Universität Kassel

On the Cooperation of Groups in Social Dilemmas

Dissertation zur Erlangung des Grades

Doktor der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.)

am Fachbereich Wirtschaftswissenschaften der Universität Kassel

von Sven Christens

Erstgutachterin: Prof. Dr. Astrid Dannenberg, Universität Kassel

Zweitgutachter: Prof. Dr. Guido Bünstorf, Universität Kassel

Drittgutