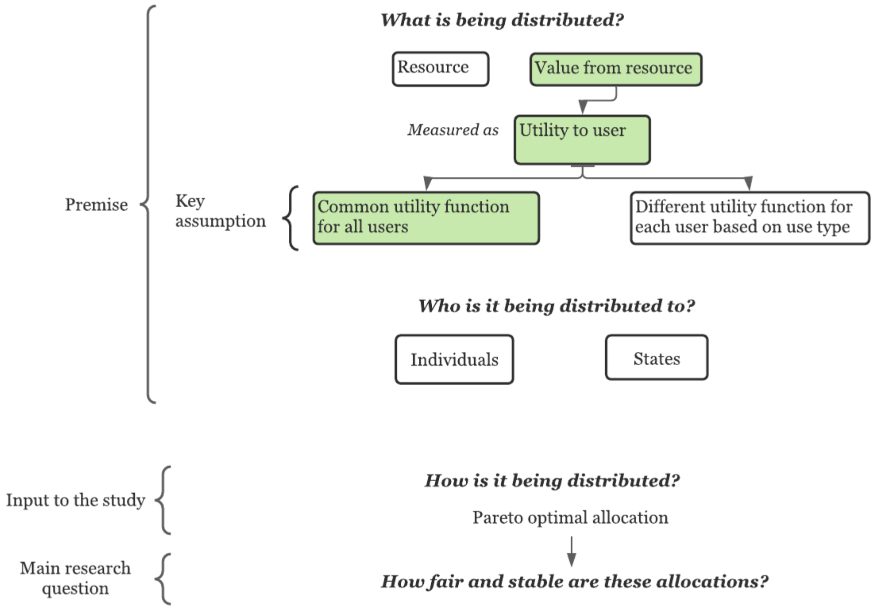


Figure 3.2: First key assumption of the research placed in context

At this point, it becomes important to note that utility theory is perched on a fundamental assumption that people are '*competent evaluators*', that is, if there are multiple policy alternatives in front of a person, if their beliefs about the consequences and properties of the different policies are true, they are good judges of what is good for them (Hausman et al., 2016). This assumption brings us to the next major question, who is the utility being distributed to or, who are the competent evaluators?

3.2. Who is the resource distributed to?

When welfare economics talks about 'people' or 'agents' as competent evaluators, they typically refer to citizens or consumers that do not necessarily need to have a direct say in the policy-making itself. From this perspective, policy fairness refers to how fair it is to *individuals*. However, when talking about transboundary water resource allocations, the idea of a *state* emerges. Two riparian states competing for a water resource implies that fairness can also be attributed to an entire nation-state as is the customary international law.

Two leading theories guide who should be considered as a recipient of a fair allocation, recognized by the United Nations: the theory of limited territorial sovereignty and the idea of a community of practice (UNWC, 2012a). The theory of limited territorial sovereignty, treated as customary international law, is the most accepted idea by states in international water-sharing agreements (Dellapenna, 1994). Rooted in Walzer's theory of morality of states, it suggests that states have a "sovereign right to exploit their resources" (Walzer, 2015). According to this theory, the idea of fairness holds within a state. Equitable and reasonable utilization is between two states and each state must refrain from causing any significant harm to another (UNWC, 2012b).

On the other hand, the idea of a community of interest is similar to the theory of cosmopolitanism, a Kantian and Stoic concept of universal human fellowship (S. G. Yalew et al., 2021). According to this idea, everyone in a shared basin has an equal right to the benefits of water irrespective of which state they belong to and the allocation must be fair to the end-user of the water resource. Given the research question is about identifying methods to evaluate fairness and stability of Pareto optimal outcomes, an important aspect of the premise is that everyone has an equal claim to benefit from the shared water resources and has agreed to find benefits from cooperation. This stands more in line with a cosmopolitan idea of water allocation.

Therefore, the water resource is being distributed to groups of individuals with different needs (different sectors) along a shared river basin, with each sector represented by one rational, competent policymaker and they all have an equal right to claim for the resource. The figure 3.3 puts this key assumption in context.

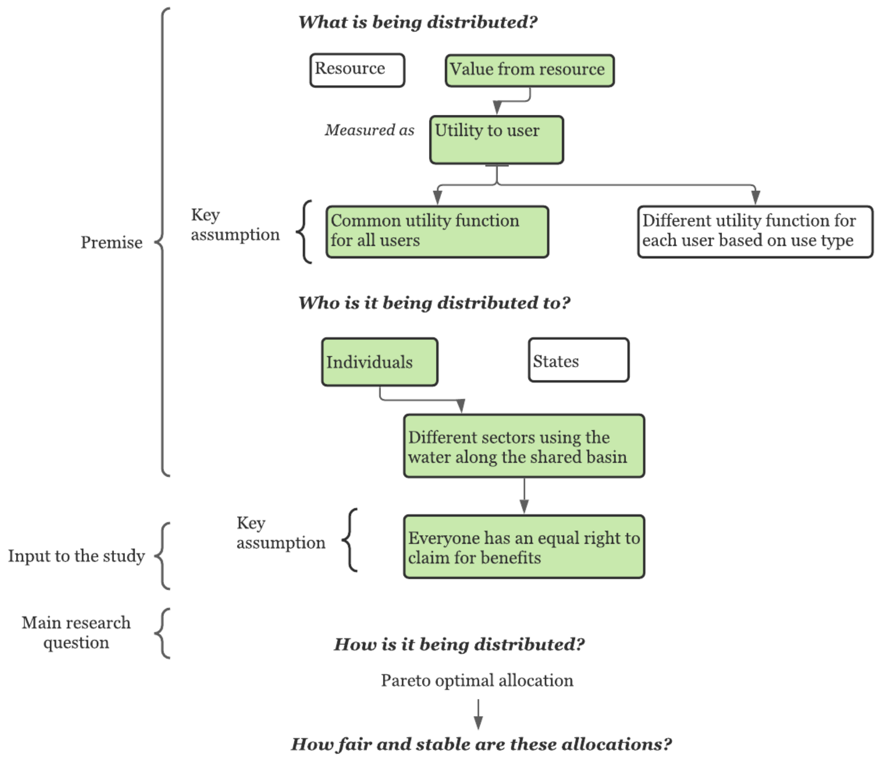


Figure 3.3: Second key assumption of the research placed in context

3.3. How is it being distributed?

At this stage, the center of this study is a shared river basin, with rational actors representing the different sectors utilizing the resource. Allocating utility to each actor can be done based on different criteria such as fairness or stability. The allocation process looks different for different criteria because it can change the allocation principle, the level at which decision making occurs, the negotiation process, and the social choice aggregation. In our case, the research question involves evaluating the fairness and stability of an existing policy. Therefore, the allocation principle remains consistent across the criteria of fairness and stability - Pareto optimality. However, the other three considerations are different when fairness and stability are being considered. As discussed in Chapter 2, game theory has been used extensively in identifying the stability of policy outcomes. In the same vein, welfare economics and moral philosophy have often informed fairness considerations in resource allocations. The two fields are compared against the three considerations in figure 3.4 and discussed here.

How is it being distributed?

Pareto optimal allocation

How stable and fair are these allocations?

Stability: Cooperative games, Non-cooperative games; Fairness: Welfare economics

	Cooperative games	*Non-cooperative games*	*Welfare economics*
Level of decision making	Decentralized among agents willing to cooperate	Decentralized among agents that do not all wish to cooperate	Centralized while considering agent preferences
Social choice aggregation	Unanimously confirm to an agreed social contract	Must unanimously agree	Social welfare functions
Negotiation process	Bargaining solutions	Individual strategies that result in stable solutions	None as it is centralized

Figure 3.4: Difference between stability and fairness evaluation in the resource allocation process

Level of decision making: This can either be centralized where one party is making a decision for everyone, by taking into account individual utility or, it can be decentralized where multiple actors come together to decide over a decision. Fairness considerations are usually agnostic to the intentions of the actors or the strategies they wish to use. This is to say that a policy is morally fair or morally *acceptable* whether or not it is *accepted* by a set of decision-makers. Therefore, when fairness is considered, the decision-making is assumed to be centralized. However, stability considerations are entirely dependent on the intentions of the actors. This is also why stable solutions can be arrived at after multiple rounds of negotiation and trading of goods (Chevaleyre et al., 2007). The intent with which actors come to the table can broadly be cooperative or non-cooperative. This influences how the different aspects of resource allocation play out and leads one to the two worlds of game-theoretic literature - cooperative games and non-cooperative games.

Negotiation process: Given decisions about fairness are centralized there is no process of negotiation associated with it. In the case of decentralized decision making, there can be a method or structure to the negotiation process that determines the final outcome. Non-cooperative game theory deals entirely with the process of negotiation. The focus is on what strategy suits the *needs* of the individual actor. The *need* can be determined based on an assumption but the most common one is that of a self-interested individual looking to enhance their utility. In cooperative game theory, players make binding agreements to adhere to the outcome of the process (social contract). Cooperative game theory then goes on to identify the conditions under which cooperative solutions can be identified and proposes 'bargaining' solutions if the parties choose to negotiate.

Social choice aggregation: Part of a larger body of work called social choice theory is the concept of choice aggregation. This is the method or function used to aggregate individual preferences or individual welfare (Sen, 1986, p. 214). When evaluating the fairness of an outcome, individual welfare must be aggregated. In this case, welfare is measured as utility given to an individual. When welfare is being aggregated, social welfare functions are used. These functions are described in detail in chapter 4. When evaluating stability, preferences need to be aggregated, in other words, if everyone is equally powerful, how many people need to agree on a policy for it to be considered stable. When individual preferences are being aggregated, there can be many kinds of choice aggregations such as majority, where a policy with the majority support is picked, plurality, where a policy with the most support is picked, or unanimity where everyone's approval is required (List, 2013). There is an inherent element of justice or fairness associated with the chosen choice of aggregation. For example, a majoritarian choice aggregation can be argued as aligned to the ideas of utilitarianism. For non-cooperative game theory, the choice aggregation is always unanimity because the actors do not wish to cooperate and each one is pursuing their interest. In a cooperative setting, the only unanimity is considered as a choice aggregation for individual preferences in this study to be able to compare it with the non-cooperative outcomes.

3.4. Research Design

The considerations described in the previous section and the guiding questions discussed in this chapter set the context around which fairness and stability can be evaluated. The research flow that follows this setting is described in figure 3.5

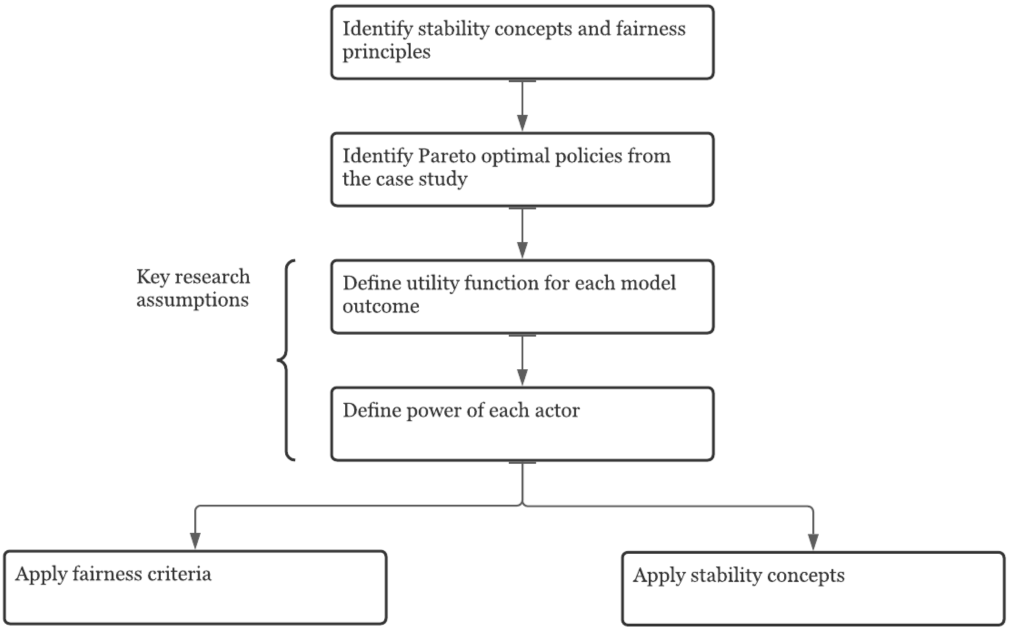


Figure 3.5: Research flow

Chapter 4 will discuss the specific stability concepts from game theory and fairness principles that will are chosen to be operationalized. Chapter 5 describes the case study on which these principles are operationalized. Finally, chapter 6 describes how the two criteria can are operationalized.

4

Theoretical Foundations: Fairness principles and stability concepts

At the beginning of this study, we discussed that the need for justice is agreed upon in transboundary river basins but its mechanisms are often debated and difficult to operationalize (S. Yalew et al., 2021). From Rawls' argument, we conceptualized justice as fairness and necessitated that a policy is more just if it results in a more stable society over time (Rawls, 1971). Operationalizing this idea is challenging for different reasons. Firstly, there are multiple ideas of fairness that are competing with each other in their conceptualization although they may agree on some underlying moral values. For example, utilitarianism and prioritarianism are consequentialist (i.e. state that fairness is in the outcome of an act but they differ in the way they are conceptualized) (Sinnott-Armstrong, 2003). Secondly, policy stability depends on multiple uncertainties such as changing public values, climate change, and in the case of transboundary rivers, changing political scenarios (UNW-DPAC, 2013; Ansink & Ruijs, 2008). It is for this reason that moral philosophers and game theorists often suggest the conditions under which a policy *could* remain stable (Verbeek & Morris, 2004). Lastly, both fairness and stability rely heavily on the context in which they are analyzed (S. Yalew et al., 2021).

To answer the larger research question of evaluating fairness and stability, it became imperative to systematically understand what are the fairness principles and stability concepts that can be used given the current context described in chapter 3 - where the utility from water is being distributed among equally powerful individuals who represent the different sectoral needs of water in a transboundary river basin.

This chapter primarily seeks to delve into the first two research questions

- What are the stability concepts suitable to evaluate allocation stability?
- What are the ethical principles suitable to evaluate allocation fairness?

To identify principles suitable for this context, three guiding criteria are used -

- **Suitability to non-transferable water sharing/ transboundary water sharing**: One of the research gaps identified within this study is to understand the fairness and stability of allocations when they cannot be transferred to other actors. Especially in transboundary river basins different benefits or risks are associated with different sides of the border and although the water can be redirected the benefit cannot be shared as-is. Additionally, this research aims to identify a method that can be applied in transboundary river basins and hence this became criteria for suitability.
- **Suitability for evaluation of a policy**: Amartya Sen distinguished the two broad purposes of aggregating utilities of different actors: deciding and evaluating. The former is to decide how to divide the water resources among a set of actors and the latter deals with what is the fairest allocation among the set of proposed allocations. As discussed before, this study seeks to evaluate

a policy in hand where the division of resources is performed using the criteria of Pareto optimality. Therefore, a principle or solution concept must be suitable for evaluation rather than deciding on how to divide the resources.

- **Level of information required**: The operationalization of fairness or stability can require specific information about the use of water in the arena or access to the actual preferences of people within the basin. Some information can be inferred from the case while some might still be inaccessible. Therefore, this became the last criteria.

4.1. Fairness principles

As briefly mentioned in chapter 3, the fairness of an allocation is evaluated using a welfare economics approach. The welfarist approach assumes that policies can be ranked on the welfare they generate to society (Sen, 1970). Welfare, in this case, is measured as utility provided to an individual actor. This process is inherently consequentialist i.e. measures the fairness based on the outcomes of a policy and does not take into account the presence of the actors themselves in the decision making, just the utility they receive.

Within this consequentialist approach, there are multiple fairness principles. From existing research in the water domain, six moral principles were identified : Envy-freeness, utilitarianism, weighted utilitarianism, egalitarianism, prioritarianism, and sufficientarianism (S. G. Yalew et al., 2021; S. Yalew et al., 2021; Jafino & Kwakkel, 2021; Doorn, 2019).

This section first discusses how the utility of water is calculated followed by a discussion on the fairness principles.

4.1.1. Utility of water

The utility of water is calculated using the theory of diminishing returns. The mathematical formulation is as shown in equation 4.1 and its corresponding graphical representation is shown in figure 4.1.

$$u(x) = \begin{cases} \frac{c_i^{1-\eta}}{1-\eta}, & \text{if } \eta! = 1 \\ log(c_i), & \text{if } \eta = 1 \end{cases} \tag{4.1}$$

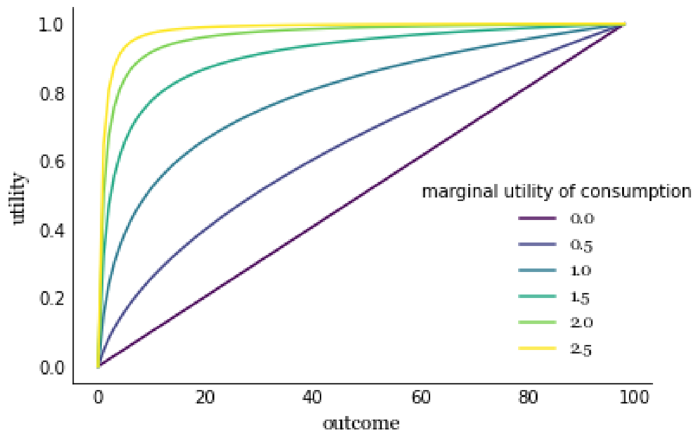


Figure 4.1: Utility curve

c_i is the consumption of water or the benefit from water by the actor i. η is the marginal utility of consumption. This is the weight given to each additional unit of resource consumed. From the figure, it is seen that when $\eta = 0$, the utility of water is linear to the consumption. This can be read as "each additional unit of resource consumed has the same value". For example, if 100 units of hydropower are being allocated to an actor, the 101st unit of hydropower adds just as much value to a user as the 1st unit of power. However, it is often not the case that the value of the 101st unit is the same as the 1st unit. In fact, it is lesser. The objective of this utility function is to capture that reducing marginal utility with each additional utility received. This function is treated as standard practice (M. Adler et al., 2017; Botzen & van den Bergh, 2014; Dasgupta, 2008).

The formulation of this utility function can have multiple positive monotonic transformations. Mathematically, this means that any positive or negative number can be added or subtracted from the utility function and its shape would remain identical. Figure 4.1 represents equation 4.1. However, if the formulation in equation 4.2 is plotted, it yields the same plot for any real value of b. This means that the utility function generates a set of preference orderings i.e. it is a relative measure where the utility of one is more in relation to another level of consumption.

$$u(x) = \begin{cases} \frac{c_i^{1-\eta}}{1-\eta} + b, & \text{if } \eta! = 1 \\ log(c_i) + b, & \text{if } \eta = 1 \end{cases} \tag{4.2}$$

It is seen that as the value of η changes, the shape of utility changes. Within climate and water literature, different studies assume values for it within the range of [0,3] (Kind et al., 2017). This value, M. D. Adler (2016), describes best, as a ”number that captures individuals' personal preferences overconsumption”.

4.1.2. Utilitarianism

Utilitarianism, an idea originally formalized by Jeremy Bentham back in the 18th century (Driver, 2009), is one of the most widely known and used theories of fairness. In this conception, it is believed that a fair policy maximizes the welfare for the most number of people irrespective of the distribution among individuals (Sen, 1970). To measure the aggregate utility a social welfare function is used (M. D. Adler, 2019). A generalized form of the social welfare function is shown in equation 4.1.

$$W = \sum_{i=1}^{n} u(x_i) \tag{4.3}$$

Aggregate welfare (W), or aggregate utility, is the sum of the utility provided by a policy for every single actor in a group of n actors. x_i is the amount of resource or benefit from the resource that an actor consumes. The fair policy is considered the one that maximizes aggregate utility. A utilitarian typically assumes decreasing marginal utility (η) i.e. they attribute a decreasing marginal utility for increasing values of consumption (M. Adler, 2012). This is achieved by using a utility function that converts the benefit or consumption of the user to utility as described in the previous section.

The most well-known operationalization of utilitarianism is Cost-Benefit analysis (CBA). CBA and the utilitarian social welfare function have been used in multiple water-related contexts such as flood risk distribution and transboundary water allocations in the Nile, the Volta and the Mekong River Basin (Kim & Glaumann, 2012; Ciullo et al., 2020; Kind et al., 2017). The idea of dividing resources using a utilitarian perspective is also referred to as the social planner method within literature (Madani & Dinar, 2012). The utilitarian social welfare function can also be used to evaluate a distribution that has been decided based on other allocation principles as well because it only requires the utility obtained by each actor after a division is performed.

An important and also the most critiqued aspect of utilitarianism is that it is agnostic to the nature of the distribution. If two actors are sharing a resource with aggregate welfare of 100, two allocations that distribute welfare as 80-20 or 50-50 are treated the same by the utilitarian social welfare function. It is non-intuitive to ascribe to an idea of fairness that does not take the differences in individual utility distributed. For this reason, multiple distributive justice principles came as a response to utilitarianism (Doorn, 2019).

One response that remains within the realms of utilitarianism is the idea of **weighted utilitarianism**. In this case, the individual utility of different actors is given weight. This weight has been determined using different methods within literature such as using the Just-Noticeable differences and relative importance of actors (Argenziano & Gilboa, 2015; Haddad, 2005). M. D. Adler (2016) suggests that η itself can act as a measure to incorporate distributional weights into utilitarian social welfare function. The functional form of weighted utilitarianism is shown in equation 4.4 where a is the weight given to actor 1 and (1-a) is the weight given to actor 2.

$$W = a(u(x_1)) + (1-a)u(x_2) \tag{4.4}$$

4.1.3. Egalitarianism

Egalitarians believe that social welfare cannot be aggregated using a single welfare function and what matters is the distribution of value between individuals. A *strict* egalitarian would side with the idea that no inequality is tolerable and all actors must get an equal distribution of utility (Ciullo et al., 2020). In general, the idea suggests that the *relative* differences in utility/well-being among individuals ought to be minimized (Doorn, 2019).

To operationalize it, i.e. to measure the level of inequality in a society, one of the most commonly used metrics is the GINI index. This index was originally conceptualized to measure income inequality in a society (Gastwirth, 1972). For the sake of this research, the Gini index is defined as equation 4.5, an operationalization borrowed from Ciullo et al. (2020).

$$G = \frac{\sum_{i=0}^{n}\sum_{j=0}^{n} abs(u_i(x_i) - u_j(x_j))}{2n^2 u_i(\bar{x_i})} \tag{4.5}$$

The numerator shows the absolute difference between the utilities of actor i and j. The denominator shows the mean value of the total utility received by all actors. Gini index (G) is the ratio of the pairwise difference in utility between the n actors to the utility everyone would have received had the distribution been perfectly equal (The mean value of utility). It is desirable to have lower values for the Gini index. A lower value implies that there is a lesser difference between the utilities given to different actors. The idea of relative difference is seen in the formulation where the difference of utility between actors is aimed to be minimized.

The egalitarian idea has two objections against it. First is the idea of a leveling-down objection. This states that egalitarianism supports the idea of reducing everyone's utility to ensure equality among all (this is operationalized by taking the absolute difference between utilities of different actors). The second is the responsibility objection, which is to say that if people are disadvantaged because of their own choices, they cannot be compensated (Doorn, 2019; Lamont, 2017).In the context of water, the responsibility objection can be seen as "if a country is using archaic agricultural practices that require a lot more water, should they be equally compensated as the others?". It was seen in a recent study performed in California that the water demand has been going down with an increase in population because of improvements in urban and agricultural efficiency, and shifts to higher-value crops and less water-intensive economic and industrial activities (Cooley, 2020). In such a case, a responsibility objection can be raised by an actor if egalitarian principles are used to identify policies.

Despite these objections, egalitarianism has been used in the context of water resource allocations because it takes into account the nature of the distribution, unlike utilitarianism (Ciullo et al., 2020). It has also been used both in deciding on the allocation and in evaluating an existing allocation. The Gini index is used for evaluation.

An idea similar to egalitarianism is the **envy-free**. This principle states that no actor must envy the utility received by another actor. A potential formulation of it is shown in equation 4.6 (Aleksandrov, Ge, & Walsh, 2019). Envy free principle requires that if an actor i, received a utility of 'x_1' and an actor j received a utility of x_2, the difference between the utility i would have got, had they been allocated x_2 must be minimized. Envy is defined as the amount of utility they would have received had they been allocated another actor's share of the water.

$$G = \frac{\sum_{i=0}^{n}\sum_{j=0}^{n} max(0, u_i(x_j) - u_i(x_i))}{\sum_{i=0}^{n}\sum_{j=0}^{n} u_i(x_j)} \tag{4.6}$$

This idea has been operationalized in settings where the division of the water resource is being determined between riparian states (Tian, Guo, Liu, Pan, & Hong, 2019b). In such a case, what is being allocated is water directly and the utility of water to one actor needs to be compared to the utility of water to another actor. However, in the case of non-transferable resources in an existing allocation, it becomes challenging to determine. For example, the envy-free index requires knowing what the utility of water equivalent to 1 unit of hydropower would be to a farmer who requires water for agricultural supply. This means reformulating the amount of water that went into producing 1 unit of hydropower, converting it into agricultural output, and calculating the utility.

4.1.4. Prioritarianism

An alternative to overcome the objections raised with egalitarianism is the idea of prioritarianism. According to this theory, allocations should prioritize the worse-off and inequalities are justified if the worse-off benefit from it. The formalization for this principle proposed by M. Adler et al. (2017) is shown in equation 4.7. Although it looks similar to the formulation used for the utility function, this is a transformation of the utility itself.

$$W = \begin{cases} \frac{(u(x_i)-u(c_{zero}))^{1-\gamma}}{1-\gamma}, & \text{if } \gamma ! = 1 \\ ln(u(x_i) - u(c_{zero})), & \text{if } \gamma = 1 \end{cases} \tag{4.7}$$

The parameter γ and c_{zero} represent *moral* or *ethical* choices that are made by the centralized decision-maker who chooses to aggregate utility and identify fair policies. γ is the priority given to the worse-off or, as M. Adler et al. (2017) described it, inequality-aversion parameter. c_{zero} can be considered as the subsistence level of consumption under which existence is seriously at risk.

The important difference between egalitarianism and prioritarianism is that the former is about reducing relative inequality while the latter is about reducing absolute inequality i.e. comparing the utility increase of an actor to their bare minimum value instead of comparing it to another actor's utility. This makes prioritarianism a principle that aggregates welfare while also being sensitive to the nature of the distribution.

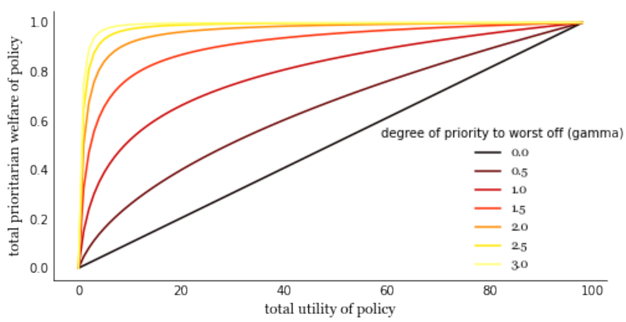


Figure 4.2: Transformed utility using prioritarianism

The visual representation of the formulation is shown in figure 4.2. The value of γ is changed while the utility obtained from c_{zero} is assumed to be zero in all the cases. The value of γ used in climate justice literature is in the range of [0,3] with a central value of 1 (M. Adler et al., 2017). The x-axis in this case is the aggregate utility of a policy. Therefore, it can be seen that when the value of γ is 0, the aggregate utility is the same prioritarian utility. In other words, prioritarianism would be the same as utilitarianism if the priority given to the worse-off and zero levels of consumption are both set to 0.

With the presence of two ethical choices, this social welfare function has been criticized to require a lot more information than the utilitarian function or the Gini index (Voorhoeve, 2014). This principle is gaining relevance in climate ethics literature, has been argued as useful in the context of water resource allocations, and has been used in a study to distribute flood risk although not for evaluation. However, the existence of a social welfare function allows for this to be used for evaluation (Ciullo et al., 2020; Doorn, 2019; M. Adler, 2012; Broome, 2008).

4.1.5. Sufficientarianism

The last principle of fairness that will be discussed is sufficientarianism. This is the idea that distributive justice aims to give *enough* of a good to a person (Doorn, 2019). Especially in the case of water, which is argued to be a human right (Gleick, 1998), sufficient amounts of water must necessarily be provided to every person to ensure survival. The idea of sufficientarianism has gained a lot of attention in the recent past in terms of its relevance to water resource allocations.

The United Nations articulates the need for water sufficiency in its Sustainable Development Goals as -

> The water supply for each person must be sufficient and continuous for personal and domestic uses. These uses ordinarily include drinking, personal sanitation, washing of clothes, food preparation, personal and household hygiene. According to the World Health Organization (WHO), between 50 and 100 liters of water per person per day are needed to ensure that most basic needs are met and few health concerns arise.

Arid countries have a water supply of fewer than 100 liters a day while countries in the west typically consume about 900 liters per person, per day (Doorn, 2019). This statement is not just meant to show inequality but to draw attention to the fact that 100 liters of water may be merely sufficient and the exact amount is still debated. Therefore, although it has been vehemently argued for in water literature, the information that is required to evaluate a policy against sufficientarianism requires specific details about the water use, the population of the area, their accessibility to water, and the amount of water lost in the process of distribution.

To conclude, 6 different fairness principles were looked at and analyzed against 3 criteria for suitability. The overview of it is shown in figure 4.3. Pareto principle is adopted as the allocation principle and hence has not been discussed. Envy-free has not been chosen for further operationalization because it is far more suited for deciding on an allocation rather than evaluating one. When used for evaluation, sufficientarianism can be used as a threshold if the information is available to a policymaker. Alternatively, it can act as the ethical consideration that informs prioritarianism (c_{zero}).

While it may be sufficient to consider one fairness principle, three were chosen to be operationalized. This is because, firstly, utilitarianism, egalitarianism, and prioritarianism have completely different formulations and capture different ideas of fairness. Secondly, different fairness principles can apply to the same context depending on the values of the people and policymakers (Jeffrey, 2018). Lastly, the use of different principles and the subsequent policies identified could also throw light on the idea of fairness that works best for the context at hand.

	Suitability to water and transboudnary river water sharing	***Suitability to evaluation of fairness***	***Information required***
Utilitarianism	Has been used extensively in literature	Social welfare function can be used	Utility obtained to every actor from an existing allocation
Egalitarianism	Has been used extensively in literature	Gini index can be used	Utility obtained to every actor from an existing allocation
Prioritarianism	Its suitability is established in water ethics while its operatioanlization in water context is still limited	Social welfare function can be used	Utility obtained to every actor , priority to be given to an actor, minimum consumption level of the actor
Sufficientarianism	Heavily supported in theory, fewer studies operationalizing it	Does not require an index or function for evaluation	Bare minimum amount of water required for different purposes to ensure human survival
Envy free	Its relevance to the water context is established	Envy free index can be used to evaluate fairness. Has been operationalized in research dealing with water allocation more than allocation evaluation	Utility obtained by every actor, utility that an actor would obtain if a different outcome is given to them

Figure 4.3: Comparing fairness principles

4.2. Stability Definitions from Game Theory

Game theory is the systematic study of interdependent rational choice. It has been extensively used to predict, explain and evaluate human behavior in situations where human actions determine the outcome of a discussion (Verbeek & Morris, 2004). In this research, human action will determine the stability of a certain policy because stability, in essence, deals with how 'accepted' a solution is. As mentioned before, the intention with which an actor comes to a discussion over policy alternatives divides game theory into cooperative games and non-cooperative games. Before delving into the suitable definitions from both fields, the use of terminology is established. Given the extensive body of scholarship in game theory, there is a plethora of terminology, interchangeably used in different contexts to define some core concepts of stability. The terminology used within the stability aspect of this research is mentioned here and the same will be used throughout the rest of the document.

Arena: Arena is the geographic area where a game is played. In this case, the decision arena is the Lower Susquehanna River Basin.
Decision Makers: Also referred to as actors, players or stakeholders are the people who eventually decide on a policy being passed in a decision arena through competition or cooperation. In this case, each objective is considered to be associated with one decision-maker.
Alternatives: Used synonymously with allocations and distributions, a solution, in this case, refers to a specific distribution of resources based on a said allocation principle.
Stable solutions: Stable solutions are those resultant outcomes that are deemed accepted by the stakeholders in an arena after playing the game.
Solution Concept or Stability definition: These are possible ways a Decision Maker might think under conflict or cooperation (Madani, 2011).
Payoffs: Payoffs are the preferences of the players over the outcomes of the game (Hausman et al., 2016)
Strategies:A Decision Maker's strategy is a complete specification of the actions the player can take whenever the player can act (Hausman et al., 2016). In this case, the player can choose between different Pareto optimal policies. Therefore, each solution is equivalent to a strategy.
Game: a single or series of interactions between decision-makers playing strategies to obtain payoffs with the knowledge that each of them has (Hausman et al., 2016).

Disagreement Point: A disagreement point in a game is the result of non-cooperation i.e. if no one agrees on the outcomes available what would be the outcome that would occur.
Unilateral improvement: A unilateral improvement for any actor is defined as a move that allows them to gain a better payoff.
State is a unique combination of strategies played by all the players.

Solution concepts from game theory are selected based on the same criteria specified at the beginning of this chapter - suitability to the context of water, suitability to the process of evaluation, and information required. While cooperative and non-cooperative game theory forms two contrasting ends of the field, they are not divided based on the presence or absence of conflict but merely by the intent of the player. A more detailed division of the field is given by Rasmusen (1989). This is shown in figure 4.4. Cooperation and non-cooperation can occur with and without conflict.

It is important here to understand what game theory considers to be conflict. When dealing with cooperation without conflict, game theory assumes that all actors have made a binding agreement to walk out with the fruits of cooperation which will be generated using an agreed solution concept such as Shapley value or nucleolus. These solutions concepts then go on to recommend how to *divide* a resource between different parties. Cooperation with conflict represents a scenario where the players, who have otherwise agreed to cooperate, choose to negotiate or bargain over the share of resource they were allocated with according to some solution concept. Non-cooperation with conflict shows a situation where multiple actors come together but strategize to make moves against each other to get the utility they want. Non-cooperation without conflict treats parties almost independent of each other and assumes they do not have to cooperate at all.

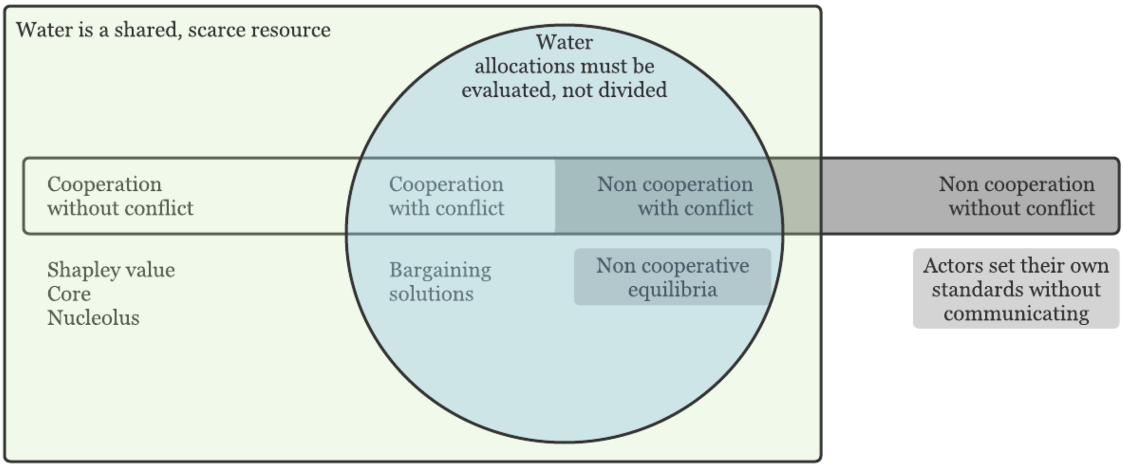


Figure 4.4: Comparing components of game theory

In this study, the resource being distributed is water. Water is a competitive resource that follows the subtraction criteria especially in shared river basins i.e. one actor using it takes away from how much another actor can use it. The competitive nature of the resource naturally creates a dependency between actors for the use of the resource. Therefore, the last situation, non-cooperation without conflict is highly unlikely to occur in a transboundary river basin. Even if actors choose to pursue unilateral movements, other actors will be affected by unilateral movements and can choose to retaliate. Which makes it a case of non-cooperation with conflict.

In the last few decades, there is a lot of media attention and political narrative around the idea of 'water wars'. Given the exacerbating issue of climate change, this idea has gained a lot more traction (Wegerich & Warner, 2010). However, the work of Aaron Wolf, which involved a review of multiple water-sharing agreements showed that there are more instances of cooperation over water than instances where riparian states make unilateral decisions that could incite deeper conflict (Wolf, 1999). This makes cooperation without conflict an interesting area of game theory to delve into.

There are two reasons cooperation without conflict is not considered suitable for this study. Firstly, the solution concepts proposed in cases with cooperation without conflict recommend how a resource can be divided and provide little insight into the stability of an existing allocation. Secondly, the existing solution concepts within this part of game theory are typically used when the goods are transferable between players. Although solution concepts do exist in cases with non-transferable utility, the use of these concepts in cases of transferable utility dominates game theory in water allocation literature (Hart, 1985; Dinar & Hogarth, 2015).

In the case of cooperation with conflict, game theory proposes bargaining solutions. This is suitable for the existing study because most water allocation agreements, even if the actors are cooperative, require frequent negotiation of arrangements even to temporarily optimize the use of water resources. This was proven true in developing the hydroelectric potential of Niagara and Columbia rivers in the United States and Canada (Dellapenna, 1994).

Finally, although non-cooperative settings have been less likely in the past, there exist situations where riparian states have taken or threatened to take or reserve the right to take unilateral action concerning water sharing. This can be seen in multiple cooperative water-sharing treaties signed between riparian states where they insist on unanimous approval to all decisions or reserve rights to making some unilateral decisions (between Spain and Portugal, or in the Indus Water treaty between India and Pakistan) (Dellapenna, 1994; Haines, 2016). Additionally, non-cooperative game theory is regarded to be far more fundamental than cooperative game theory. This is because cooperative game theory assumes that actors will adhere to binding agreements decided at the beginning of the allocation and it does not specify what kind of binding agreements make an allocation stable. Non-cooperative game theory does not make any such assumptions (Hausman et al., 2016). For this reason, the solution concepts under non-cooperative conflict are also considered suitable.

Cooperative game theory typically requires exact amounts of utility obtained by each actor and non-cooperative game theory only requires a preference order over the outcomes obtained. Given the availability of this information, the solution concepts from both these fields can be used.

4.2.1. Cooperative bargaining games

Bargaining games have also been used in the past in conjunction with ethics (Hausman et al., 2016) and bargaining solutions are created by incorporating ethical principles into their formulation. For example, there are egalitarian bargaining solutions proposed by Kalai (1977) and Myerson (1977).

The bargaining solution used in this case is a more recent concept termed fallback bargaining (Brams & Kilgour, 2001). The earliest original reference to a similar idea is the 'Kant-Rawls Compromise' proposed by Hurwicz and Sertel (1999). It is not entirely clear if this formally falls under social choice theory or game theory but it has been used to identify 'stable' or 'accepted' solutions in water resource allocation problems concerning both contexts (Sheikhmohammady & Madani, 2008; Gold et al., 2019; Madani, 2011; Madani et al., 2015; Kıbrıs & Sertel, 2007).

Fallback bargaining is a method of identifying a bargaining solution where each decision maker begins by indicating their preferences over alternatives. Then, they fall back in lockstep to less and less preferred alternatives - starting with first choices, then second and so on till a compromise is reached (Brams & Kilgour, 2001). This method seeks to minimize the maximum dissatisfaction of the players involved (Madani, 2011). It has been identified as a method that results in 'reasonable' policies being picked from a set of finite policies (Kıbrıs & Sertel, 2007).

The level of dissatisfaction is represented by the preference order specified by the player over the set of alternatives presented to them. This can easily be visualized as shown in figure 4.5. As is the case with social choice theory and game theory, multiple mathematical axioms exist to prove the existence of an outcome and the maximum iterations after which it will be found (Brams & Kilgour, 2001).

For policy relevance, it is interesting to know that fallback bargaining can be applied by incorporating different choice aggregations. When unanimity is applied, it is termed as the 'Kant-Rawl's Compromise' because every player agrees to it and it minimizes the maximum dissatisfaction. In the case of figure 4.5, the black-colored policies are not picked because they are the least preferred outcomes for both the players. However, as the number of players increases, the outcome which is not least preferred by a 'majority' of players can also be incorporated and the corresponding outcome can be picked. Additionally, fallback bargaining can return more than one outcome and they would both be considered stable.

Another variation of fallback bargaining is with an *impasse*. This means that each player can specify the preference or cardinal value below which they do not intend to bargain and an outcome can be picked accordingly. This is useful especially for water distributions because water is argued as a human right (Gleick, 1998) and providing a sufficient amount of water, especially for consumption and domestic purposes, could be a mandatory minimum that must be provided. It was seen that if fallback bargaining is performed with an impasse, or a minimum value, the worst outcome from the bargain will be at the level of the impasse (Brams & Kilgour, 2001).

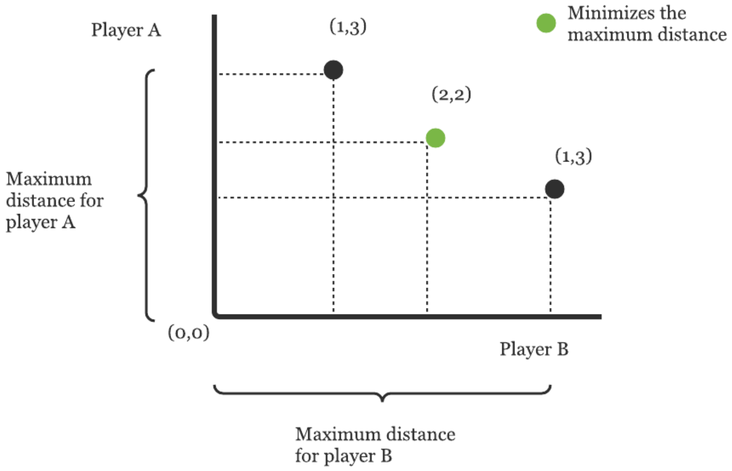


Figure 4.5: Visualizing the concept of fallback bargaining

Fallback bargaining is said to be immediately implementable if no actor has an incentive to lie about their preferences (Brams & Kilgour, 2001). It is therefore important to be in a setting where individuals are truthful about their preferences. This is the only cooperative setting in which stability is identified within this study. In future research, additional bargaining solutions can be incorporated to corroborate the results from this method to others.

4.2.2. Competitive games

Non-cooperative or competitive game theory deals with such situations where self-interested actors engage in negotiation. Stability definitions from the Graph Model for Conflict Resolution, a branch of non-cooperative game theory are used to identify competitive stable equilibria among the Pareto optimal allocations Madani and Hipel. As described in figure 3.4, non-cooperative game theory allows us to identify the types of strategy a player adopts in a negotiation. The types of strategy a player adopts depend on their available knowledge about other actors' preferences, their motivations to reduce their payoffs to hamper the progress of another actor and their level of foresight. The level of foresight is typically defined by the number of rounds of play a game has. Based on these three criteria (Madani, 2011) identified various stability definitions that are applicable to generic water resource management games. An overview of the stability definitions is presented in figure 4.6.

Solution concept	Stability description	Characteristics		
		Foresight	Disimprovement	Knowledge of preferences
Nash Stability	Decision maker cannot unilaterally move to a more preferred state	Low (1 move)	Never	Own
General Meta-Rationality (GMR)	All unilateral improvements are blocked by subsequent unilateral moves by others	Medium (2 moves)	By opponent	Own
Symmetric Meta-Rationality (SMR)	All unilateral improvements are still blocked even after possible responses by the original player	Medium (3 moves)	By opponents	Own
Sequential Stability (SEQ)	All unilateral improvements are blocked by subsequent unilateral improvements by others	Medium (2 moves)	Never	All

Figure 4.6: Stability definitions adopted from Madani & Hipel (2011)

The Nash equilibrium is one of the most widely used solution concepts in game-theoretic solutions (Madani & Hipel, 2011). However, the application of Nash equilibrium to games that are not one-shot may result in predictions that are not valid in practice (Ostrom, 1998). This is because, as can be seen in figure 4.6, it assumes no foresight, a situation where no other actor's preferences are known and no interest in harming oneself to gain benefit. The lack of foresight in decision-making receives the most amount of criticism for the application of Nash equilibrium to long-term policy decision negotiations. While the other stability definitions attempt to bridge the gap in a Nash equilibrium and have been used in water resource conflict resolution before (Nandalal & Hipel, 2007; Li, Kilgour, & Hipel, 2004), by definitions, a Nash equilibrium will also be stable in other equilibrium conditions (Madani & Hipel, 2011).

The critique towards the use of Nash equilibrium is not just because it is unrealistic, but also because by choosing the Nash equilibrium players could do worse off for themselves than if they choose to fully cooperate. it was also seen in water treaties signed where players take unilateral action that the ultimate result is unsatisfactory and optimal utilization of resources only happens by accident and not by design (Dellapenna, 1994).

Nevertheless, non-cooperative game theory remains a fundamental theory in studying human behavior under conflict and despite the obvious merits of cooperation, states and policymakers tend to make unilateral actions for individual benefit. Therefore, to account for the limitations as much as possible, other stability definitions were proposed by Kilgour, Hipel, and Fang (1987) and Fang, Hipel, and Kilgour (1989) which can also be seen in figure 4.6. While it is not always possible to map these exact assumptions of foresight and knowledge of preferences to the real world, the solution that is stable against more stability definitions can be said to be more stable (Madani & Lund, 2011b).

5

Case Study: The Susquehanna River Basin

5.1. Case Description

The Susquehanna River is the largest river lying entirely in the United States. The water spans over an area of 71,000 square kilometers, an area almost half the state of New York. It runs through New York, Pennsylvania, and Maryland. It is a source of water supply to 4.1 million people. In the Lower Susquehanna River Basin is the Conowingo Reservoir, an inter-state water body shared between Pennsylvania and Maryland. The Conowingo dam was constructed for hydropower generation purposes and is currently the largest non-federal dam in the US. This region acts as a complex multiobjective system due to competing demands between the hydropower production, water demand from Baltimore and Chester, cooling water demand of the Peach Bottom Atomic Power Station, recreation water levels, environmental regulations, and managing flood risk levels. A schematic representation of the case is shown in figure 5.1.

The Susquehanna River Basin Commission has historically used computer-aided adaptive management to manage these multi-sector needs. More recently, Giuliani et al. (2014) discussed the tradeoffs between the outcomes in the Lower Susquehanna Basin, highlighting conflicts between hydropower revenue, nuclear power cooling water, and environmental requirements. Salazar et al. (2016) used this case was also used as a benchmark case study to explore the capability of Multiobjective Evolutionary Algorithms in capturing trade-offs.

This is a transboundary competitive river basin with multi-sector, incommensurable demands. In the past, multiobjective optimization has been used in this case to identify strategies for reservoir management (Giuliani et al., 2014). This makes it a suitable benchmark case to test the use of game theory and evaluating the fairness of resource distributions given a set of Pareto optimal solutions.

5.2. Model Description

The model used to emulate this case is built upon the model used by Salazar et al. (2016). Water is pumped between the Muddy Run reservoir and the Conowingo reservoir to take full advantage of the intra-daily fluctuations in energy prices and maximize hydropower profit. During off-peak hours water is pumped up to the Muddy Run facility and during peak hours it relies on gravity-based return flows to reach Conowingo. Although direct rainfall is considered negligible evaporation losses are considered within the study to account for prolonged drought which the Susquehanna river basin is known to face. The relationships are as described below

$$s_{t+1}^{CO} = s_t^{CO} + q_t^{CO} + q_t^{CO,L} - r_{t+1}^{CO} - E_{t+1}^{CO} - q_{t+1}^{p} + r_{t+1}^{MR} - leak$$
$$s_{t+1}^{MR} = s_t^{MR} + q_t^{MR} + q_{t+1}^{p} - r_{t+1}^{MR} - E_{t+1}^{MR}$$

s^i is the volume at each reservoir (i = Conowingo Reservoir (CO), Muddy Run Reservoir (MR)), q_t^{CO} and $q_t^{CO,L}$ are the main and lateral flows from the previous time step, r_{t+1}^{CO} is the actual release,

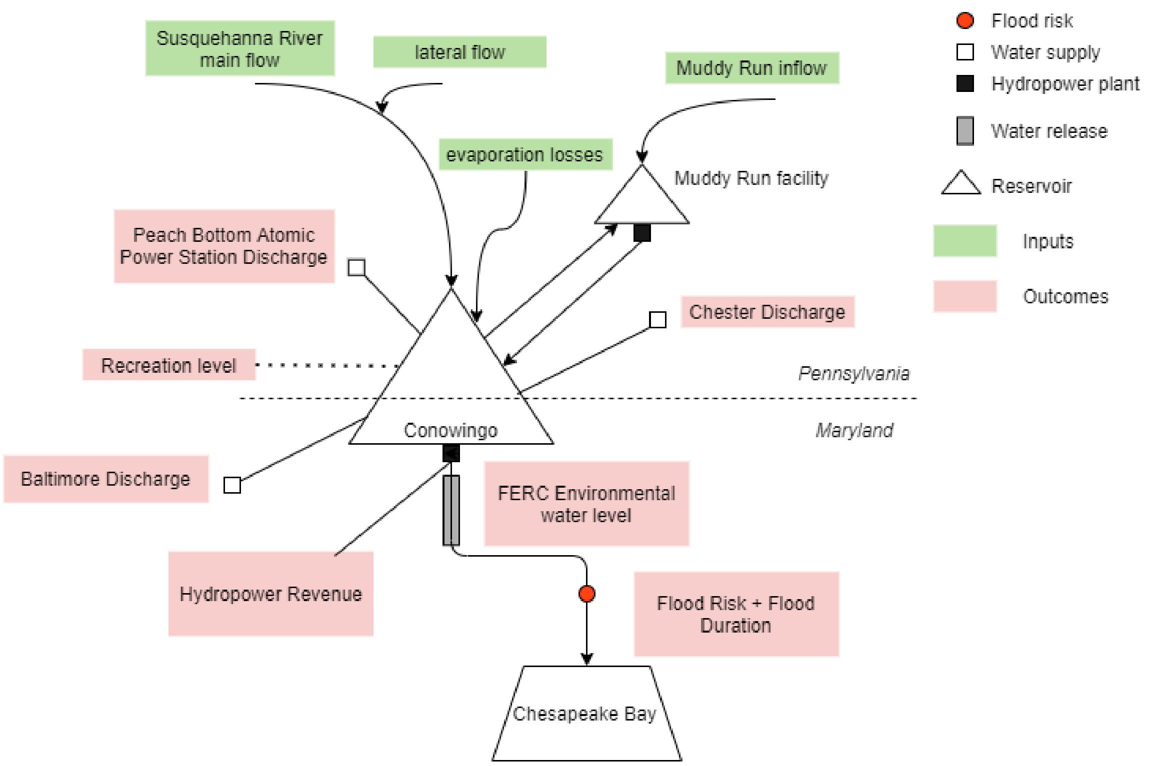


Figure 5.1: Schematic Diagram of the Lower Susquehanna River Basin

q_{t+1}^{MR} is the flow from the Muddy Run facility and E_{t+1}^{i} is evaporation losses between timestep [t+1, t]. The actual release $r_{t+1}^{CO} = f(s_t^i, u_t^i)$, where s_t^i is the storage and u_t^i is the release policy decision.

The model has 8 different outcomes described below.
Hydropower Revenue ($\$/MWh$; to be maximized) is the economic revenue obtained from the hydropower production at Conowingo.

$$J^{hyd} = \sum_{t=1}^{H} HP_t.p_t$$

$$HP_t = \eta.g.\gamma_w.deltaH.q_t^{turb}$$

η is the turbine efficiency, g is acceleration due to gravity, γ_w is the water density (1000 kg/m^3), deltaH is the net hydraulic level in metres (reservoir level - tail water level) and q_t^{turb} is the turbine flow in m^3/s.
Water demand at Baltimore, Chester and Atomic Power Plant (to be maximized) is the daily average volumetric reliability.

$$J^i = \frac{1}{H}\sum_{t=1}^{H}(\frac{Y_t^i}{D_t^i})$$

i represents the demand by the three different locations - Baltimore, Chester and Atomic Power Plant. Y_t^i is the daily water delivery in m^3 and D_t^i is the daily demand in m^3 in all the three places.
Recreation (to be maximized) is the storage reliability in weekends of the tourist season.

$$J^{SR} = 1 - \frac{n_F}{N_{we}}$$

n_F is the number of weekend days in tourist season below the target level and N_{we} is the total number of weekends in the tourist season.
Environmental Shortage (to be minimized) is the daily average shortage index relative to the FERC

flow requirements.

$$J^{SI} = \frac{1}{H}\sum_{t=1}^{H}(\frac{max(Z_t - Y_t, 0)}{Z_t})^2$$

where Y_t is the daily release and Z_t is the target release according to the FERC flow requirement both in m^3.

Flood Risk (ft ; to be minimized) is the maximum water level above critical level for the dam (109.2 ft) that the downstream areas would witness.

$$J^{Fl} = max(H, H_{Flood})$$

where H is the level in the reservoir and H_{Flood} is the critical level beyond which flooding occurs.

5.3. Pareto optimal policies

The Pareto optimal outcomes are identified using Multi-Objective Evolutionary algorithms as demonstrated by Salazar et al. (2016). The tradeoffs among the 107 optimal policies identified are shown in figure 5.2. Of all the optimal policies, the worst-case scenarios are considered within this selection. The entire analysis further will be based on the worst-case outcomes under the assumption of a low-risk appetite for all the actors.

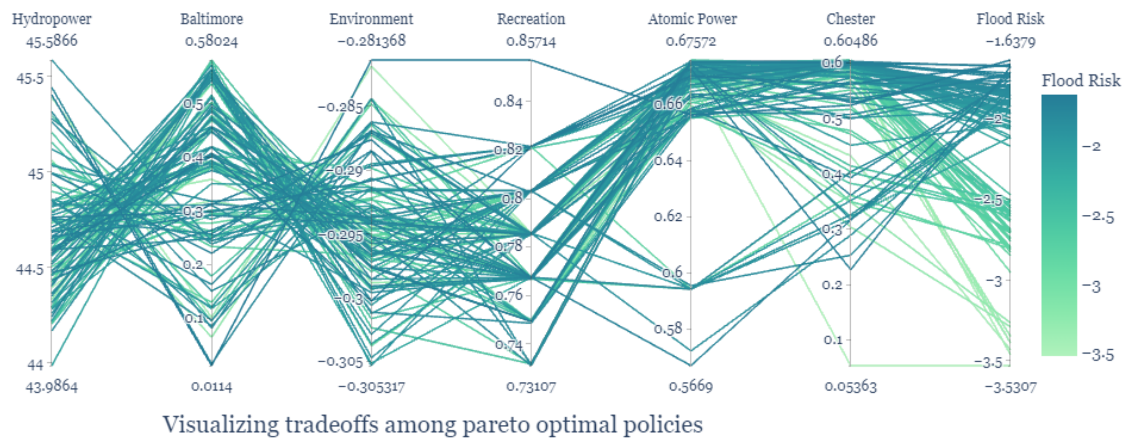


Figure 5.2: Tradeoffs among Pareto optimal outcomes (desirable direction for all outcomes is upwards)

The figure 5.2 colors the payoffs by flood risk. It can be seen that

- Hydropower revenue and environment being desirable could tend to an undesirable outcome for Baltimore, recreation, and Chester.
- There is also a slight tradeoff between atomic power discharge and Chester discharge
- Although Baltimore and Chester have the same outcome i.e. water reliability, Baltimore sees more situations with relatively low reliability.
- Atomic Power, Chester, and flood risk are generally having better outcomes than the others where there is a larger variation.

The distribution of these outcomes across policies is visualized in Appendix A. Along with that, the outcomes are mapped to more detailed descriptions in appendix A.2. The Susquehanna model was transferred to python as part of this study. Details of the improvements to code are in appendix B.1.

6

Method: Identifying stable and fair policies

This chapter discusses the other two sub research questions that deal with operationalizing the theory discussed in Chapter 3

- How can stability concepts be applied to evaluate a given water resource allocation?
- How can fairness principles be applied to evaluate a given water resource allocation?

Within the Susquehanna Case Study described in Chapter 5, each objective is assumed to be represented by one actor. Additionally, 107 Pareto optimal policies were found after a basin-wide optimization. These policies will be evaluated to determine their fairness and stability. In the process of ascertaining a method, multiple choices and assumptions are made. The choices will be tracked and explained throughout this chapter.

6.1. Utility of water

In section 6.1 we discussed the idea of a utility function that converts model outcomes to the utility that an actor receives. The model outcomes from the Susquehanna case study are mapped to the utility using the equation 4.2. In the case of the Susquehanna river basin, three choices are made when converting outcomes to utility.

- **Choice 1**. The value of marginal utility (η) is assumed to be 1.2. Within literature the value of η is typically in the range of [0,3] (M. Adler et al., 2017). The value of 1.2 is a median value given all the values assumed within literature and has been used in the context of flood risk distribution by Kind et al. (2017). This value is varied as part of a sensitivity check and the outcomes are presented in Chapter 7.
- **Choice 2**. The utility equation does not define what negative utility can mean. However, mathematically, especially when the model outcomes are between 0 and 1, the equation can tend to produce negative values ($\ln(x)<0$ if $0<x<1$). To avoid this a positive monotonic transformation of the utility function is used as shown in equation 6.1. This, although not intuitive, is purely done to maintain the utility values above 0 and does not impact the outcomes in any way, as explained in section 6.1. Such an assumption was also made and substantiated in the study conducted by Kind et al. (2017).
- **Choice 3**. Flood risk and environment are negative utilities in the case of the Susquehanna river basin. This means they do not add value but subtract value from the overall utility generated in the basin. To handle this, their utilities are treated as negative i.e. if the utility from hydropower is x, the utility for flood risk for similar consumption levels is -x. Their utility function will therefore be a mirror image of the positive utilities.

$$u(x) = \begin{cases} \frac{c_i^{1-\eta}}{1-\eta} + 20, & \text{if } \eta! = 1 \\ log(c_i) + 20, & \text{if } \eta = 1 \end{cases} \tag{6.1}$$

Figure 6.1 shows the shape of the utility curve for each of the model outcomes. Each point represents a Pareto optimal policy. The values from the utility function are compared against a simple linear transformation of the utilities. They are normalized for comparison. Notice that for larger outcomes of flood risk and environment, the utility is lower. Additionally, Baltimore shows the full extent of the curve because only Baltimore has values that are spread from almost 0 volume reliability to 60% volume reliability. The curve for the other outcomes does not change significantly because the outcome values are quite close to each other. For example, the hydropower revenue only changes between \$ 44.24 mn. to \$ 45.4 mn.

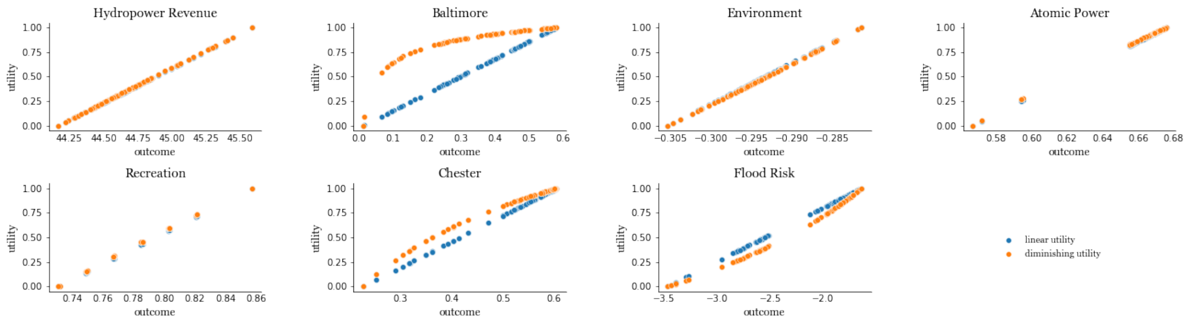


Figure 6.1: Utility curves for Susquehanna river basin model outcomes

6.2. Fairness of optimal policies

The utility functions calculated using the method discussed in the previous section are used as input to the fairness social welfare functions and Gini index. The operationalization of utilitarian social welfare function and Gini index do not require any additional information apart from the utilities and are directly plugged into the equations 4.3 and 4.5. However, the prioritarian social welfare function requires additional choices to be made to operationalize it.

- **Choice 4**. The value of γ or the priority given to the worse off is assumed to be 1, the median value used in literature (M. Adler et al., 2017). The sensitivity of the results to this choice is tested and discussed in chapter 7
- **Choice 5**. The value of c_{zero} has some additional specifications that it needs to abide by according to the chosen formulation. The formulation does not deal with negative utility values. Therefore, the difference between the utility to the actor and the utility of their bare minimum consumption must be greater than zero. This means that the bare minimum level of consumption must be lesser than the lowest value of consumption among all the policies (M. Adler et al., 2017). It is assumed that the bare minimum consumption is the worst-case scenario for every actor where they receive no utility at all. The mathematical function for prioritarianism does not handle zeroes or negative values. Therefore, the worst case is assumed to be very close to zero in the case of Baltimore, Chester, Atomic power, hydropower, and recreation. In the case of flood risk and environment, a high negative value is assumed to be the worst case. (See appendix A.3)

The outcome from the prioritarian social welfare function will henceforth be called prioritarian utility. This value represents the aggregate prioritarian utility of a policy. When the values of γ and *czero* are changed, the prioritarian utility changes. The expected shape of the prioritarian utility with change in aggregate utility is shown in section 4.1.4. By making the two choices mentioned above and comparing the aggregate utility of policies to prioritarian utility, the figure 6.2 is generated. This shows that when γ is 0 and czero is close to zero, the prioritarian utility of a policy is the same as aggregate utility, which is as it is expected to be. However, as the value of γ changes, the prioritarian utility starts to change as well. It is important to remember here that the value of czero is different for all 7 outcomes. This means that changing any of them slightly could impact the final value of prioritarian utility. The sensitivity of this is discussed briefly in chapter 7 and in detail in appendix A.3.

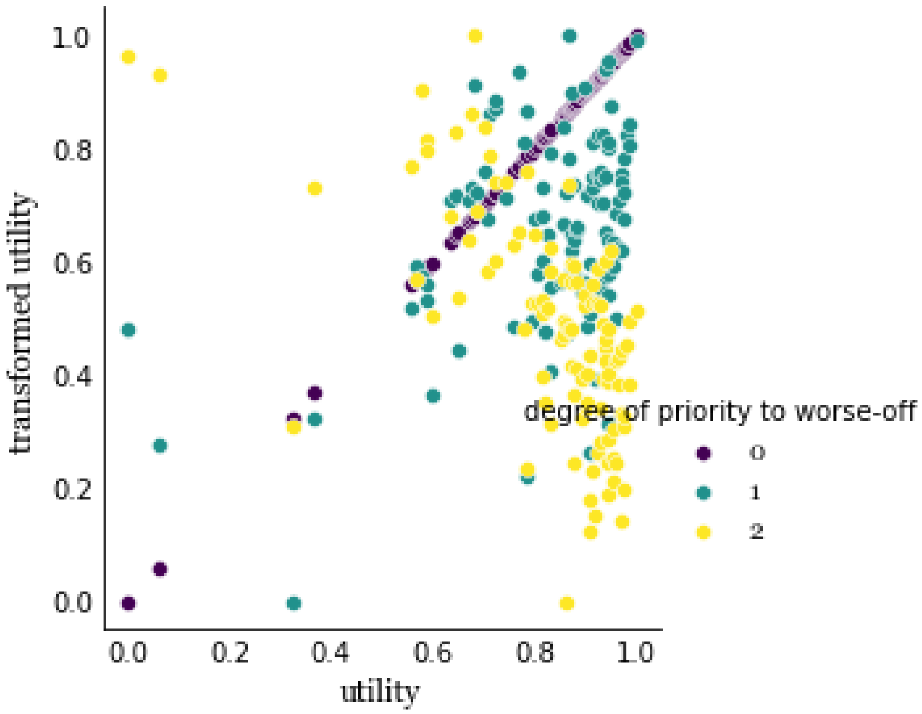


Figure 6.2: Verifying the change of prioritarian utility with change in aggregate utility and priority given to the worse-off (γ)

The three metrics operationalized to evaluate fairness are - aggregate utility (utilitarianism), Gini index (egalitarianism), and prioritarian utility (prioritarianism). They will be referred to as fairness metrics. The direction of desirability for Gini index is lower i.e. lower values are better because that signifies lesser inequality. For the other two, higher values are preferred because that indicates maximizing utility.

6.3. Stability in competitive games

This section describes the method used to evaluate the non-cooperative stability of the Pareto optimal outcomes obtained. The policies under evaluation are Pareto optimal. This means no actor can pursue an individual policy, they can either choose to agree to the optimal policy or not. Therefore, to identify the stability of solutions it is necessary to understand what will be the disagreement point because this can change the way players make their decision (Chun & Thomson, 1990). A disagreement point is essentially what the players will be left with if they do not come to a conclusion or do not agree to cooperate. For the Susquehanna Case, a baseline value is borrowed from a previous study using the same case performed by Giuliani et al. (2014). The value of the baseline can be seen in figure 6.3. This is the current baseline performance of the reservoir policy adopted by the Susquehanna River Basin Commission. This value will henceforth be termed as 'Susquehanna Status Quo'.

- **Choice 6**. The Susquehanna status quo policy does not contain values for flood risk. However, within the construction of the model outcomes, both flood risk and recreation use the same input parameters. Therefore, given recreation is zero, flood risk is assumed to be zero.
- **Choice 7**. All players are considered to be equally powerful within this study. Typically, the stability of a solution can change if there is a player with more power. Within game theory literature, the power is occasionally calculated using what one individual actor brings to the table or the grand coalition of all the actors (Madani & Dinar, 2012). However, in the distribution of

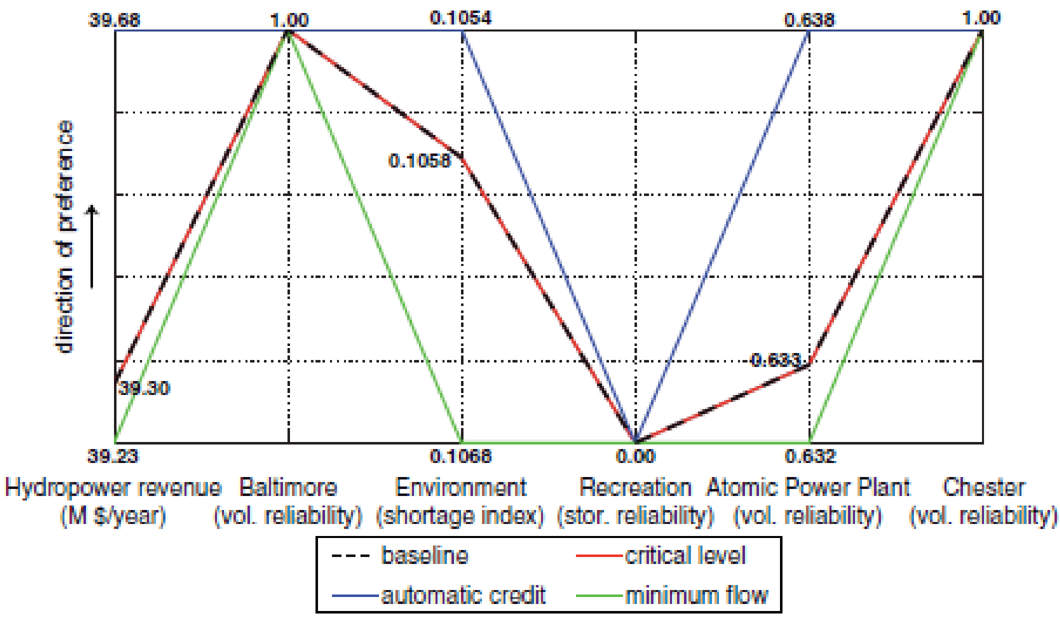


Figure 6.3: Baseline values adopted from (Giuliani et. al., 2014)

a natural resource such as water, there is no particular contribution that one player makes to the entire table. Therefore, the power of the actors needs to be determined using other means and can be done as part of future work.

- **Choice 8**. If a player is present in a discussion they must make a move or decide whether they support the policy or not. In real life, there can be dummy players who do not make a choice, or indecision itself can be a choice. However, because indecision results in the same outcome as not opting for the Pareto optimal policy, it is not considered within this study. To capture the relative influence of players, the games are conducted with and without every player being present on the table.

The method used to identify non-cooperative stable solutions is illustrated by considering a small subset of policies and 2 players - Baltimore and Chester.

Consider a set of 3 policies from the list of policies shown in figure 6.4 - policy 7,15,16. Non-cooperative game theory is typically expressed in normal form or matrix form. For this game, the normal form looks like 6.5. The columns are the *moves* that Chester can take or the policies that Chester chooses to side with. The rows are the moves that Baltimore can make. Each cell represents the outcome that will occur if Chester and Baltimore make the respective moves. Given it is a reservoir shared by multiple actors, any policy that is decided must be a collective unanimous decision (represented in green in figure 6.5. The optimized policies generated from the multi-objective optimization cannot be used if there is no agreement upon them. Therefore, in all cases where there is no agreement i.e. when they do not make the same moves, a baseline policy is used as a disagreement point. Consider the zero baseline condition seen in figure 6.4 to be the baseline.

Once a baseline is identified, the following steps were performed

- Step 1: The number of states in the game is identified (top right matrix in figure 6.5)
- Step 2: Each state is mapped to the outcome it will generate, in this case, states 2,3,4,6,7, and 8 all lead to the outcome of the baseline being picked.
- Step 3: Each outcome is mapped to the preference rank calculated from the outcomes.(bottom right matrix in figure 6.5)

	hydropower_revenue	atomic_power_plant_discharge	baltimore_discharge	chester_discharge	recreation	environment	flood_risk
policy7	26.805644	0.642111	0.253106	0.499968	0.105000	-0.258350	-0.654308
policy15	31.595089	0.358669	0.236321	0.292729	0.142143	-0.098953	-1.141051
policy16	24.307824	0.671614	0.548312	0.597803	0.122500	-0.266191	-1.408397
policy28	24.068095	0.634545	0.445319	0.545038	0.250000	-0.230026	-1.381300
policy30	28.546794	0.415784	0.575099	0.333086	0.142143	-0.152607	-0.989359
policy46	27.471527	0.523745	0.460119	0.400510	0.214286	-0.195564	-0.905152
status_quo	39.300000	0.633000	1.000000	1.000000	0.000000	0.105800	1.633000
zero_baseline	0.000000	0.000000	0.000000	0.000000	0.000000	-1.000000	-2.000000

Figure 6.4: Outcome from all policies

This ranking is then used as input to the GCMR-II.py package. This package, originally developed by Fang, Hipel, Kilgour, and Peng, was used in similar studies ascertaining stable solutions in water resource management problems (Nandalal & Hipel, 2007; Madani & Lund, 2011a). This package was modified to suit the current case i.e. create states that are possible to exist (in line with choice 8). Once the required states are ascertained, the player's preference over those states is calculated. If the outcomes are the same for two policies for an actor, the preference over it is considered to be equal. The package takes the preferences over states and generates a payoffs matrix (bottom left matrix in figure 6.5) and a reachability matrix. A reachability matrix identifies how a player can move. This is visualized in figure 6.6.

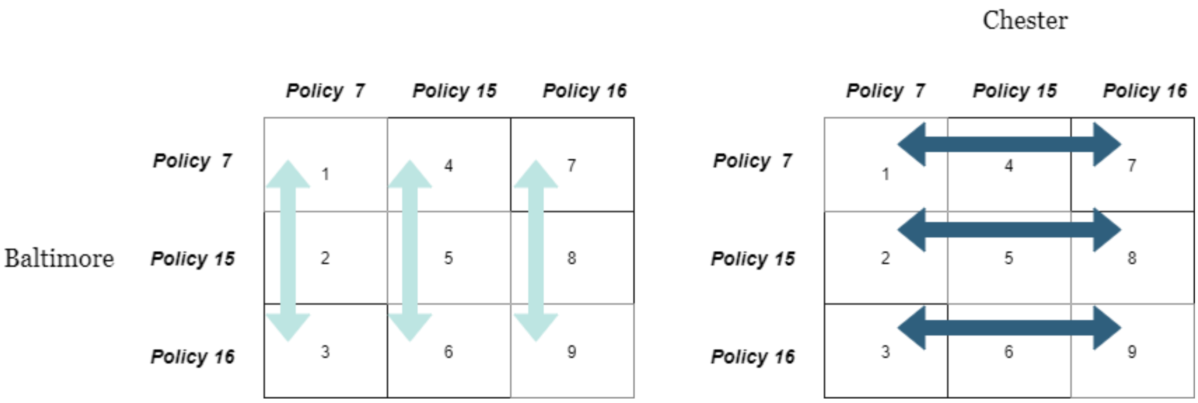


Figure 6.6: Reachability matrix for Baltimore and Chester. The light blue arrows represent the states between which Baltimore can make transitions. The dark bluw arrows show the same for Chester

It can be seen that Baltimore can only move between states 1,2 and 3, between 4,5, and 6, and between 7,8, and 9. It cannot move between state 5 and state 1 or between state 5 and state 9. This is to say that each actor can *change their strategy* but cannot force another actor to also change theirs. From a policy standpoint, this means that the idea of discussing mid-way during the negotiations may be represented in a multiple-round game but cannot be specified explicitly. Once these payoffs and reachability matrices are generated, the different stability equilibria are identified by the gmcr-II.py package and as discussed in Chapter 3. The stability under different equilibrium conditions is established, If it is stable according to all definitions it is marked as strongly stable. Else, it is considered weak stability.

Game theory, theoretically, is always predominantly as a game between 2 players and 2 decisions. However, it can be expanded to multiple players and multiple decisions. As can be seen here, with 2 players and 3 decisions, 3^2 states are generated. Therefore, expanding it directly to 116 decisions for 7 players can become computationally very expensive. To account for this, stability is determined in multiple rounds. In each round, there are a specified number of players evaluating two policies at a time. The most stable policies are then selected to have another round of play. The number of rounds is repeated until the number of stable policies starts to stabilize. This is illustrated in figure 6.7

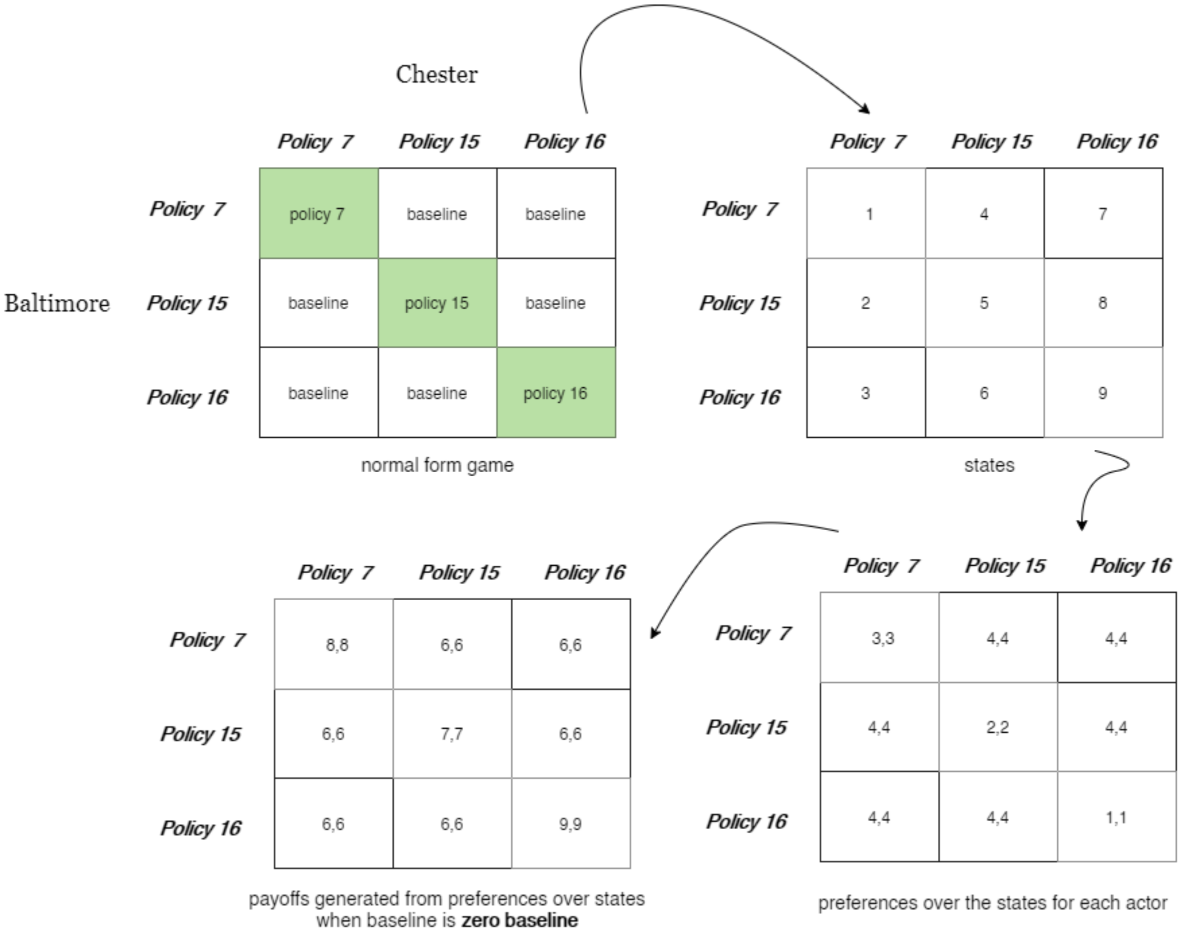


Figure 6.5: Matrix form of the Susquehanna River Basin game with Baltimore and Chester playing over three decisions

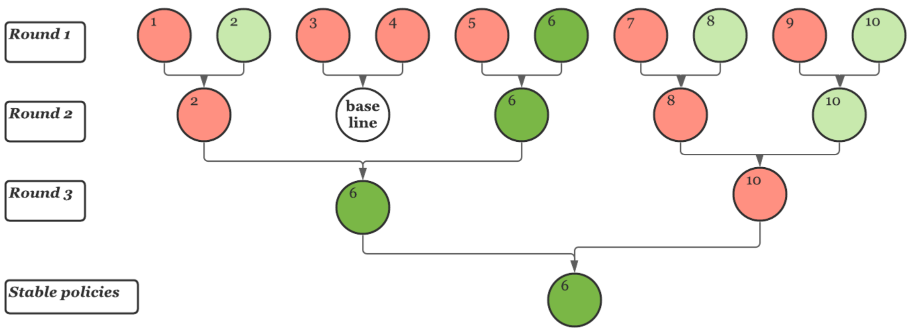


Figure 6.7: Structure of game play to determine the stability of policies in non cooperative settings (numbers represent policies, red policies are not stable and green policies are. Some policies are more stable than others which is shown by the shades of green)

This method, once set up, can be run for any number of people and any number of rounds. One limitation of the method is that it does not take into consideration the degree to which one outcome is preferred over another. This can be modified as part of future work.

6.4. Stability using fallback bargaining

To identify the stable outcome through fallback bargaining, a network approach was used. The method used to identify outcomes from fallback bargaining is elucidated here and illustrated with an example of finding the fallback bargaining outcome for 6 different policies.

- Step 1: Generate preference rankings over all the policies for every actor
- Step 2: Represent each policy as a node.
- Step 3: Specify a directed edge between policies if an actor prefers one policy to another. The edge will be sourced from the lower preferred policy to a high preferred policy.
- Step 4: Create this graph for every actor in the arena (see figure 6.8)

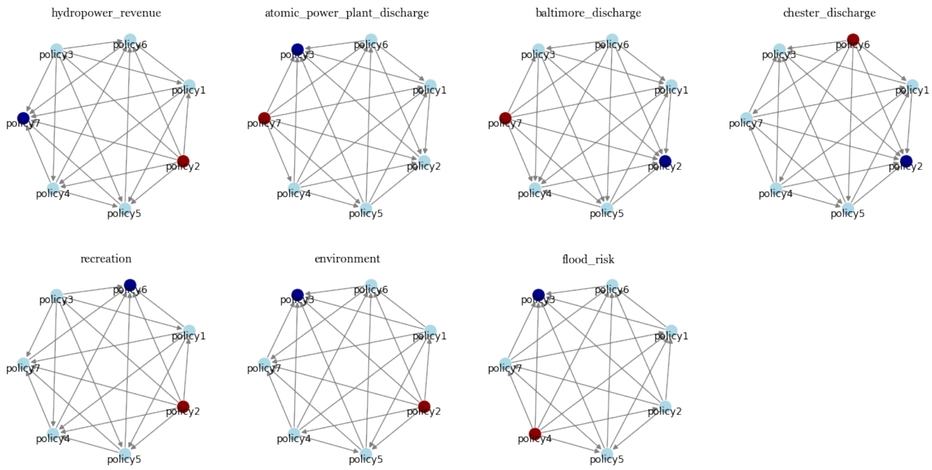


Figure 6.8: Preference directions of individual actors. Red represents policy least preferred and dark blue is the most preferred policy. The arrow is directed from the less preferred policy to the more preferred policy.

- Step 5: If a policy is least preferred by an actor, assign the 'least preferred' property of the node to a value of 1. Increment it by 1 for every actor who least prefers it. This is done to operationalize the unanimous agreement social choice rule.
- Step 6: For every node, calculate how many edges are entering it vs how many are leaving it. The number of edges entering it is the number of policies over which this policy is preferred by the actors. The number of edges leaving it is the number of policies over which this is preferred. (resultant = entering arrows - leaving arrows)
- Step 7: Find the policy with the highest resultant value and which no actor has as the least preference. The resultant value can be seen as a heuristic for collective preference. The concept of 'falling back in lockstep' as defined by (Brams & Kilgour, 2001) can be seen with lower resultant values that meet the aggregation condition.

  This process can be repeated by considering multiple baseline conditions. At step 5, other properties can be introduced to each node such as ensuring it is above a certain value. This can be used to create a situation of an 'impasse' below which bargaining cannot take place (described in detail in Chapter 3).

  Figure 6.9 shows the outcome of fallback bargaining for 7 policies considered in figure 6.8. The order of policies remains the same so in all three baseline conditions, policy 3 is the outcome from fallback bargaining. It can be seen that every actor's least preferred policy is marked in red. No actor got their least preferred policy. It is the most preferred policy for the three actors but just above the least preferred policy (6th preference) for hydropower and recreation in a no baseline condition. This exact procedure is adopted for all the 116 Pareto optimal policies.

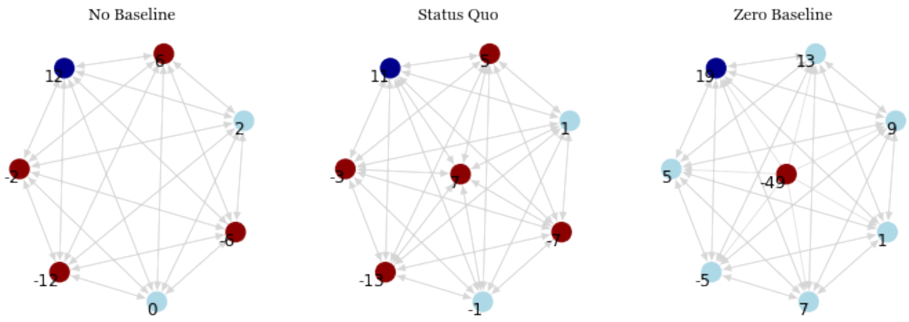


Figure 6.9: Fallback bargaining outcome for the 7 policies considered in figure 6.8. The placement of policies is the same as that in the previous figure. Red is a least preferred policy, dark blue is the outcome from fallback bargaining

At the end of operationalizing fairness and stability concepts, multiple choices were made. These choices are listed in figure 6.10. The ethical and moral choices such as the priority given to the worse-off, bare minimum level of consumption and the value of marginal utility assumed for the utility function impact the outcomes the most. The way this method pans out for the Susquehanna case and the impact of these choices is discussed in the next chapter.

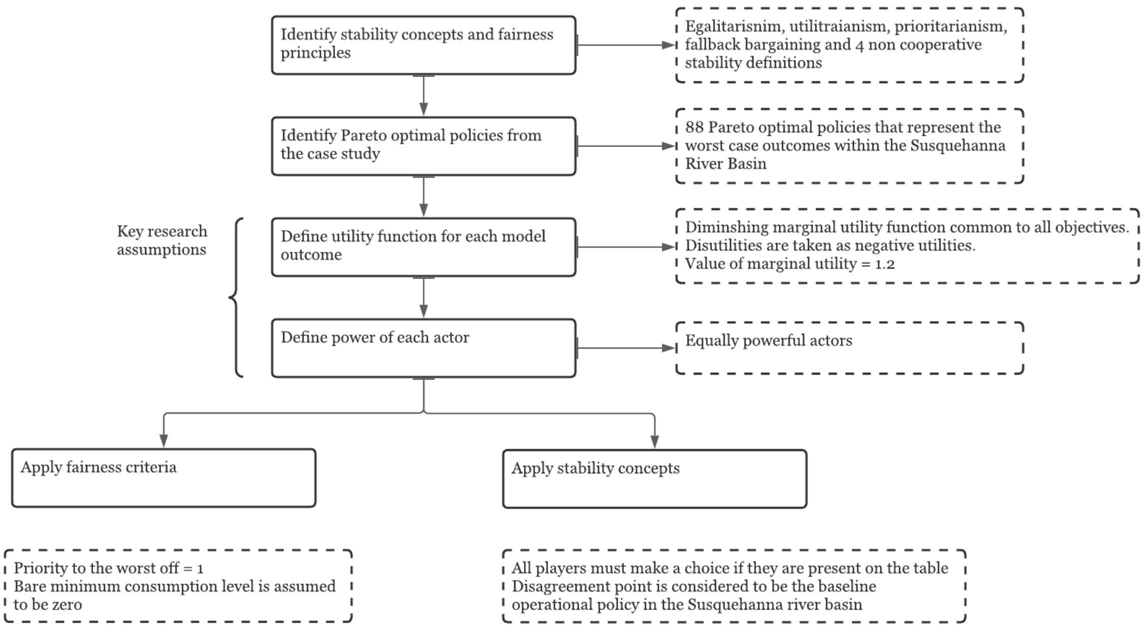


Figure 6.10: Research choices

7
Results

This research aims to identify a method to evaluate the fairness and stability of Pareto optimal outcomes. This chapter discusses the results of applying the methods proposed in Chapter 6.

7.1. Evaluating policy fairness

Fairness has been evaluated by operationalizing the ethical principles of utilitarianism, egalitarianism, and prioritarianism. The three principles are measured based on three metrics - aggregate utility, Gini index, and prioritarian utility respectively. Henceforth these metrics will be referred to as the fairness metrics. The findings from applying these metrics to the 106 Pareto optimal policies for the Susquehanna River basin are described here.

7.1.1. Findings from operationalizing ethical principles

There are fewer desirable egalitarian policies than utilitarian or prioritarian policies

Figure 7.1 shows the top 95$^{\text{th}}$ percentile policies along each fairness metric. The policies within this range are considered 'most' utilitarian, egalitarian, or prioritarian among the available set of policies. There are 17 policies in the top 95$^{\text{th}}$ percentile of aggregate utility and 13 in prioritarian utility. There is an overlap of 5 policies between the most utilitarian and prioritarian policies. On the other hand, there are only two policies in the top 95$^{\text{th}}$ percentile of the Gini index indicating that extremely equal policies are fewer.

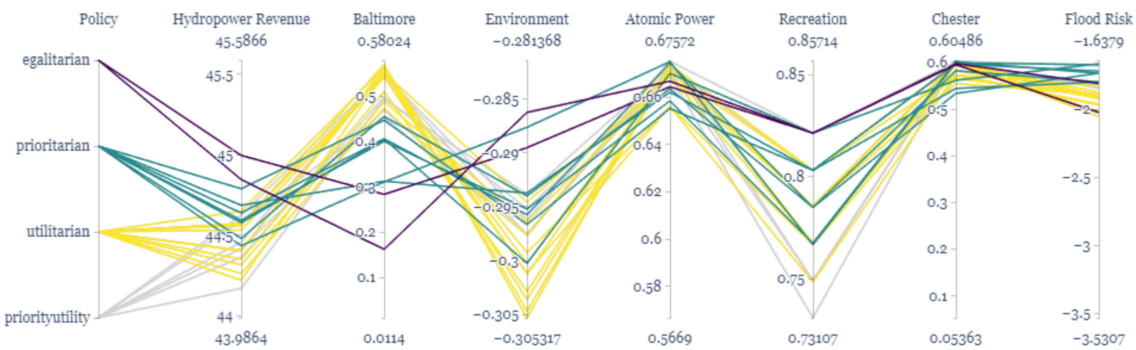


Figure 7.1: Graph shows policies in the top 95$^{\text{th}}$ percentile of utility, gini index and prioritarian utility. Plot shows tradeoffs of model outcomes between actors. 5 policies overlap between top buckets of utility and prioritarian utility, they are shown in grey.

A strictly egalitarian policy tolerates no inequality (Ciullo et al., 2020). Which means the Gini index must be zero. However, there are inherent tradeoffs within the Susquehanna river basin. Providing water to Baltimore, Chester and Atomic power take away from the water that can be released downstream for hydropower revenue, environmental shortage, or recreation. Even if there is sufficient water to meet every actor's needs, satisfying the needs of hydropower and environmental shortage means an increase in flood risk as formalized in the model described in chapter 5. This would mean that the Gini index being zero or close to zero is less likely because of the inherent physical tradeoffs unless everyone's utilities are reduced which is not the case.

However, the policies with the lowest Gini index, true to the definition of the metric, have smaller outcome tradeoffs. Aggregate utility remains agnostic to the distribution between actors, as their formulation suggests. The most prioritarian policies, excluding the shared policies with utilitarianism (green lines in figure 7.1), have smaller tradeoffs. The utilitarian policies have larger tradeoffs. This is suggestive of the fact that equal policies need not have very high utility.

There can be tradeoffs between equality, aggregate utility, and prioritarian utility. High aggregate utility can come at the cost of one actor.
To understand the tradeoffs between gini index, aggregate utility and prioritarian utility figure 7.2 is used.

Figure 7.2: Tradeoffs between aggregate utility, gini index and prioritarian utility. Values are normalized for comparison, direction of desirability is upwards

There is a tradeoff between aggregate utility and prioritarian utility, albeit small. As mentioned before, only 5 policies overlap between the top buckets of aggregate utility and prioritarian utility.
It is interesting to note that the policies with high aggregate utility have a tradeoff with the Gini index indicating that high utility can be achieved at the cost of equality. However, the grey policies in figure 7.2 indicate that high equality can be achieved without large compromises on aggregate utility. The policies with the least aggregate utility show a large tradeoff. These policies achieve equality at the cost of certain actors. Figure 7.3 shows the policy with least aggregate utility and Gini index. The policy with the least utility allocates nearly nothing to Baltimore. However, the inequality caused by this is overcome by the minimal inequality between the other actors.

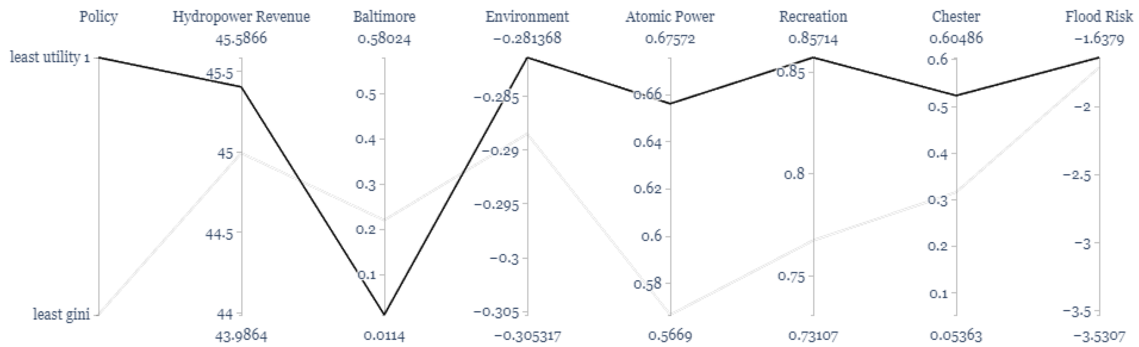


Figure 7.3: Least utilitarian and egalitarian policies.

7.1.2. Sensitivity of policy fairness to ethical choices

Policies within the 95^{th} percentile of aggregate utility, when the value of marginal utility is 1.2 and priority to the worse-off is 1, tend to be less equal than when the aggregate utility is slightly reduced as seen in figure 7.2. However, when the underlying ethical choices change, the values generated by these metrics change, thus changing the fairness of a policy according to the different metrics.

The sensitivity of the fairness metrics to three choices made is analyzed

- **Marginal utility (η)** - additional utility added to the user from one unit of water consumed. Changing this changes the individual utility allocated to each actor and thus, impacts all three fairness metrics - utility, Gini index, and prioritarian transformed utility. The default value is set to 1.2 (see Chapter 6). The range identified from literature is between 0 and 3 (M. Adler et al., 2017).
- **Priority given to the worse-off (γ)** - priority given to the actors that are worse-off. Only the prioritarian utility function uses this value. The default value is set to 1.
- **Zero consumption level (*czero*)** - value of the minimum amount of the resource the actor can get. Only the prioritarian utility function uses this value. This value is different for each actor. The default value is considered the worst-case scenario each actor (see Chapter 6).

Figure 7.4 shows how the relationship between aggregate utility, Gini index, and prioritarian utility changes with change in marginal utility (η) on columns and with priority to the worse-off (γ) on rows. The value of *czero* is maintained in the worst case.

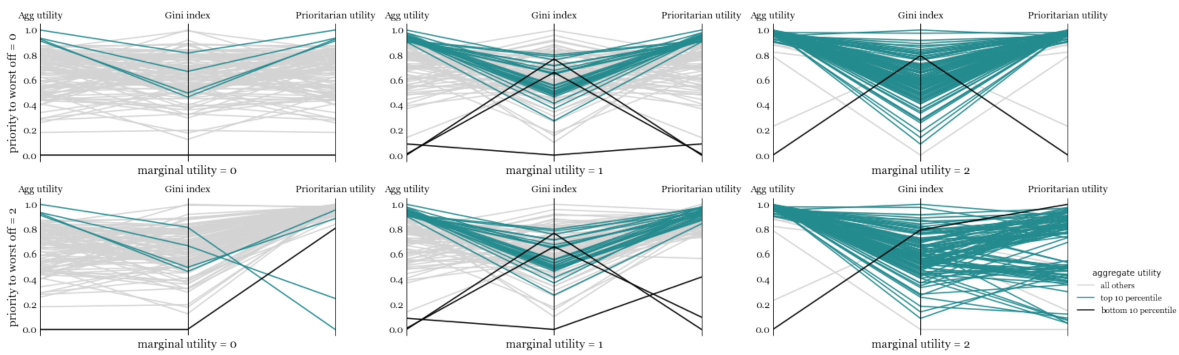


Figure 7.4: Sensitivity of tradeoffs to change in marginal utility (columns) and priority to the worse-off (rows)

Similar values of aggregate utility can be distributed in multiple ways
As η increases (column-wise comparison), there are more policies with high values of aggregate utility, and the tradeoffs are larger. These policies start to span the entire range of the Gini index, for example, when marginal utility is 2, a policy with high aggregate utility can be the most equal policy or extremely unequal among the available policies. This indicates that for similar aggregate utility, there can be many equal and unequal distributions.

Increasing bare minimum consumption level prioritizes an actor while lowering overall prioritarian utility generated
The ethical choices impact the prioritarian fairness metric the most. When γ is increased, prioritarian utility starts to weigh the utility to the worse-off higher. With a change in both η and γ, a policy with high aggregate utility need not have high prioritarian utility. The sensitivity of prioritarian utility to *czero* was tested by changing the values of *czero* individually for every actor. The results are presented in appendix A.3. It was seen that if hydropower, recreation, or environment are prioritized individually, the prioritarian policies in the 95^{th} percentile of prioritarian utility are more preferable to these actors. The number of policies in the 95^{th} percentile also goes down when these actors are prioritized showing that few policies favor these actors. When the *czero* of these actors is changed, the tradeoff with aggregate utility becomes larger indicating that the other actors (Baltimore, Chester, atomic power, and flood risk) contribute more to the aggregate utility. Additionally, this analysis shows that prioritarianism is not agnostic to the underlying distribution.

The sensitivity of policy fairness to these choices can change the number of policies that are in the top percentile of the fairness metrics. However, whatever the choices may be, it is possible to identify if there are policies that overlap between fairness metrics.

7.1.3. Identifying fair policies

At this juncture, it is established that the inherent tradeoffs within the system reflect in the fairness metrics. Additionally, the tradeoffs between the Gini index, prioritarian utility, and aggregate utility depend on the choices that are made. Under each of the potential choices, compromise policies can be identified. Compromise policies can be identified by increasing the size of the region of interest until a sufficient number of policies are found. These policies can further be examined to eliminate the ones with undesirable tradeoffs or apply other qualitative information available to a policymaker.

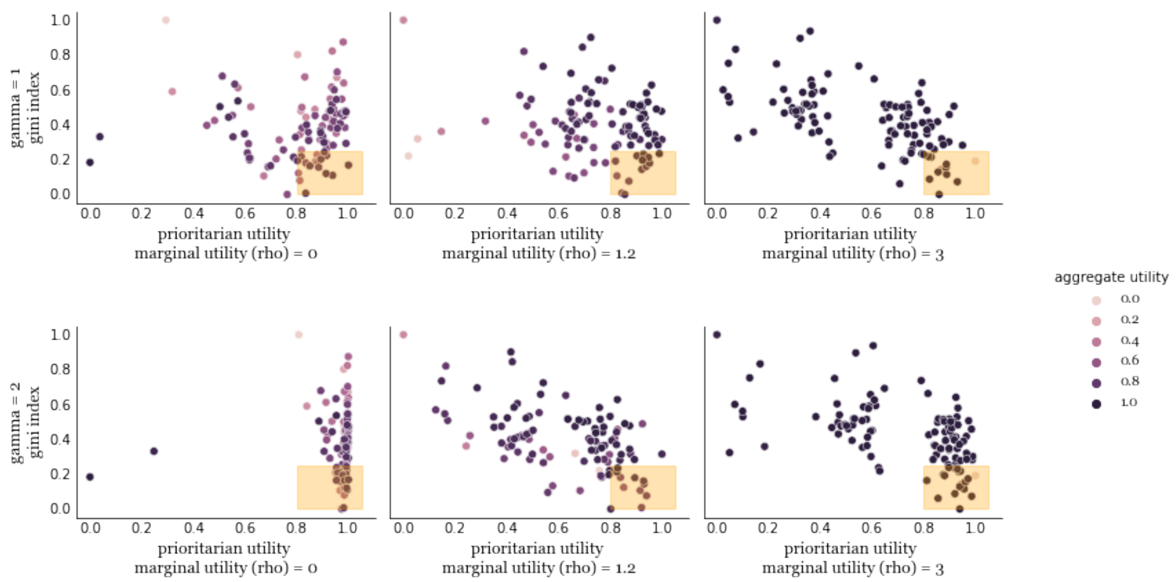


Figure 7.5: Shows prioritarian utility on x axis and gini index on y. The orange box is the region of interest. Aggregate utility is on colour.

For the ethical choices considered within this study (marginal utility = 1.2, priority to worse-off =

1), the policies that are in the region of interest (top 20th percentile of Gini index, prioritarian utility, and aggregate utility) are shown in figure 7.6

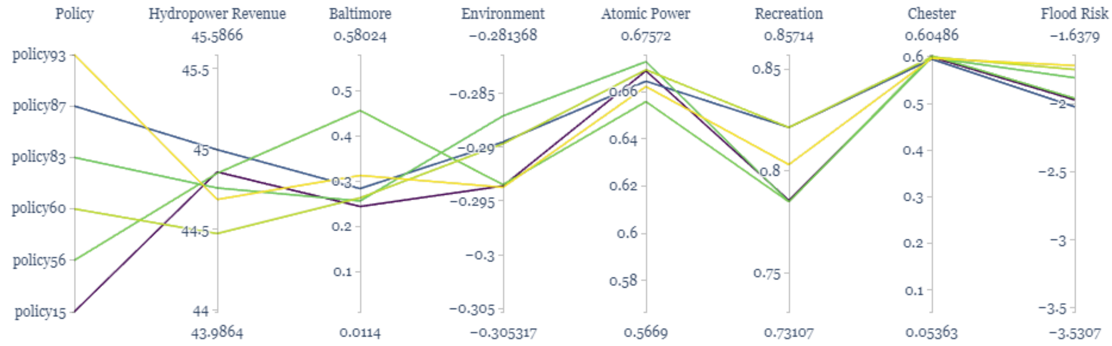


Figure 7.6: Policies in the region of interest when the marginal utility is 1.2 and priority to the worse-off is 1

If the ethical choices change, a different set of policies will be considered stable.

7.2. Evaluating policy stability

The stability of policies was evaluated by using stability concepts from Game Theory. Stability was evaluated in both cooperative and non-cooperative settings. Non-cooperative stability definitions answer the question - 'Is there a stable policy among the Pareto optimal policy?'. On the other hand, cooperative fallback bargaining answers the question - 'What is the most stable policy among the Pareto optimal policies?'. This subtle but crucial difference implies that non-cooperative settings do not guarantee the existence of a stable solution but cooperative solution concepts do. Cooperative stability identifies the policy that is more likely to be stable if actors agree to certain conditions while non-cooperative stability definitions identify the stable policy assuming that each actor wants to pursue their interests.

For this reason, the result from non-cooperative stability definitions is the presence (or the lack thereof) of stable policies. The result from fallback bargaining is the most stable outcome.

7.2.1. Findings from applying non-cooperative stability definitions

Identifying stability requires the clarification of two game aspects

- **Decision Makers**: Who is on the table to make the decision about stability?
- **Status quo/disagreement point**: What will the players be left with if they do not come to a consensus?

Stable solutions in a non-cooperative setting are identified across 6 disagreement points and all different combinations of decision-makers at the table. The disagreement points considered are shown in figure 7.7.

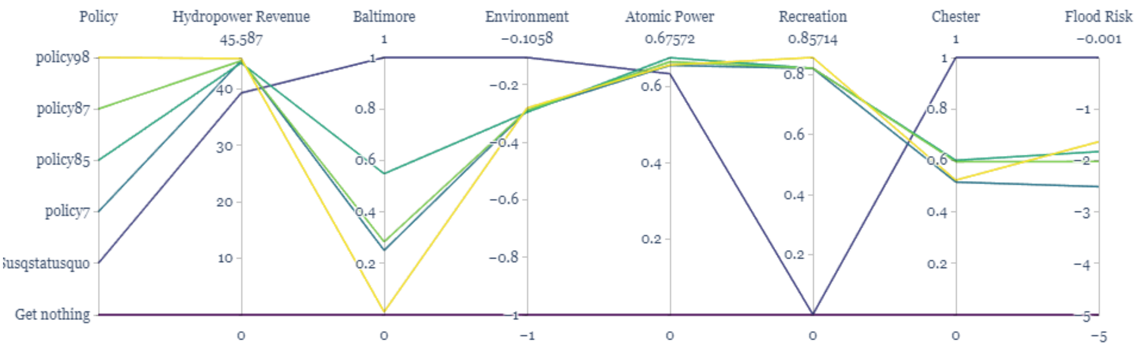


Figure 7.7: Disagreement points considered for stability evaluation

Policy7 and policy98 are the most egalitarian and utilitarian policies respectively when a linear utility function is considered. Policy87 and policy85 are the most egalitarian and utilitarian policies respectively when a diminishing utility function (η=1.2) is considered. Susquehanna's status quo is the current operating policy in the river basin borrowed from Giuliani et al. (2014). Across these disagreement points, competitive and cooperative stability of policies is identified.

Competitive stability of a policy is prone to status quo bias

Within the Susquehanna river basin, it is assumed that the only way a Pareto optimal policy is accepted by all competitive players is if everyone agrees on a policy i.e. a policy is stable if it is unanimously agreed upon. If a policy is not unanimously agreed upon, the status quo (or disagreement point) is chosen as the only stable policy. This means when 7 actors are deciding on 3 policies, there are only 3 scenarios in which a consensus is reached. On the other hand, there are $7^3 - 3$ scenarios where the status quo remains in place.

Using the Graph Model for Conflict Resolution, it was found that under a get nothing status quo i.e. a policy where every actor gets nothing, all the Pareto optimal policies are stable. However, if any of the Pareto optimal policies or the current Susquehanna status quo is used as a disagreement point, the disagreement point itself emerges as the only stable policy. This is shown in figure 7.8. This figure indicates the presence of a status quo bias. The status quo bias is a preference for the current state that biases the actor from changing their stance (Kahneman, Knetsch, & Thaler, 1991).

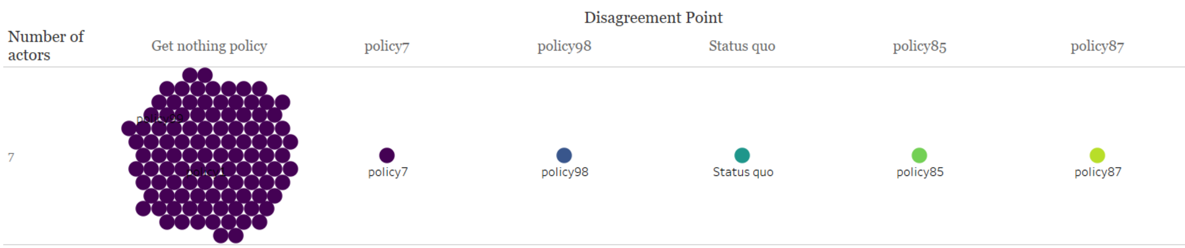


Figure 7.8: Shows the number of stable solutions under different disagreement points when there are 7 decision makers.

If there is a status quo/disagreement point that benefits even a few decision-makers, the status quo is likely to be more stable than the policies under consideration. In other words, it can be expensive to get 7 people to agree on a new policy if the existing policy already benefits a few decision-makers. To understand this behavior better, stable policies were generated for all combinations of players being present on the table.

Status quo bias reduces with a reduction in the number of decision makers

As the number of decision-makers on the table goes down, the status quo bias reduces. This is because it is more likely that a policy satisfies fewer players, especially when there are inherent tradeoffs like in the Susquehanna basin. This is shown in figure 7.9. The data behind figure 7.9 is generated but applying the Graph Model for Conflict Resolution to different player combinations. On the rows, the figure shows the number of actors present on the table. For example, 6 actors making the decision implies stable policies from all games where 6 of the 7 actors are involved.

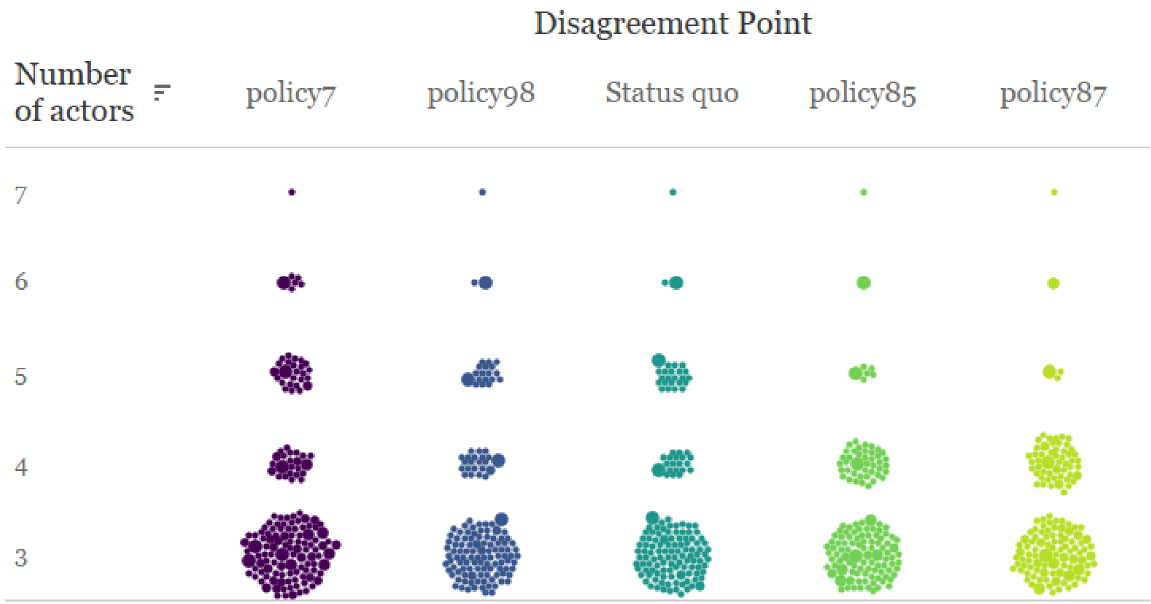


Figure 7.9: Shows the number of stable solutions under different disagreement points with different number of decision makers. Size shows the degree of stability of a policy.

The size of the circles represents strong or weak stability. It was seen that a policy is either stable in one condition (the General Metarationality which is a stability definition under 2 moves of foresight, no knowledge of opponent's preferences, and the chance of opponents to choose their disimprovement), or it is stable in all the conditions. This is in line with the way the equilibria are related to each other i.e. if it is a Nash equilibrium it will likely be an equilibrium in all other conditions (Madani & Hipel, 2011). The columns are the disagreement points. Under the get nothing policy, all policies are stable and hence are not shown in the figure.

The choice of disagreement point influences the number of stable policies and their degree of stability. However, the number of players on the table has a larger influence over the existence of stable policies. For example, when only Chester, Atomic power, and environment are involved in the decision making, policy32 and policy3 are considered highly stable when the status quo is policy87. These policies are not stable when other actors start to get involved. These policies are shown in figure 7.10.

7.2.2. Findings from using fallback bargaining

As mentioned earlier in this section, cooperative stability mainly answers the question 'how stable are the policies under consideration'. Therefore, the outcome from fallback bargaining is not whether there are stable solutions or not (like in non-cooperative games), but the level of stability of a policy measured as fallback rank or fallback resultant. To see if the fallback ranks change considerably upon the introduction of new policies, two disagreement points are considered while performing fallback bargaining - Get nothing policy and the Susquehanna status quo.

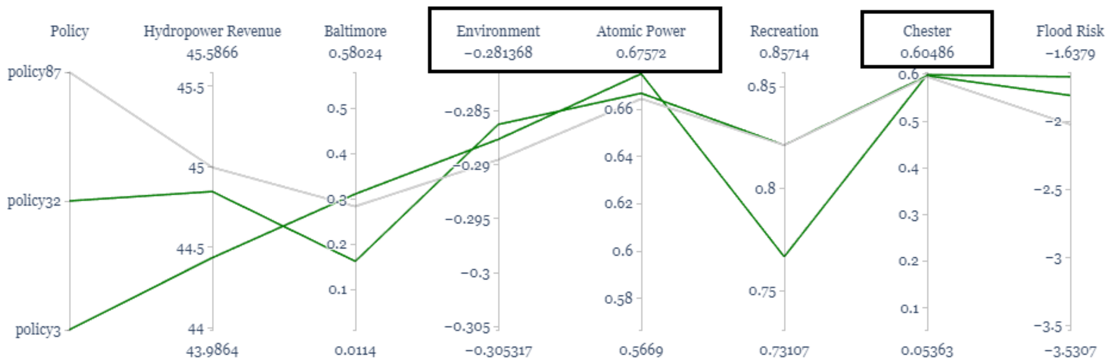


Figure 7.10: Stable solutions under the status quo of policy 87 when Chester, Atomic power and environment are the sole decision-makers.

Outcome from fallback bargaining remains stable when new disagreement points are introduced

The fallback ranking of policies is plotted in figure 7.11. The x-axis shows how preferred a certain policy is. All policies with a negative preference are undesirable. The desirable policies are divided into two - the top 95$^{\text{th}}$ percentile and the rest. The colors are determined by the rank of the policy when no disagreement point is considered. The same colors are used for the other two scenarios. A shift in the colors of the policies indicates a movement of policies from being desirable to moderately desirable or undesirable.

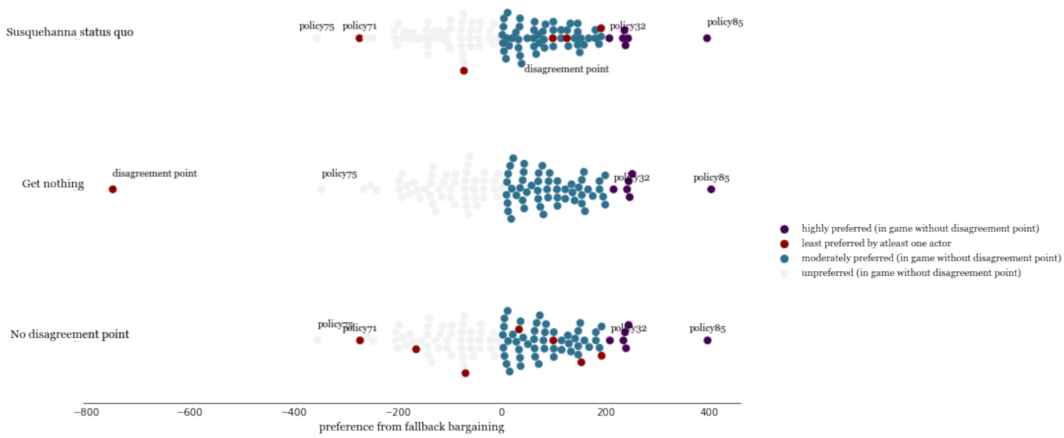


Figure 7.11: Stability ranks of policies under different disagreement points. The colours are the desirable and undesirable policy ranks calculated under the 'No disagreement point' scenario. Movement in coloured policies signifies instability of fallback rank.

When a 'Get nothing' policy is introduced, the least preferred policy for all actors becomes the disagreement point. The remaining ranks are fairly unchanged. When the Susquehanna status quo policy is introduced, there is movement in the policy ranks albeit very small. The disagreement point itself is least preferred by some actors and hence cannot be the best outcome from fallback bargaining. The top 2 policies remain the same in all scenarios.

The top 2 and bottom 2 from fallback bargaining are shown in figure 7.12. Policy85 significantly outranks the other policies. The Gini index of policy85 lies in the 75^{th} percentile, the aggregate utility lies in the 95^{th} percentile and the prioritarian utility lies in the 85^{th} percentile indicating that it has a high aggregate and prioritarian utility but is not as equal as other policies.

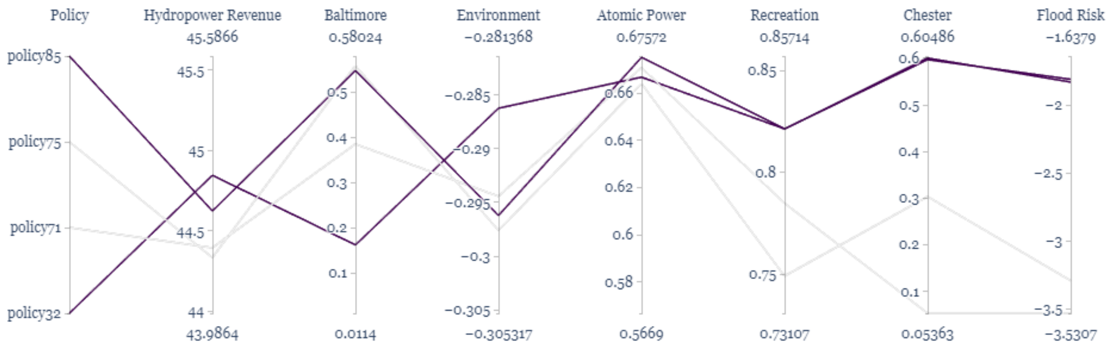


Figure 7.12: Best and worst outcomes from fallback bargaining

8
Discussion

This chapter discusses the implications of the results presented in Chapter 7. The aim of the study (main research question) is to identify a method to evaluate the stability and fairness of non-transferable water resource allocations in a transboundary river basin. In the previous chapters, a method was identified and applied to the Susquehanna river basin. In this chapter, the findings are placed in the larger context of resource allocation and distributive justice in transboundary river basins. In the process, we discuss how the proposed method can be used, its core components, and key limitations.

8.1. The problem of incommensurability

One aspect of the research question is the allocation of 'non-transferable water resources'. This follows from one of the research gaps identified in this study - a dearth of methods that evaluate stability and fairness when the allocated goods are non-transferable. This dearth is often attributed to the problem welfare economists refer to as 'incommensurability', for which there is no one standard definition. The definition given by Craswell (1998) was found to best explain this concept. Two options are said to be incommensurable if no scale or metric would rank one option higher than another and more importantly if there is no reason to *justify* the choice of whichever option is ranked higher. For example, in the Susquehanna case, flood risk and environmental shortage can be called incommensurable if there is no metric or justification to rank a policy reducing flood risk higher than one that reduces environmental shortage.

While there is much truth to the idea that two options like flood risk and environmental shortage can be incommensurable, to evaluate fairness and stability, or simply to incorporate the tradeoffs between different uses of water into policy formulation, the different values need to be reconciled (UNESCO & World Water Assessment Programme, 2021). In this method, non-transferable goods are reconciled using utility functions. A utility function can be used to bypass incommensurability and provide a common scale for comparison if the functions of different water uses are specified appropriately. It is seen in section 7.1.2 that changing the utility function can impact the fairness of policies significantly.

In this study, we use the same utility function for all 7 uses of water. This choice is a simplification. Water valuation literature highlights two aspects to ascertain the value of different uses of water. First, each type of water use must be valued differently (UNESCO & World Water Assessment Programme, 2021). This means a utility function for agricultural use of water can be completely different from that for flood risk. This can be achieved by appropriately changing the value of marginal utility in the existing function or using a modified function. The second aspect of valuation is that it can be measured using different kinds of information such as volume reliability and water quality assessment by tracking total dissolved solids (TDS) concentration in water or accessibility to water (Karamouz et al., 2011). In this study, we use the model outcomes as the raw information upon which a utility function is constructed. If it is important to consider water quality, accessibility, or other metrics, they are best incorporated into the problem formulation.

It is important to mention here that the utility function is a common metric or scale in Craswell's definition of incommensurability. It acts as a common scale for comparison but does not provide any justification for ranking one value over another. This is to say that a utility function can purport that 100 units of hydropower revenue are worth more than 10 units of water supply, but it does not suggest that the 100 units of hydropower must be the chosen policy. This justification can be given by the chosen fairness principles, stability concepts, or simply the judgment of the policymaker. Such a justification becomes extremely difficult when dealing with irreversible losses. Doorn (2018) defines an irreversible loss as "a person or group of people suffering from the loss of an incommensurable good that cannot be substituted or compensated by an equivalent of the same good elsewhere." This brings us to the second simplification in this study - treating losses merely as negative utilities.

Within this study, environmental shortage and flood risk are treated as negative utilities. This means gaining rewards is compared with minimizing losses. If the loss is irreversible, as can be the case with environmental shortage, no utility function can quantify the value of the loss and the comparison becomes meaningless. To overcome this limitation, in future work, losses can be identified as irreversible and reversible (Doorn, 2018). Thresholds can be identified for the irreversible loss and only beyond this threshold can a comparison be permitted between risks and rewards.

8.2. Reflecting on the fairness evaluation method

While utility functions form the premise of the study, the method to evaluate fairness is directly tied to the research question. Within the method, three fairness principles were operationalized: utilitarianism, egalitarianism, and prioritarianism using three fairness metrics. They were considered suitable because of their relevance to the context of water and their ability to evaluate an existing allocation. The operationalization of fairness using the three fairness metrics showed results that are in line with the core idea behind the fairness principles. Aggregate utility was agnostic to the distribution, Gini index was entirely sensitive to it and prioritarian utility aggregated the utility while being sensitive to the underlying distribution.

Among the three fairness principles chosen, operationalizing prioritarianism was the most challenging as it required ethical or moral choices to be made, an act that leaves a door open to uncertainty. The Prioritarian social welfare function proposed and used by M. Adler et al. (2017), has been critiqued in the past to be impractical because the average consumption or the bare minimum consumption of a resource is to be known for all the actors (*czero*) (Voorhoeve, 2014). With the current framework suggested in the method, the sensitivity of prioritarian fairness to ethical choices can be estimated. However, to inform policy, the values of *czero* must be determined.

The prioritarian formulation used in this method was used to estimate the social cost of carbon. Before this, it was used for income and wealth distribution problems (M. Adler, 2012). In those cases, the value of minimum consumption is subjective and depends on a person/citizen. However, in the case of water, the minimum consumption levels can be objectively determined at least for some uses. For example, agricultural use of water can be limited to as much as a certain crop type requires. It does not entirely depend on the choice of the person farming the land. On the other hand, for human consumption, there are efforts underway to agree and enforce a minimum supply of water that must be allocated for human consumption (Gleick, 1998; UNW-DPAC, 2013). If such values are determined or agreed upon by policymakers, this formulation of prioritarianism will be extremely beneficial to evaluate the fairness of water allocation because the window of uncertain values that minimum consumption can take is reduced. Additionally, ideas of sufficientarianism can also be included within this formulation to inform the bare minimum levels of consumption.

While the metrics were true to their definition, what was found interesting is that with the increase in aggregate utility, there can be more equal solutions. This is in line with the idea that cooperating to share benefits makes the 'pie' bigger creating more utility available to distribute. Such situations could lead to win-win arrangements for all actors (Davidsen, 2010). While increasing the size of the pie *can* result in equally distributed policies, a high utility can also be achieved at the cost of one actor. To identify win-win situations, therefore, it was found useful to use both aggregate utility and Gini index. Given high utility can also be achieved at the cost of an actor, finding these win-win situations demands a ”compromise” to be made along with the value of a fairness metric. This brings us to the next point of discussion in the fairness evaluation method: what does it mean to be in an acceptable range of aggregate utility or Gini index?

8.3. Moving past binary bias: Fairness as a matter of degree

Within behavioral psychology and social sciences, an idea that has been recently popularized by Grant (2021) is 'binary bias'. Binary bias is ”a tendency to impose categorical distinctions on continuous data. Evidence is compressed into discrete bins, and the difference between categories forms the summary judgment” (Fisher & Keil, 2018). The idea of using binary thinking to structure information is ubiquitous: Heroes and villains, us and them, good and bad, right and wrong, cooperation and conflict.

Fairness deals with the moral question of right and wrong. Utilitarians, Kantians, and many other moral theorists believe every act is either right or wrong in the binary sense (Hillerbrand & Peterson, 2014). If such a binary approach is used, the evaluation of every Pareto optimal policy will be binary, i.e. fair or unfair. In this study, utilitarian SWF and prioritarian SWF generate numeric values that indicate the total utility of a policy. These aggregations are over utility values, which are a relative measure i.e. utility values are essentially preferences over a set of policies (Hausman et al., 2016). This means that the aggregate values they generate are meaningless by themselves but only make sense in comparison. The principle behind utilitarianism is to maximize utility. This translates to identifying the policy that is 'more fair' than others because it provides 'more aggregate utility'. This suggests that the idea of fairness is not viewed as a binary but as a spectrum where the degree of fairness is determined. The method does not say that a few policies are fair while others are unfair but simply ranks them by the degree to which they align with the fairness principles.

The argument that a policy is fairer than others according to a fairness principle does not remove the idea that there are no unfair policies at all. As mentioned before, although utilities are preferences over outcomes, identifying bare minimum values of consumption of water can help ascertain the bare minimum utility that must be given. In such a case the utility values become meaningful without having to compare them with the utility of another policy. Similarly, if there are irreversible losses, the value of utility loses meaning and cannot be used at all. In both these cases, the idea of 'unfair' can be operationalized.

Figure 8.1 shows the idea of fairness as a matter of degree. After thresholds are applied for irreversible losses and minimum levels of water consumption are assumed, there would be multiple Pareto optimal policies to choose from. It is in such a case that the idea of degrees of fairness holds. A lot of policies may generate high utility (as is the case with the Susquehanna river basin), but some may have more utility than others. Similarly, there may be multiple equal policies, but some are more equal than others. By considering fairness as a matter of degree, there is room to create compromise solutions that balance the tradeoffs between equality, utility, and priority. Using the visual analysis proposed within the method, the policies with high values of the Gini index, aggregate utility, and prioritarian welfare can be identified. These policies can then be deliberated upon further.

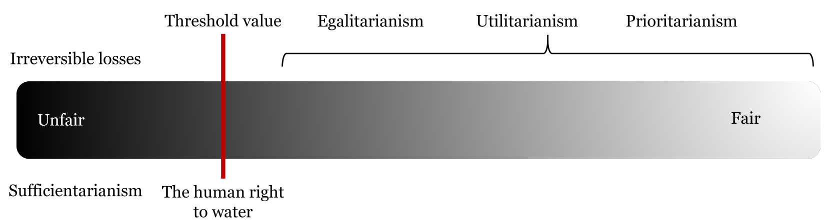


Figure 8.1: Viewing fairness as a matter of degree

8.4. The hidden cost of cooperation

Non-cooperative stability definitions were chosen to be operationalized for two reasons. First is the idea that water is a scarce resource and multiple international water treaty negotiations involve the presence or threat of unilateral actions. Second is the fact that non-cooperative game theory answers the question 'is this a stable policy?', as opposed to cooperative game theory which only ascertains the conditions under which a stable policy could exist.

When non-cooperative stability definitions were operationalized and applied to the Susquehanna river basin, no stable policies were found. This absence of stable policies is attributed to a status quo bias - a preference for the current state that biases the actor from changing their stance (Kahneman et al., 1991). This bias increased as the number of players at the table went up. If there is a status quo in place where no actor is getting any utility, then all the Pareto optimal policies are stable no matter how much utility they provide to the actors. However, once a policy is in place, even if it benefits only a few actors, agreeing on a new policy becomes expensive. In discussing how to cope with public value conflicts, De Graaf, Huberts, and Smulders (2016) suggest that a "choice for a particular value has to be made only once, after which it becomes routine". If there is a status quo in place that exists because of the alignment of actors on some value, it becomes difficult to change it.

For example, let us take the game with all 7 players where the disagreement point is the Susquehanna status quo. In the Susquehanna status quo, Baltimore gets maximum utility and recreation gets nothing. However, when all 7 equally powerful actors negotiate over a new policy, no new stable policy emerges. When multiple players become involved, it becomes a multi-issue game, and the outcome or the act of deciding on a policy will depend on the dynamics of the game and the interpersonal relationship between the players (De Bruijn & Ten Heuvelhof, 2018). This brings us to the limitation of the method.

It is often not the case in real life that 7 players are all equally powerful in a transboundary river basin. It is also not likely that each actor will represent only one objective. It will more often be the case that some actors will have more than one objective to meet and each of them will have different levels of influence. This is a limitation of the method. It would be interesting as part of future work to map multiple objectives to a single actor, prioritize the objectives and then identify stable solutions. The GMCR package can also be modified to include such a scenario. (More about technical improvements and limitations is discussed in appendix B.2).

While non-cooperative stability definitions yielded no stable solution for 7 actors, fallback bargaining guarantees the existence of a stable solution. This is because it answers the question 'What is the most stable policy out of all the existing policies?' Additionally, the outcome from fallback bargaining was seen in the 75^{th} percentile of equality and 95^{th} percentile of utility. Like in non-cooperative settings, fallback bargaining also assumes players want to maximize their utility and uses their ranking over the policies to determine which is the most stable policy. This means that a more utilitarian policy could be considered stable if it does not give the worst outcome to any player.

The idea that fallback bargaining, or any other cooperative solution concept for that matter, guarantees the existence of a stable solution could be misleading. This is because the cost of achieving this stable solution is masked under the assumption cooperative game theory makes that all players *will agree* to a social contract. The cost of this agreement is not captured within any method that operationalizes cooperative stability. In this case, the social contract that fallback bargaining imposes is simply to avoid the 'worst possible outcome' for all actors. Although that may sound reasonable, the current status quo policy in the Susquehanna basin maximizes the utility to Baltimore by giving the worst outcome to recreation. Avoiding the worst-case outcome for one actor when multiple actor's needs are co-dependent (i.e., when there are inherent tradeoffs) may not be easy to achieve. Like in the case of non-cooperative stability, incorporating power dynamics between players and relative importance of objectives can also improve the stable outcome generated by fallback bargaining and can be pursued as part of future work.

At this juncture, it is worth reflecting on why both cooperative and non-cooperative stability concepts are both simultaneously suitable for a case in hand. Within transboundary water interaction, research has identified that various forms of conflict over water occur almost without exception along-side various forms of cooperation. However, when it comes to analysis, they are often treated separately because of the existing research apparatus for scenarios of cooperation and conflict. This separation can mean that fewer ugly faces of conflict and fewer pretty faces of cooperation are overlooked (Zeitoun & Mirumachi, 2008). The presence of a stable solution in a cooperative setting and the absence of one in a non-cooperative setting is indicative not just of the status quo bias but also the hidden cost of cooperation. Cooperative solutions assume the best of people by treating them as those who adhere to social contracts once they agree to do so. Non-cooperative solutions assume the worst of world leaders who need to share water resources by treating them as self-interested actors. Both can be argued as simplistic assumptions as in the real world, people would agree on a few things while vehemently pushing for their interests in other cases. Considering both within the same method is argued as looking at the best case and worst case stability outcomes.

8.5. Implications to distributive justice in transboundary river basins

After discussing the merits and demerits of the method itself, attention is drawn back to the implications of this research to distributive justice water resource allocation in transboundary river basins. Both in domestic and international settings, water resource management is an inherently political process (Warner & Wegerich, 2010). Public policies are the means through which politics allocates values (Easton, 1965). The biggest challenge for governance, however, is to balance contradicting values (De Graaf et al., 2016). Applying this method to the Susquehanna river basin highlighted different contradictions at multiple stages.

From the onset, the identified Pareto optimal outcomes highlighted a tradeoff between environmental shortage and water supply (contradicting objectives). Subsequently, some policies displayed a tradeoff between utilitarianism, egalitarianism, and prioritarianism (contradicting values). Finally, the fair policies were not found to be stable in a non-cooperative setting (contradicting criteria). However, the proposed method, at multiple points, highlights policies where a balance can be achieved between different objectives and values. Using the three fairness principles as axes along which a desirable policy space is identified, the tradeoffs between policies can be minimized while also balancing the three values.

The gap between fairness and stability can be bridged if the actors in the transboundary river basin are entirely cooperative with each other. Using Rawls' idea of justice, cooperative stability definitions are the closest to just policies. This is because actors agree to a certain idea of fairness, in this case, 'not giving the worst outcome to any actor' and identifying a stable policy. The fairness of the outcome from fallback bargaining was also found to be aligning with utilitarian values. The Gini index was in the 75th percentile (i.e. there were more equal policies but this is more stable, and hence, more just). However, it was seen that cooperative stability is not representative of the true stability of a policy.

It is important to mention here that this method was used to evaluate existing Pareto optimal policies. Pareto optimality can be argued as an idea of fairness by itself and it has been identified as a necessary but not sufficient condition for fairness (Cudd, 1996). In that light, ideas of efficiency and fairness may align with each other. Stability, on the other hand, need not align with Pareto optimality, or efficiency, at all (Read et al., 2014). Therefore, to identify fair policies that are stable, it is worth considering a different allocation principle within the formulation itself. This can be done as part of future work.

Overall, the idea of fairness and stability have been operationalized to evaluate existing allocations. To identify just policies using such a method, different allocation principles must be considered to identify the policies that are both fair and stable. Following this, the policies in this intersection must be tested for stability over time. Although that is a long road to tread upon, this method can be seen as a first step.

9 Conclusion

For decades resource allocation has been a critical problem of public policy. For all those decades, there has been cooperation and conflict over sharing water resources equitably and reasonably, especially between riparian states. Rawl's argued that a just distribution is one that is fair and accepted by society. In more pragmatic terms, economists argue that distributions must be efficient and thus, Pareto optimal. This research began with the idea that it is desirable to distribute water, a risk, a scarce resource and a human right, efficiently, fairly and in a manner that is accepted by people across riparian states. It was found that no one method incorporates all three considerations explicitly in identifying resource allocations.

To bridge the research gap, the following research question is formulated

Given a Pareto optimal solution set of non-transferable water resource allocations in a shared river basin, how can stability and fairness be evaluated?

The research intends to propose a method that can be used for the fairness and stability evaluation of efficient policies. When dealing with fairness and stability, the most crucial aspect is context. Three guiding questions are used to set the context of this research -

- **What is being distributed** - The utility obtained from water is distributed between actors. Utility in this context is the relative preference ordering of an actor over a set of policies i.e. it is a measure of how much the end-user values the resource allocation given to them. This is measured using a diminishing marginal utility function where every additional unit of utility has a lesser value. The use of this utility function introduces an important research choice - the value of marginal utility. The value of marginal utility (η) determines how the value of every additional unit of water changes for an actor.

- **Who is it being distributed to** - The allocation is assumed to be made to equally powerful actors in a shared river basin who have an equal right to the water resource in line with the ideas of *cosmopolitanism.*

- **How is it being distributed?** - The allocation principle i.e. the principle according to which water is divided between actors is taken as Pareto optimality. This means that an efficient set of policies are considered for further analysis.

In this context, the research aims to develop a method to evaluate the fairness and stability of efficient policies. This method is applied to the case study of the Susquehanna River Basin. An existing model of the Susquehanna river basin from Salazar et al. (2016) is used to generate optimal policies. 106 Pareto optimal policies are identified from this case study which distribute benefit from water optimally among 7 objectives - Baltimore water supply, Chester water supply, atomic power plant discharge, hydropower revenue, recreation water supply, environmental shortage and flood risk. The actors, in this case, are assumed to be 7 people who each represent one model objective. To answer the main research question, 4 sub-questions are used. Each of the sub-questions is discussed here.

SQ1: What are the fairness principles that can be applied to evaluate fairness?

There are multiple fairness principles found in the literature that can be used to evaluate fairness. To ascertain the suitability of a principle three criteria are used - suitability to the context of water resource allocation, suitability for evaluation and amount of information required for operationalization. The idea of suitability for evaluation was introduced as a criterion to demarcate evaluation and the process of division. Fairness principles can be used to divide a resource as well as to evaluate an existing division. Given this research deals with the evaluation of an existing optimal allocation, this criterion was included.

From the literature in the water resource allocation context, 6 fairness principles have been identified - utilitarianism, egalitarianism, prioritarianism, sufficientarianism, envy-free and weighted utilitarianism.

Utilitarianism is a principle that states that fair policies are those that maximize the utility for the maximum number of people. It is the most widely operationalized principle in ethics literature. It has been operationalized using cost-benefit analysis or the utilitarian social welfare function in water-related contexts. For evaluation in this case the utilitarian social welfare function is used. The policy that maximizes the aggregate utility, calculated by simply adding the individual utilities, can be considered fair according to utilitarianism. This means that it only requires the individual utility values as input. Utilitarianism is agnostic to the nature of the distribution. It only deals with an aggregate value of all the utilities given to each of the actors. This indifference to the nature of the distribution brought in multiple other ideas of fairness as a response to utilitarianism.

Weighted utilitarianism allows giving different weights to the actors to whom utility is being distributed. However, it was found that such a differential weight can also be achieved by modifying the value of marginal utility (η) for each actor. Given the sensitivity to η is already within the study, this principle was not considered for further analysis.

Egalitarianism is a principle that states that policies that divide resources equally are fair. For evaluation, Gini index can be used to ascertain how equal a policy is. Gini index is a ratio of the sum of the relative difference of utilities between two actors to the mean utility that an actor can get if it was equally distributed. This measure, therefore, ensures equality by reducing relative differences between actors. This also uses only individual utilities as input and has been used in water allocations before. This principle does not object to equality being attained by reducing everyone's utility. As a counter to this, prioritarianism is proposed.

Prioritarianism is a principle that prioritizes the worse-off. The worse-off are those that obtain very little utility above the bare minimum consumption level that they need to survive. This principle, therefore, reduces absolute inequality. This principle is used lesser than egalitarianism and utilitarianism but is quickly gaining relevance in climate ethics and water ethics. Prioritarian utility is a bridge between utilitarianism and egalitarianism where it aggregates the utility of individual actors but is also sensitive to the underlying distribution. It achieves this by including two ethical choices within the formulation of a Prioritarian social welfare function - priority to the worse off (γ) and bare minimum level of consumption (*czero*) The social welfare function subtracts the bare minimum utility from the utility every individual receives, prioritizes it to the degree of γ and aggregates this value. This principle requires the most amount of information to operationalize.

Sufficientarianism A a response to the commodification of water, was the argument that water is a human right. The United Nations declared water as a human right and identified a bare minimum amount of water that every individual must receive. This principle, therefore, becomes the most suited principle for the idea of water. However, it requires information about the number of people that use the water for domestic purposes, if they have access to it and how much water is lost in ensuring water supply. For these reasons, sufficient principle is not explicitly considered. However, it is extremely suitable for fairness evaluation.

Envy free is a principle that states that no actor must envy the utility another actor received. To operationalize envy-free, the utility an actor receives is subtracted from the utility they would have got had they been allocated a different actor's share. The application of this concept has predominantly been seen in cases where water is being distributed directly and not when the utility of benefit from water is being distributed. This is therefore considered more suitable when dividing a resource rather than evaluating an existing allocation when the goods are not transferable.

To conclude, utilitarianism, prioritarianism, egalitarianism and sufficientarianism are extremely suitable for evaluating the fairness of non-transferable water resource allocations. Given the information for sufficientarianism was not available, it was excluded from further analysis. However, its potential place in evaluating fairness is discussed as part of the discussion and future work.

SQ2: What are the stability concepts that are suitable to evaluate stability?

Game theory was used to identify stability concepts that are suitable for this research. Game theory has been widely applied in the context of transboundary water allocations with and without transferable resources. Within game theory, 4 areas were identified to ascertain suitability based on the same criteria as fairness - cooperative games without conflict, cooperative games with conflict, non-cooperative games without conflict, non-cooperative games with conflict.

Cooperative games without conflict are situations where all actors come together, agree to abide by a social contract and ascertain how to divide a resource fairly. This area of game theory predominantly suggests how a resource can be divided and hence is not considered suitable.

Cooperative games with conflict are situations where actors choose to cooperate and agree to divide the resource according to a certain fairness principle. However, upon allocation, they choose to bargain over the allocations instead of abiding by the social contract entirely. Bargaining solutions have been used in ethics literature before and use the utility allocated to each actor as input to identify bargaining solutions. For this study, fallback bargaining is considered.

Non-cooperation with conflict are situations where actors do not wish to cooperate i.e. they pursue their own interests but are communicating or making moves within the same arena. This is considered suitable because multiple water-sharing treaties and arrangements have the presence or threat of unilateral action by multiple actors. Although cooperation over water is more common, non-cooperation is not unheard of. Additionally, non-cooperative game theory is considered more fundamental than game theory as it does not assume the presence of a social contract. 4 stability definitions from game theory are considered within this study each with differing amounts of foresight, knowledge of opponents preferences and players willingness to indulge in disimprovement of their outcomes to hamper other's movements.

Non-cooperation without conflict is a scenario where actors do not wish to cooperate or communicate. Water is a scarce resource that follows the subtraction criteria i.e. one actor using it reduces the amount there is for another actor. This makes the game arena competitive by nature. Therefore, non-cooperation without conflict is unlikely and hence not considered suitable.

To conclude, the areas of bargaining games and non-cooperative games with conflict are considered suitable for this case because water is a competitive resource and in an international setting, negotiations are bound to happen over allocations. Additionally, they have been applied in the water context before and only require ranking over policy outcomes as inputs to determine stability.

SQ3: How can fairness principles be used to evaluate fairness?

Three fairness metrics were identified to evaluate fairness - utilitarian social welfare function for utilitarianism, Gini index for egalitarianism and prioritarian social welfare function for prioritarianism. The final measure generated by the two social welfare functions is termed aggregate utility (for utilitarianism) and prioritarian utility. These metrics were used to analyse the fairness of Pareto optimal policies in the Susquehanna river basin. The insights from applying it to the case allow this research question to be answered.

- **Using fairness metrics to identify tradeoffs between fairness principles**
  Inherent tradeoffs were identified between the three principles as their formulations concern different aspects of fairness. The fairness metrics allow visualization of the extent to which tradeoffs exist between fairness principles. This helps identify policies that are fair according to one metric but are not fair according to another.

- **Using fairness metrics to balance tradeoffs between fairness principles**
  Given utility is being distributed between actors and utility is a measure of relative preferences, it was seen that prioritarian utility, aggregate utility and Gini index can be used to determine the relative fairness of a policy i.e. the policy with the highest utility is not seen as the only fair policy but it is the most fair policy while others can be considered less fair. If the metrics are used to determine the degree of fairness according to each principle, compromises can be made along utility, equality and prioritarian utility to identify policies that perform well on all three metrics.

- **Using fairness as a threshold to identify unfair policies**
  An additional way of using fairness principles is as threshold values. The threshold values can be identified using principles like sufficientarianism or identifying the minimum value that users require. The idea of degrees of fairness also falls through when considering irreversible losses. If there are irreversible losses, the value beyond which an irreversible loss occurs can be used as a threshold. The policies that adhere to these threshold values can then be considered in evaluating degrees of fairness.

- **Using fairness metrics bearing ethical and research choices in mind**
  Prioritarianism was found to be the most challenging principle to operationalize owing to its sensitivity to ethical choices. The choice of priority given to the worse-off, and bare minimum consumption levels for every actor are both ethical or moral and need to be made by the policy-maker. these choices significantly impact the fairness and tradeoffs generated by prioritarianism. The existing method can be expanded upon further to explore the consequences of these ethical choices on prioritarian policies.
  The other choice made within this research is that of marginal utility. The best way to evaluate fairness would be to determine a different value of marginal utility for every actor based on the type of outcome they receive. This is currently not done in the method and acts as a limitation.

To conclude, fairness metrics can be used as a scale or a threshold in evaluating fairness. They can also be used to compare fairness across principles. However, it is critical to bear in mind that they are sensitive to the ethical choices made within the research and the utility functions used within the study.

SQ4: How can stability concepts be used to evaluate policy stability?

Fallback bargaining and 4 non-cooperative solution concepts are applied to ascertain stability. Stability analysis requires the presence of a disagreement point. This is the outcome if the actors do not agree on a stable policy. The current Susquehanna status quo policy is considered as a disagreement point along with four other Pareto optimal policies within this study.

- **Using non-cooperative stability to identify if there is a stable policy**
  Non-cooperative stability identifies if there is a stable solution among the set of policies. This means the outcome of applying non-cooperative stability definitions can result in not finding a stable policy at all. This was the case in the current study when 7 players are playing the game.

- **Using non-cooperative stability definitions with the appropriate status quo**
  It was seen that the choice of disagreement point or status quo can significantly impact the presence and number of stable policies. When no policy is in place i.e. all actors get 0 utility, all Pareto optimal policies are stable. However, if there is an existing policy in place there is a large status quo bias where the cost of bringing a new policy in place is high.

- **Using fallback bargaining to identify the most stable policy**
  Fallback bargaining can be used to identify the most stable policies if the actors decide not to select a policy that gives the worst outcome to any actor. It was found that even in the presence of an existing status quo, the outcome from fallback bargaining are stable i.e. they did not exhibit status quo bias. This method is, therefore, useful in identifying at least one stable outcome.

- **Using both cooperative and non-cooperative stability definitions**
  It was seen that although cooperative solutions assure the presence of a stable solution they hide the cost of getting all the players to agree on a bargaining method. Additionally, it is often the case that cooperation and non-cooperation co-exist in the same transboundary river basin. Cooperative settings assume that all players are amicable and will adhere to the chosen bargaining method while non-cooperative settings assume that players are only self-interested and will not adhere to any contract. For these reasons, it is believed that using them both could provide an overview of the best and worst cases of stability.

The method, overall, operationalizes fairness and stability to evaluate efficient transboundary water resource allocations. It was found that by using fairness as a degree and identifying policies with high utility and equality, win-win situations can be created for all actors. Stability of policies is ascertained disparately from fairness and it was found that it is difficult to get 7 actors to agree on putting up a new policy when an existing policy is already in place. Cooperative bargaining solutions were found to be the sweet spot where the outcome can be considered fair if it does not violate any threshold or bare minimum values of consumption and can be considered stable if all actors choose to cooperate.

9.1. Limitations and Future Work

As is the case with most research endeavours, this study also opened up more research questions in the process of answering some and makes multiple choices and assumptions along the way which act as limitations. This section discusses the key limitations and areas of future work.

Although all the choices and assumptions have been discussed, substantiated and checked for sensitivity wherever necessary, some choices are still simplifications and overlook the nuances of the real world that must make their way into evidence-based policy making. One such simplification is treating all the objectives with the same utility function and the dis-utilities as simply negative utilities. There is existing empirical research that can help ascertain the utility of water for specific purposes such as the work proposed by Karamouz et al. (2011) that ascertains utility functions for urban and agriculture use. Ascertaining utility for all the different purposes is challenging but the need for such research has been identified by organizations such as the United Nations in their attempt to compile ideas for the valuation of water (UNESCO & World Water Assessment Programme, 2021). Pursuing this would involve mapping each of the outcomes to their exact use cases and then identifying utility functions for all of them.

Prioritarianism, although the most challenging to implement exhibits a lot of promise in being a bridge between egalitarianism and utilitarianism. Delving deeper into the process in which ethical choices can be made about the base level of consumption and priority to the worse off can reduce the uncertainty caused due to changing ethical choices. In the context of water, it would be interesting to see if sufficientarianism can inform the ethical choices that need to be made for the prioritarian social welfare function.

The most fitting next step for this research would be to incorporate utility definitions, fairness principles and stability concepts into the problem formulation itself. This would mean changing the problem from an evaluation to a division of resources. Taking such a step would avoid hinging on Pareto optimal outcomes which are known to not guarantee stability (Read et al., 2014). Incorporating these principles into the formulation allows for the creation of rival problem formulations with policies generated using different allocation principles which can then be evaluated using the proposed method.

Finally, the time component of all of these fairness principles and stability has been kept out of the study. Incorporating that and observing how fairness and stability change over time could allow this research to be meaningfully tied back to distributive justice in transboundary water resource allocations.

References

Adler, M. (2012). *Well-being and fair distribution: beyond cost-benefit analysis.* Oxford University Press.

Adler, M., Anthoff, D., Bosetti, V., Garner, G., Keller, K., & Treich, N. (2017). Priority for the worse-off and the social cost of carbon. *Nature Climate Change*, *7*(6), 443–449.

Adler, M. D. (2016). Benefit–cost analysis and distributional weights: an overview. *Review of Environmental Economics and Policy*, *10*(2), 264–285.

Adler, M. D. (2019). *Measuring social welfare: An introduction.* Oxford University Press, USA.

Aleksandrov, M., Ge, C., & Walsh, T. (2019). Fair division minimizing inequality. In *Epia conference on artificial intelligence* (pp. 593–605).

Anand, P. B. (2007). Right to water and access to water: an assessment. *Journal of International Development: The Journal of the Development Studies Association*, *19*(4), 511–526.

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

Argenziano, R., & Gilboa, I. (2015). *Weighted utilitarianism, edgeworth, and the market* (Tech. Rep.). Discussion paper, Mimeo.

Botzen, W. W., & van den Bergh, J. C. (2014). Specifications of social welfare in economic studies of climate policy: overview of criteria and related policy insights. *Environmental and Resource Economics*, *58*(1), 1–33.

Brams, S. J., & Kilgour, D. M. (2001). Fallback Bargaining. , 30.

Broome, J. (2008). The ethics of climate change. *Scientific American*, *298*(6), 96–102.

Chevaleyre, Y., Endriss, U., Lang, J., & Maudet, N. (2007). A short introduction to computational social choice. In *International conference on current trends in theory and practice of computer science* (pp. 51–69).

Chun, Y., & Thomson, W. (1990). Bargaining with uncertain disagreement points. *Econometrica: Journal of the Econometric Society*, 951–959.

Ciullo, A., Kwakkel, J. H., De Bruijn, K. M., Doorn, N., & Klijn, F. (2020). Efficient or Fair? Operationalizing Ethical Principles in Flood Risk Management: A Case Study on the Dutch-German Rhine. *Risk Analysis*, *40*(9), 1844–1862. Retrieved 2021-03-06, from https://onlinelibrary.wiley.com/doi/10.1111/risa.13527 doi: 10.1111/risa.13527

Cooley, H. (2020). Urban and agricultural water use in california, 1960–2015.

Craswell, R. (1998). Incommensurability, Welfare Economics, and the Law. *University of Pennsylvania Law Review*, *146*(5), 1419. Retrieved 2021-05-09, from https://www.jstor.org/stable/3312810?origin=crossref doi: 10.2307/3312810

Cudd, A. E. (1996). Is pareto optimality a criterion of justice? *Social theory and practice*, *22*(1), 1–34.

Dasgupta, P. (2008). Discounting climate change. *Journal of risk and uncertainty*, *37*(2), 141–169.

Davidsen, P. (2010). Hydrosolidarity as water security in the okavango river basin. *The politics of water: A survey*, 68–95.

De Bruijn, H., & Ten Heuvelhof, E. (2018). *Management in networks.* Routledge.

De Clippel, G., & Rozen, K. (2019). Fairness through the lens of cooperative game theory: An experimental approach. *Available at SSRN 3527069.*

De Graaf, G., Huberts, L., & Smulders, R. (2016). Coping with public value conflicts. *Administration & society*, *48*(9), 1101–1127.

Dellapenna, J. W. (1994). Treaties as Instruments for Managing Internationally-Shared Water Resources: Restricted Sovereignty vs. Community of Property. , *26*, 31.

Dinar, A., & Hogarth, M. (2015). Game Theory and Water Resources Critical Review of its Contributions, Progress and Remaining Challenges. *Foundations and Trends® in Microeconomics*, *11*(1-2), 1–139. Retrieved 2021-05-09, from http://www.nowpublishers.com/article/Details/MIC-066 doi: 10.1561/0700000066

Doering, K., Quinn, J., Reed, P. M., & Steinschneider, S. (2021, July). Diagnosing the Time-Varying Value of Forecasts in Multiobjective Reservoir Control. *Journal of Water Resources Planning and Management*, *147*(7), (ASCE)WR.1943–5452.0001386, 04021031. Retrieved 2021-05-21, from http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0001386 doi: 10.1061/(ASCE)WR.1943-5452.0001386

Doorn, N. (2013). Water and justice: Towards an ethics of water governance. , 18.

Doorn, N. (2018). Distributing risks: Allocation principles for distributing reversible and irreversible losses. *Ethics, Policy & Environment*, *21*(1), 96–109.

Doorn, N. (2019). *Water ethics: An introduction*. Rowman & Littlefield International.

Driver, J. (2009). The history of utilitarianism.

Easton, D. (1965). A systems analysis of political life.

Fang, L., Hipel, K., Kilgour, D., & Peng, X. (2003, February). A decision support system for interactive decision making - Part II: analysis and output interpretation. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *33*(1), 56–66. (Conference Name: IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)) doi: 10.1109/TSMCC.2003.809360

Fang, L., Hipel, K. W., & Kilgour, D. M. (1989). Conflict models in graph form: Solution concepts and their interrelationships. *European Journal of Operational Research*, *41*(1), 86–100.

Fisher, M., & Keil, F. C. (2018). The binary bias: A systematic distortion in the integration of information. *Psychological Science*, *29*(11), 1846–1858.

Fu, J., Zhong, P.-A., Zhu, F., Chen, J., Wu, Y.-n., & Xu, B. (2018). Water Resources Allocation in Transboundary River Based on Asymmetric Nash–Harsanyi Leader–Follower Game model. , 270.

Gastwirth, J. L. (1972). The estimation of the lorenz curve and gini index. *The review of economics and statistics*, 306–316.

Giordano, M., Drieschova, A., Duncan, J. A., Sayama, Y., De Stefano, L., & Wolf, A. T. (2014). A review of the evolution and state of transboundary freshwater treaties. *International Environmental Agreements: Politics, Law and Economics*, *14*(3), 245–264.

Giuliani, M., Herman, J. D., Castelletti, A., & Reed, P. (2014). Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management. *Water Resources Research*, *50*(4), 3355–3377. Retrieved 2021-03-06, from http://doi.wiley.com/10.1002/2013WR014700 doi: 10.1002/2013WR014700

Gleick, P. H. (1998). The human right to water. *Water policy*, *1*(5), 487–503.

Gold, D. F., Reed, P. M., Trindade, B. C., & Characklis, G. W. (2019). Identifying Actionable Compromises: Navigating Multi-City Robustness Conflicts to Discover Cooperative Safe Operating Spaces for Regional Water Supply Portfolios. *Water Resources Research*, *55*(11), 9024–9050. Retrieved 2021-05-09, from https://onlinelibrary.wiley.com/doi/10.1029/2019WR025462 doi: 10.1029/2019WR025462

Grant, A. (2021). *Think again: The power of knowing what you don't know*. Viking.

Grasso, M. (2007). A normative ethical framework in climate change. *Climatic change*, *81*(3), 223–246.

Haddad, B. M. (2005). Ranking the adaptive capacity of nations to climate change when socio-political goals are explicit. *Global Environmental Change*, *15*(2), 165–176.

Haines, D. (2016). The indus waters treaty has always been controversial, but modi is wise to resist calls to abrogate it. *South Asia@ LSE*.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of political economy*, *63*(4), 309–321.

Hart, S. (1985). An axiomatization of harsanyi's nontransferable utility solution. *Econometrica: Journal of the Econometric Society*, 1295–1313.

Hausman, D., McPherson, M., & Satz, D. (2016). *Economic analysis, moral philosophy, and public policy*. Cambridge University Press.

He, H., Yin, M., Chen, A., Liu, J., Xie, X., & Yang, Z. (2018, September). Optimal Allocation of Water Resources from the "Wide-Mild Water Shortage" Perspective. *Water*, *10*(10), 1289. Retrieved 2021-06-19, from http://www.mdpi.com/2073-4441/10/10/1289 doi: 10.3390/w10101289

Hillerbrand, R., & Peterson, M. (2014). Nuclear power is neither right nor wrong: The case for a tertium datur in the ethics of technology. *Science and Engineering Ethics*, *20*(2), 583–595.

Hipel, K. W., Fang, L., & Kilgour, D. M. (2020). The graph model for conflict resolution: Reflections on three decades of development. *Group Decision and Negotiation*, *29*(1), 11–60.

Hurwicz, L., & Sertel, M. R. (1999). Designing mechanisms, in particular for electoral systems: the majoritarian compromise. In *Contemporary economic issues* (pp. 69–88). Springer.

Jafino, B. A., & Kwakkel, J. (2021). *Equitable adaptation planning under deep uncertainty for the upper vietnam mekong delta* (Tech. Rep.). Copernicus Meetings.

Jeffrey, P. (2018). Distributive justice, stability, and sustainable water resources management.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, *5*(1), 193–206.

Kalai, E. (1977). Proportional solutions to bargaining situations: interpersonal utility comparisons. *Econometrica: Journal of the Econometric Society*, 1623–1630.

Karamouz, M., Akhbari, M., & Moridi, A. (2011, May). Resolving Disputes over Reservoir-River Operation. *Journal of Irrigation and Drainage Engineering*, *137*(5), 327–339. Retrieved 2021-06-14, from http://ascelibrary.org/doi/10.1061/%28ASCE%29IR.1943-4774.0000292 doi: 10.1061/(ASCE)IR.1943-4774.0000292

Kasprzyk, J. R., Reed, P. M., & Hadka, D. M. (2016). Battling Arrow's Paradox to Discover Robust Water Management Alternatives. *Journal of Water Resources Planning and Management*, *142*(2), 04015053. Retrieved 2021-05-09, from http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000572 doi: 10.1061/(ASCE)WR.1943-5452.0000572

Kijne, J. W. (2003). *Unlocking the water potential of agriculture.* Food & Agriculture Org.

Kilgour, D. M., Hipel, K. W., & Fang, L. (1987). The graph model for conflicts. *Automatica*, *23*(1), 41–55.

Kim, K., & Glaumann, K. (2012). Transboundary water management: who does what, where? *Analysing the Data in SIWI's Transboundary Water Management Database Swedish Water House, Stockholm.*

Kind, J., Wouter Botzen, W., & Aerts, J. C. (2017). Accounting for risk aversion, income distribution and social welfare in cost-benefit analysis for flood risk management. *Wiley Interdisciplinary Reviews: Climate Change*, *8*(2), e446.

Kleinman, A. P. (1969). The production function and the imputation of the economic value of irrigation water. , 246.

Kwakkel, J. H. (2017). The exploratory modeling workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. *Environmental Modelling & Software*, *96*, 239–250.

Kwakkel, J. H., Walker, W. E., & Haasnoot, M. (2016). *Coping with the wickedness of public policy problems: approaches for decision making under deep uncertainty.* American Society of Civil Engineers.

Kıbrıs, ., & Sertel, M. R. (2007, February). Bargaining over a finite set of alternatives. *Social Choice and Welfare*, *28*(3), 421–437. Retrieved 2021-06-01, from http://link.springer.com/10.1007/s00355-006-0178-z doi: 10.1007/s00355-006-0178-z

Lamont, J. (2017). *Distributive justice.* Routledge.

Li, K. W., Kilgour, D. M., & Hipel, K. W. (2004). Status quo analysis of the flathead river conflict. *Water Resources Research*, *40*(5).

List, C. (2013). Social choice theory.

Madani, K. (2010). Game theory and water resources. *Journal of Hydrology*, *381*(3-4), 225–238. Retrieved 2021-05-09, from https://linkinghub.elsevier.com/retrieve/pii/S0022169409007653 doi: 10.1016/j.jhydrol.2009.11.045

Madani, K. (2011). Hydropower licensing and climate change: Insights from cooperative game theory. *Advances in Water Resources*, *34*(2), 174–183. Retrieved 2021-03-06, from https://linkinghub.elsevier.com/retrieve/pii/S0309170810001843 doi: 10.1016/j.advwatres.2010.10.003

Madani, K., & Dinar, A. (2012). Cooperative institutions for sustainable common pool resource management: Application to groundwater: COOPERATIVE INSTITUTIONS FOR CPR MANAGEMENT. *Water Resources Research*, *48*(9). Retrieved 2021-05-09, from http://doi.wiley.com/10.1029/2011WR010849 doi: 10.1029/2011WR010849

Madani, K., & Hipel, K. W. (2011). Non-Cooperative Stability Definitions for Strategic Analysis of Generic Water Resources Conflicts. *Water Resources Management*, *25*(8), 1949–1977. Retrieved 2021-03-06, from http://link.springer.com/10.1007/s11269-011-9783-4 doi: 10.1007/s11269-011-9783-4

Madani, K., & Lund, J. R. (2011a). A monte-carlo game theoretic approach for multi-criteria decision making under uncertainty. *Advances in water resources*, *34*(5), 607–616.

Madani, K., & Lund, J. R. (2011b, May). A Monte-Carlo game theoretic approach for Multi-Criteria Decision Making under uncertainty. *Advances in Water Resources*, *34*(5), 607–616. Retrieved 2021-06-02, from https://www.sciencedirect.com/science/article/pii/S0309170811000352 doi: 10.1016/j.advwatres.2011.02.009

Madani, K., Shalikarian, L., Hamed, A., Rowney, C., Madani, K., Hamed, A., ... Rowney, . (2015, April). Bargaining Under Uncertainty: A Monte-Carlo Fallback Bargaining Method for Predicting the Likely Outcomes of Environmental Conflicts. In (pp. 201–212). doi: 10.1007/978-3-319-14215-9_11

Marchau, V. A., Walker, W. E., Bloemen, P. J., & Popper, S. W. (2019). *Decision making under deep uncertainty: from theory to practice*. Springer Nature.

McCaffrey, S. (2001). The contribution of the un convention on the law of the non-navigational uses of international watercourses. *International Journal of Global Environmental Issues*, *1*(3-4), 250–263.

McCracken, M., & Wolf, A. T. (2019). Updating the register of international river basins of the world. *International Journal of Water Resources Development*, *35*(5), 732-782. Retrieved from https://doi.org/10.1080/07900627.2019.1572497 doi: 10.1080/07900627.2019.1572497

Myerson, R. B. (1977). Graphs and cooperation in games. *Mathematics of operations research*, *2*(3), 225–229.

Nandalal, K. D. W., & Hipel, K. W. (2007, July). Strategic Decision Support for Resolving Conflict over Water Sharing among Countries along the Syr Darya River in the Aral Sea Basin. *Journal of Water Resources Planning and Management*, *133*(4), 289–299. Retrieved 2021-05-22, from http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9496%282007%29133%3A4%28289%29 doi: 10.1061/(ASCE)0733-9496(2007)133:4(289)

Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action: Presidential address, american political science association, 1997. *American political science review*, 1–22.

Pacific Institute. (2019). *Water conflict chronology.* Pacific Institute, Oakland, CA. Retrieved 31/05/2021, from https://www.worldwater.org/water-conflict/

Rasmusen, E. (1989). *Games and information: An introduction to game theory.* Blackwell Oxford.

Rawls, J. (1958). Justice as fairness. *The philosophical review*, *67*(2), 164–194.

Rawls, J. (1971). *A theory of justice.* Harvard university press.

Read, L., Madani, K., & Inanloo, B. (2014). Optimality versus stability in water resource allocation. *Journal of environmental management*, *133*, 343–354.

Reed, P., Hadka, D., Herman, J., Kasprzyk, J., & Kollat, J. (2013). Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in Water Resources*, *51*, 438–456. Retrieved 2021-05-09, from https://linkinghub.elsevier.com/retrieve/pii/S0309170812000073 doi: 10.1016/j.advwatres.2012.01.005

Renzetti, S., & Dupont, D. (2003). Ownership and performance of water utilities.

Salazar, J. Z., Reed, P. M., Herman, J. D., Giuliani, M., & Castelletti, A. (2016). A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control. *Advances in water resources*, *92*, 172–185.

Sen, A. (1970). *Collective choice and social welfare.* San Francisco: Holden Day. (Edinburgh: Oliver and Boyd, 197l; Amsterdam: North-Holland, 1979. Swedish translation: Bokforlaget Thales, 1988.)

Sen, A. (1986). Foundations of social choice theory: An epilogue. In J. Elster & A. Hylland (Eds.), *Foundations of social choice theory.* Cambridge: Cambridge University Press.

Sheikhmohammady, M., & Madani, K. (2008). Bargaining over the Caspian Sea — The Largest Lake on the Earth. In *World Environmental and Water Resources Congress 2008* (pp. 1–9). Honolulu, Hawaii, United States: American Society of Civil Engineers. Retrieved 2021-05-09, from http://ascelibrary.org/doi/10.1061/40976%28316%29262 doi: 10.1061/40976(316)262

Sinnott-Armstrong, W. (2003). Consequentialism.

Tian, J., Guo, S., Liu, D., Pan, Z., & Hong, X. (2019a). A Fair Approach for Multi-Objective Water Resources Allocation. *Water Resources Management*, *33*(10), 3633–3653. Retrieved 2021-02-25, from https://doi.org/10.1007/s11269-019-02325-5 doi: 10.1007/s11269-019-02325-5

Tian, J., Guo, S., Liu, D., Pan, Z., & Hong, X. (2019b). A fair approach for multi-objective water resources allocation. *Water Resources Management*, *33*(10), 3633–3653.

UNESCO, & World Water Assessment Programme. (2021). *Valuing water.* Retrieved 2021-06-20, from https://unesdoc.unesco.org/ark:/48223/pf0000375724_eng (OCLC: 1244428616)

UNWC. (2012a). Equitable and reasonable utilisation. Retrieved from https://www.unwatercoursesconvention.org/documents/UNWC-Fact-Sheet-4-Equitable-and-Reasonable-Utilisation.pdf

UNWC. (2012b). Theories of resource allocation. Retrieved from https://www.unwatercoursesconvention.org/documents/UNWC-Fact-Sheet-10-Theories-of-Resource-Allocation.pdf

UNW-DPAC. (2013). Cooperation in action: approaches, tools and processes. Retrieved from https://www.un.org/waterforlifedecade/water_cooperation_2013/pdf/water_cooperation_in_action_approaches_tools_processes.pdf

Verbeek, B., & Morris, C. (2004). Game theory and ethics.

Villamayor-Tomas, S., Avagyan, M., Firlus, M., Helbing, G., & Kabakova, M. (2016). Hydropower vs. fisheries conservation: a test of institutional design principles for common-pool resource management in the lower mekong basin social-ecological system. *Ecology and Society*, *21*(1).

Voorhoeve, A. (2014). Matthew D. Adler: Well-being and fair distribution: beyond cost-benefit analysis: Oxford, xx, 635 p. ISBN: 978-0-19-538499-4. *Social Choice and Welfare*, *42*(1), 245–254. Retrieved 2021-05-09, from http://link.springer.com/10.1007/s00355-013-0721-7 doi: 10.1007/s00355-013-0721-7

Walzer, M. (2015). *Just and unjust wars: A moral argument with historical illustrations*. Basic books.

Warner, J., & Wegerich, K. (2010). Is water politics? towards international water relations. *The politics of water: A survey*, 3–18.

Wegerich, K., & Warner, J. (2010). *The politics of water: A survey*. Routledge.

Wolf, A. T. (1999). Criteria for equitable allocations: the heart of international water conflict. *Natural Resources Forum*, *23*(1), 3–30. Retrieved 2021-05-26, from http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1477-8947.1999.tb00235.x (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1477-8947.1999.tb00235.x) doi: https://doi.org/10.1111/j.1477-8947.1999.tb00235.x

Yalew, S., Kwakkel, J., Zatarain Salazar, J., & Doorn, N. (2021). Intergenerational justice and trade-off analysis in water resources management of the nile. In *Egu general assembly conference abstracts* (pp. EGU21–7604).

Yalew, S. G., Kwakkel, J., & Doorn, N. (2021). Distributive justice and sustainability goals in transboundary rivers: Case of the nile basin. *Frontiers in Environmental Science*, *8*, 281.

Zeitoun, M., & Mirumachi, N. (2008). Transboundary water interaction i: Reconsidering conflict and cooperation. *International Environmental Agreements: Politics, Law and Economics*, *8*(4), 297–316.

A

Verification and sensitivity analysis

A.1. Utility function

This section discusses the distribution and tradeoffs seen among the actors when using payoffs and using the utility function. There is a slight difference in the tradeoffs with Chester where multiple model outcomes are mapped to high utility values. Similarly with Baltimore. This is because these two outcomes are the ones with the highest range of values. Therefore, the utility values start to plateau after a certain value of model outcome. The distributions of the payoffs and the utilities is observed next.

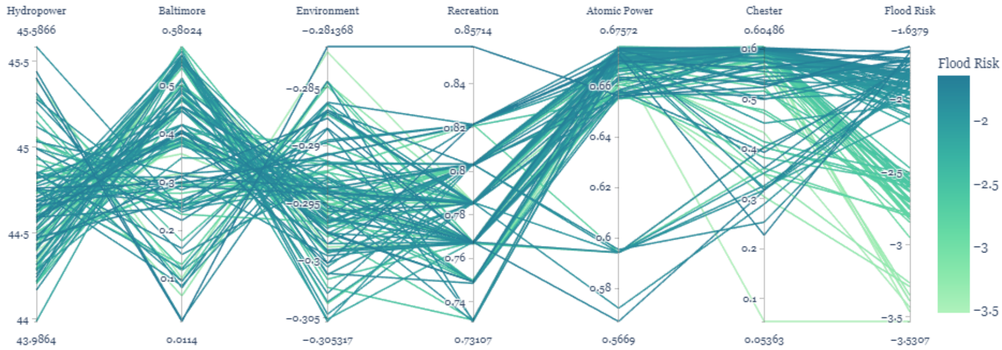


Figure A.1: Tradeoffs among the model outcomes

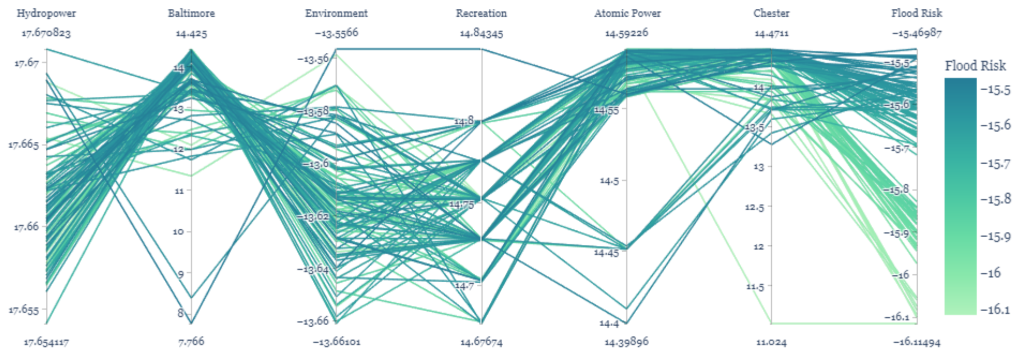


Figure A.2: Tradeoffs among the utilities of model outcomes

The distribution of each outcome is shown in figure A.3 and figure A.4

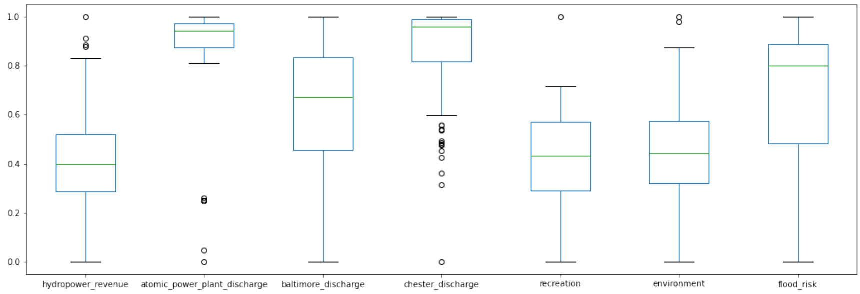


Figure A.3: Distribution of normalized payoff values for each outcome

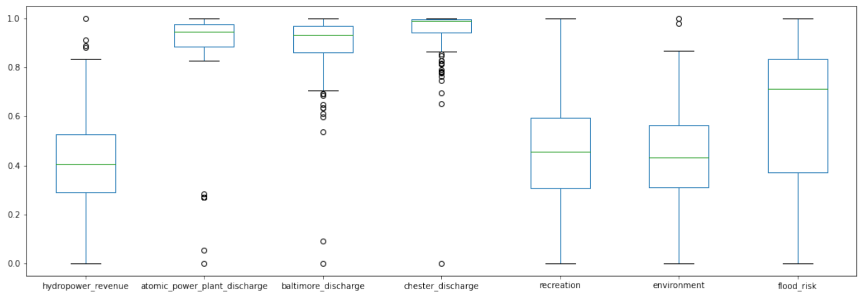


Figure A.4: Distribution of normalized utility values for each outcome

A.2. Mapping outcomes to observable values

It is often difficult to imagine what the payoff numbers mean in terms of tangible values. Especially in terms of fairness, it could come in handy to know what each of the model outcomes translate to in real life. In order facilitate that, the model outcomes were mapped to real world values. Table below shows the real world equivalents of the model outcomes.

Outcome	Unit	Description
Hydropower revenue	$/MWh	The plant generates an average of 4000 MWh per day. This makes each unit decrease an annual difference of roughly $1.4 million.
Atomic power plant discharge	volume reliability	A 1% decrease in reliability is roughly $109m^3$ annual water shortage
Baltimore Discharge	volume reliability	A 1% decrease in reliability is roughly $1693m^3$ annual water shortage
Chester Discharge	volume reliability	A 1% decrease in reliability is roughly 187m3 annual water shortage
Environmental Regulation	shortage index	A 1% increase in shortage is roughly $730m^3$ annual water shortage
Recreation	storage reliability	A 10% descrease in reliability would mean three days of low water for recreation of the 28 days for which it is required.
Flood duration	days	Each outcome represents the number of days

A.3. Fairness sensitivity to czero

By default, the values of zero levels of consumption are shown here -

- Baltimore, atomic power plant discharge, chester discharge - 0.1% volume reliability
- Hydropower revenue - $ mn 0.001
- Environment - -1 shortage index
- recreation - 0.1% shortage reliability
- Flood risk - 100ft of dam water excess (Which does not necessitate opening all the release gates of Susquehanna i..e it is not as bad as the worst flood in the area)

The value are assumed close to zero because the utility function does not accept a 0 value.

The sensitivity of the top 95th percentile prioritarian policies to the changes in zero levels and the tradeoff with other fairness principles is shown here.

From the plots the following observations can be made -

- The top percentile policies are extremely sensitive to the value of lowest level of consumption.
- The policies that were considered prioritarian when the zero consumption of hydropower was increased provide hydropower with greater hydropower than otherwise. Compare figure A.5 and A.7.
- In such a case there is also a larger tradeoff with utilitarian policies Compare figure A.6 and A.8.
- Changing the zero consumption of some outcomes like hydropower, recreation or environment reduces the number prioritarian policies in the top percentile and increases the tradeoff with aggregate utility. This is reflective of the inherent tradeoff within the system, where if environment, recreation or hydropower are prioritized there can be lesser policies with high aggregate utility i.e. the tradeoff for the others increases.
- This can be tied back to the distribution of utilities shown in figure A.4 or the distribution of payoffs shown in figure A.3. The mean utilities of Baltimore, Atomic power, Chester and Flood Risk is higher than hydropower, recreation or environment.

1. **c_{zero} consumption of hydropower is \$40mn**

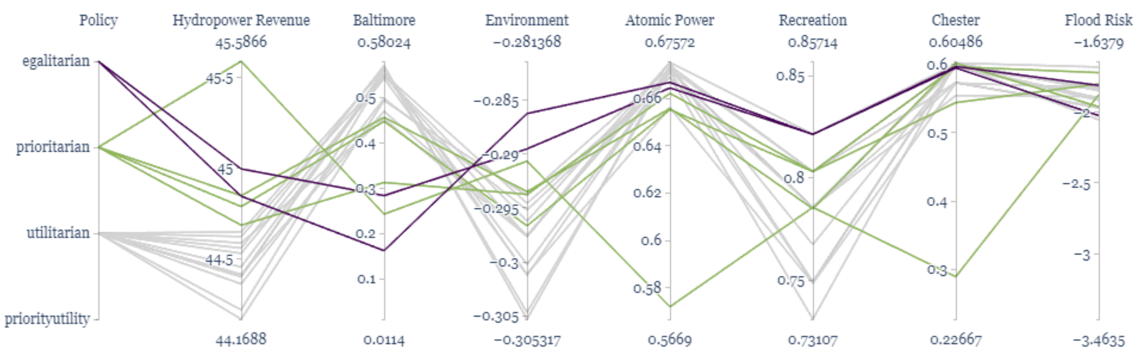


Figure A.5: Top 95th percentile outcomes when c_{zero} for hydropower is \$40mn

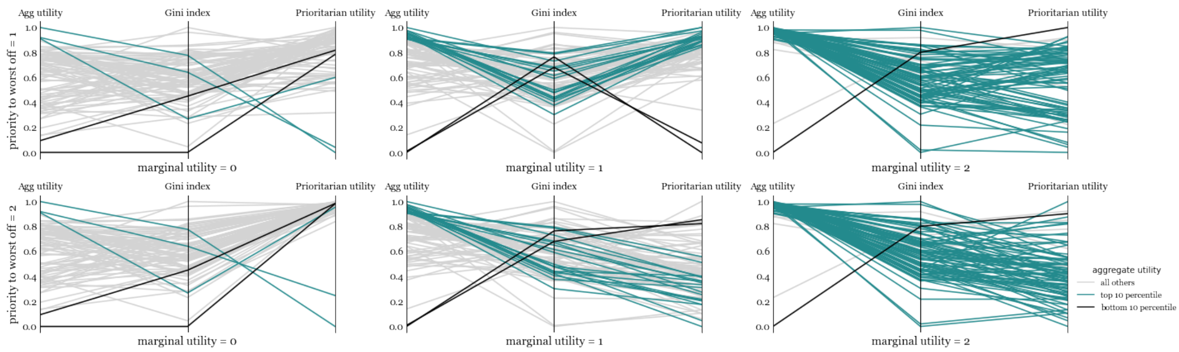


Figure A.6: Fairness principle tradeoffs when c_{zero} for hydropower is $40mn

2. c_{zero} consumption of atomic power plant is 40%

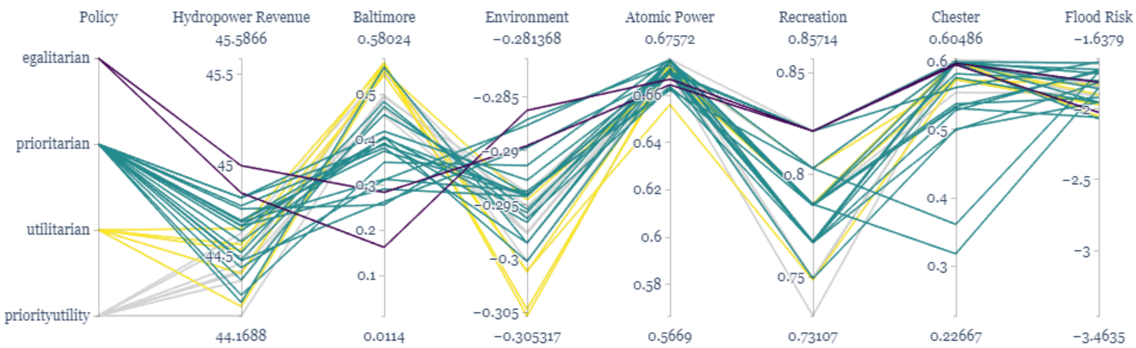


Figure A.7: Top 95th percentile outcomes when c_{zero} for atomic power plant is 40%

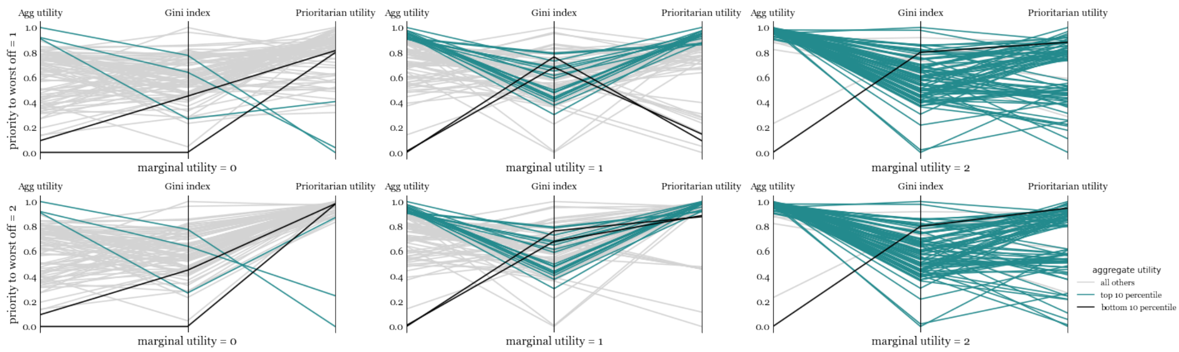


Figure A.8: Fairness principle tradeoffs when c_{zero} for atomic power plant is 40%

3. c_{zero} consumption of chester is 20%

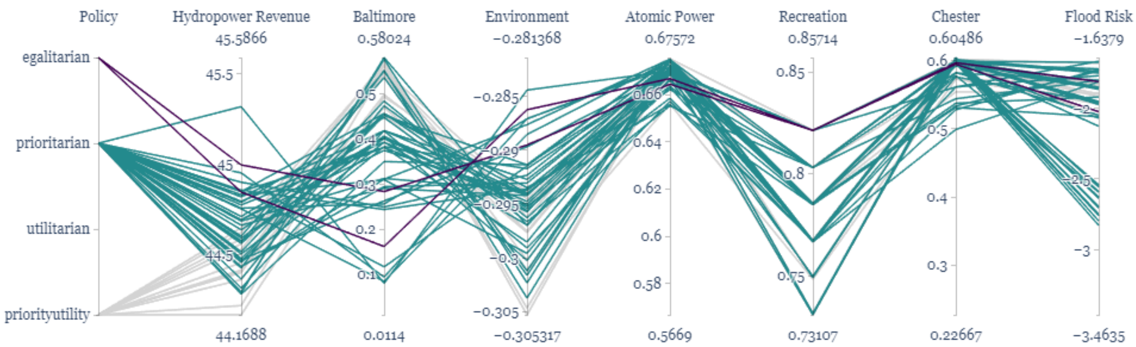


Figure A.9: Top 95th percentile outcomes when c_{zero} for chester is 20%

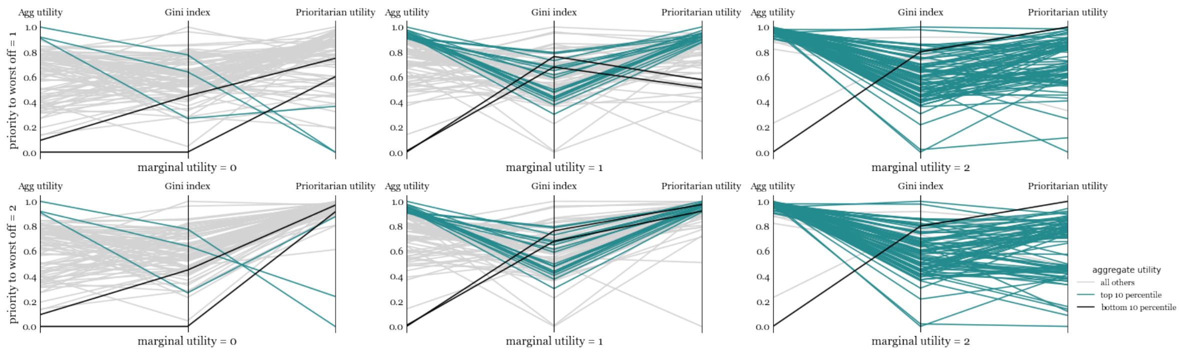


Figure A.10: Fairness principle tradeoffs when c_{zero} for chester is 20%

4. c_{zero} consumption of recreation is 50%

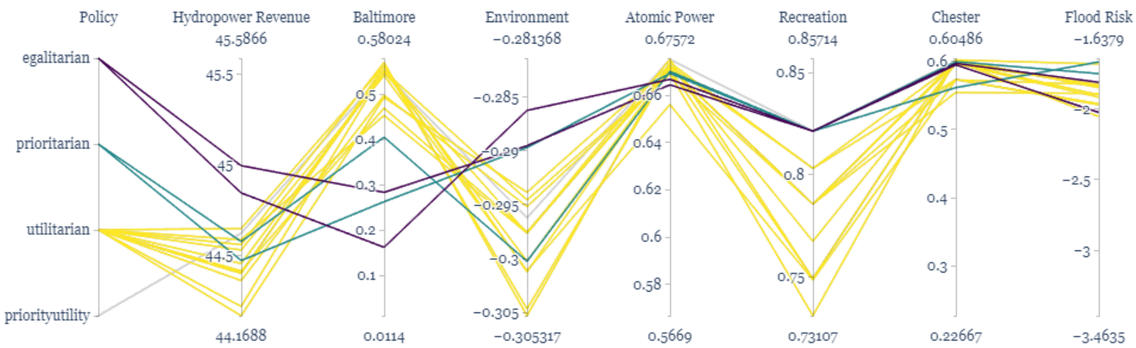


Figure A.11: Top 95th percentile outcomes when c_{zero} for recreation is 50%

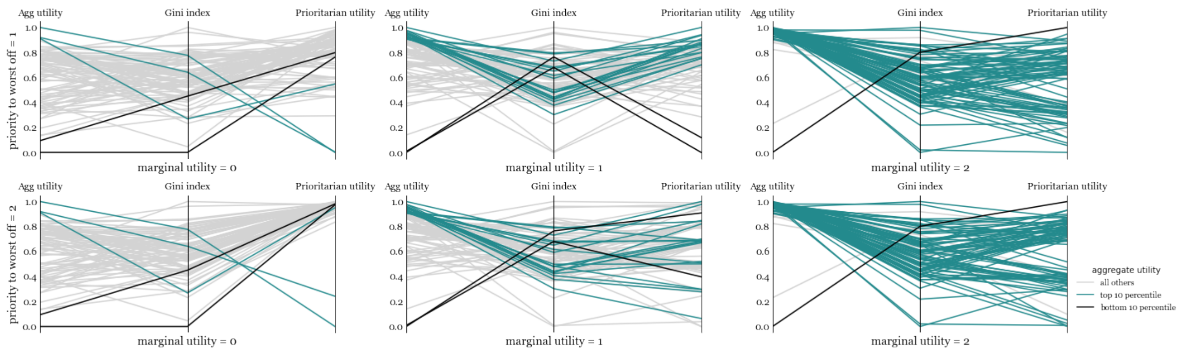


Figure A.12: Fairness principle tradeoffs when c_{zero} for recreation is 50%

5. c_{zero} consumption of environment is 40%

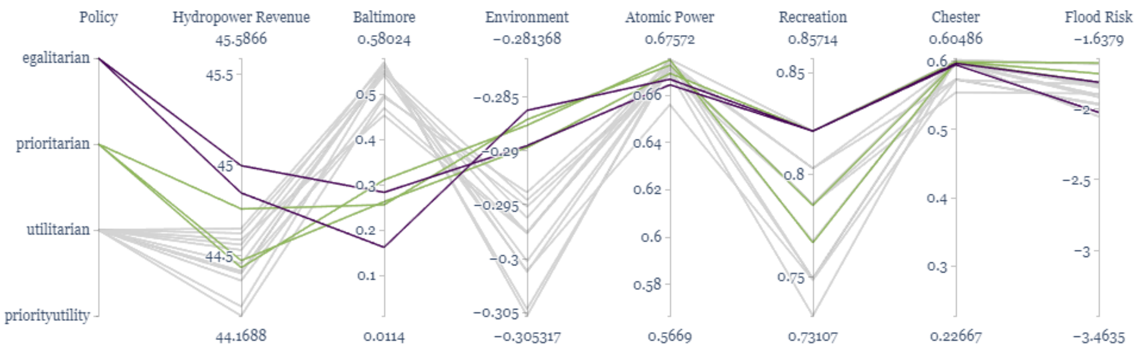


Figure A.13: Top 95th percentile outcomes when c_{zero} for environment is 40%

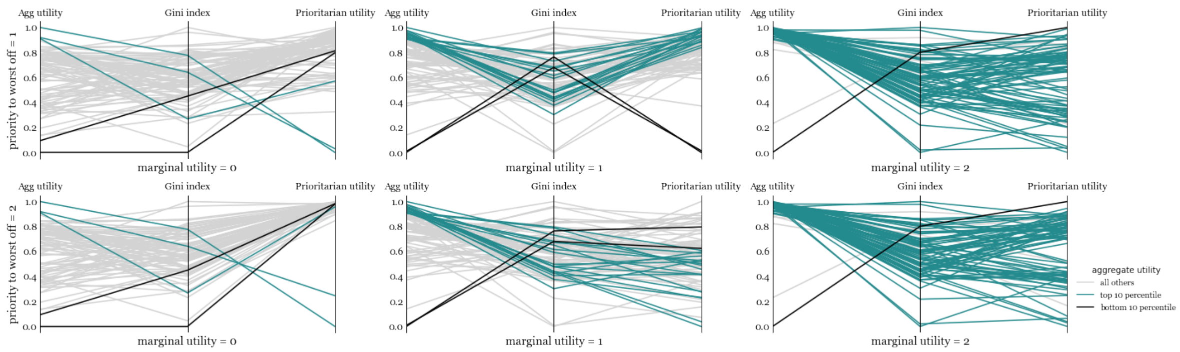


Figure A.14: Fairness principle tradeoffs when c_{zero} for environment is 40%

6. c_{zero} consumption of flood risk is 4ft

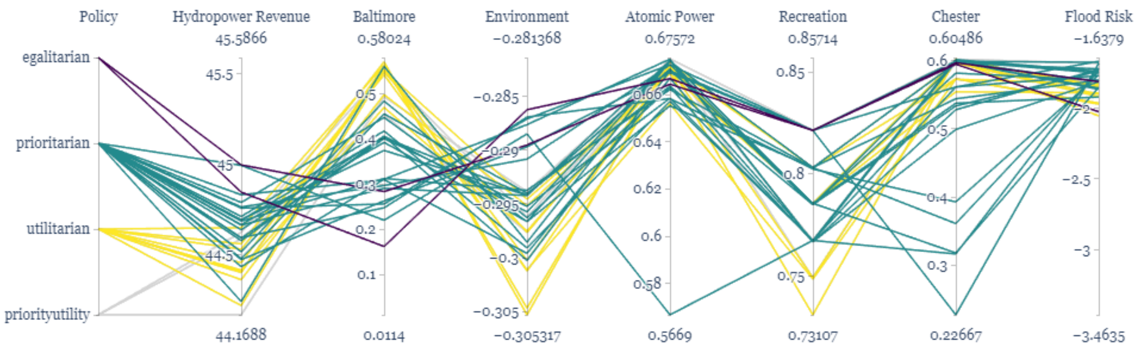


Figure A.15: Top 95th percentile outcomes when c_{zero} for flood risk is 4ft

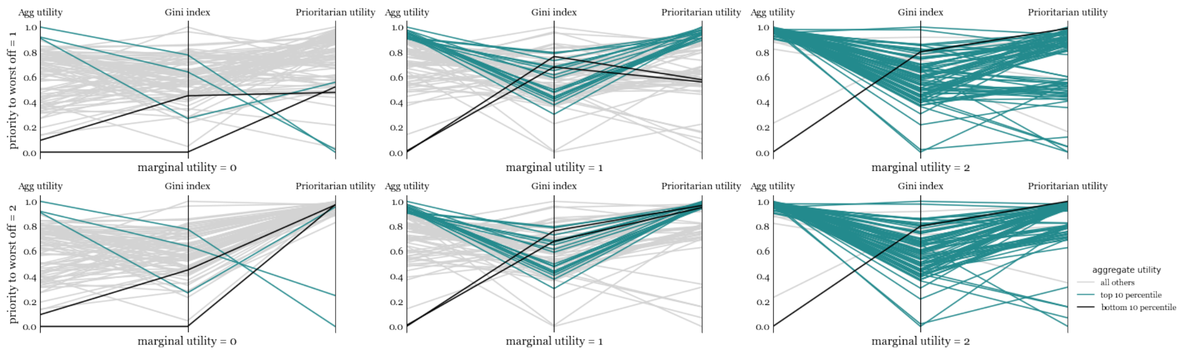


Figure A.16: Fairness principle tradeoffs when c_{zero} for flood risk is 4ft

The figure A.17 shows how the number of policies in the region of interest go down significantly when with fewer policies being high on prioritarian utility. Only the case of $\eta = 1.2$ and $\gamma = 1$ have policies that have high aggregate utility, priority and gini index.

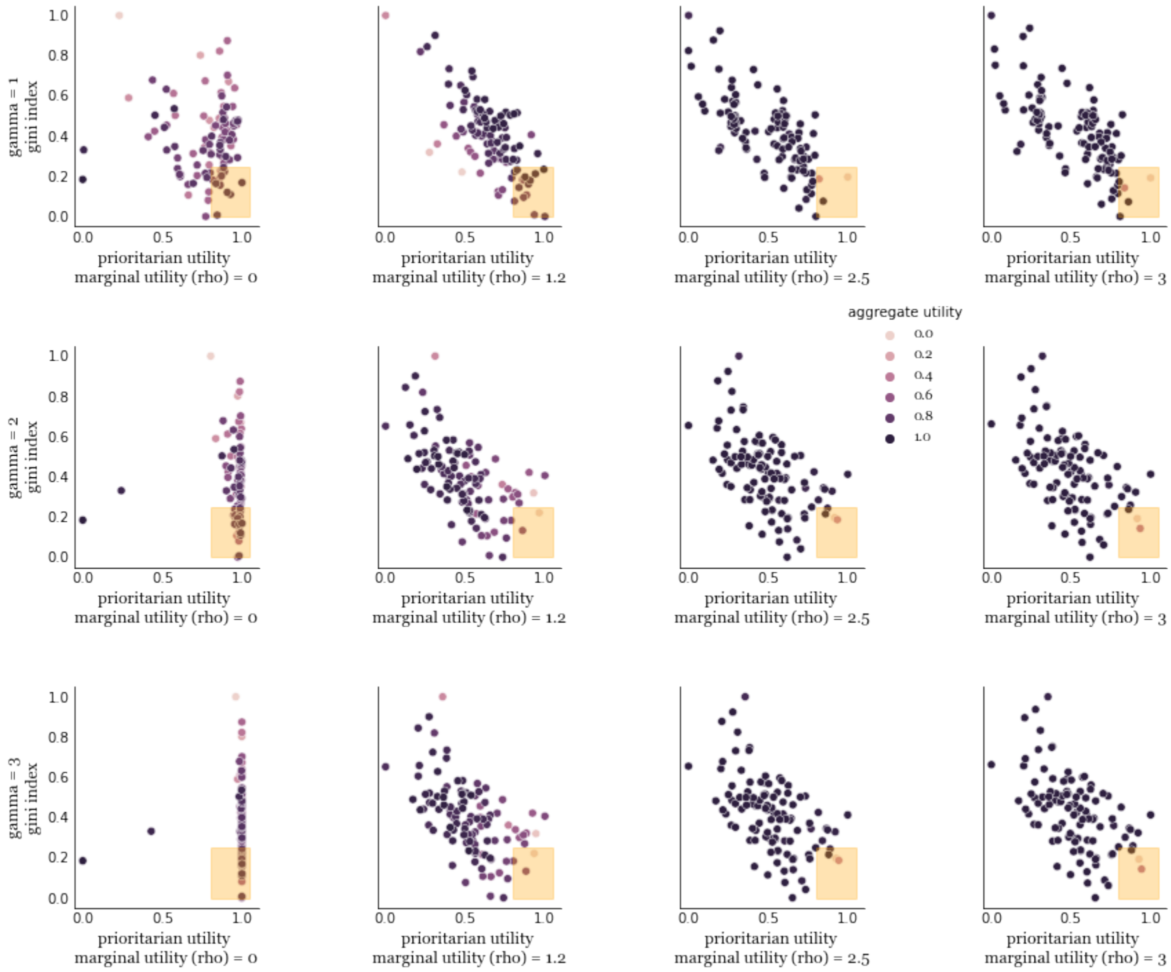


Figure A.17: Identifying policies when the zero consumption of hydropower is $mn 40

B

Technical improvements, limitations and future work

All the operationalization for this research has been performed on python. The notebooks and data files associated with it are all present on the Github repository mentioned in the beginning of the thesis. This chapter of the appendix aims to give an overview of the improvements made within the work that was borrowed from other researchers and identifies a few potential lines of future work.

B.1. Susquehanna Model on Python

The original Susquehanna model borrowed from (Salazar et al., 2016) was in C++. We moved it from C++ to python through a luft and shift process.

Technical Improvements

- The physical constrains of the Susquehanna river basin were updated based on the latest information used in the work of (Doering et al., 2021).
- The RBF.py file was modified to be vectorized in order to improve performance
- The model was connected to the EMA workbench in order to run perform multi objective optimization (Kwakkel, 2017).

Limitations and Future work

- The model is currently extremely slow. The use of object oriented programming, and/or cython could significantly help with the performance.
- The model generates worst case scenarios using Monte Carlo simulation. These scenarios are currently being stored in data files. This can be connected to the EMA in order avoid that and also open doors to new samplign techniques for scenarios.
- The flood duration and flood risk in the model is currently measured in ft of water excess in the dam. Converting this to flood height would make it easier to estimate the intensity of the flood.

B.2. GMCR package

The Graph Model for Conflict Resolution was created by Fang et al. (2003). In their subsequent versions, they release a python package available on GitHub https://github.com/onp/gmcr-py.git. They also have an application with a user interface available on request https://www.eng.uwaterloo.ca/ rkinsara/index.html.

Technical Improvements

- Modifying the python package allows for automation of running the games which is why the .py file was used. The original code allows actors not to make a move at all. This was modified for the current version of the package. However, it can be used if there is a case that requires it.
- The entire process of running sequential games and including and not including different actors is automated and available in a Jupyter notebook titled non_cooperative_stability.ipynb.
- Preferences of actors was modified to include equally preferred states.

Limitations and Future Work

- Currently, the package does not consider the weight of the preferences of actors. This can be modified within the code as part of future work.
- The GMCR package can be modified to map multiple objectives to a single actor, prioritize the objectives and generate new preference rankings. This would be valuable in reducing the number of players and making the game a little more realistic. Such an approach has been used before by Madani and Lund (2011b).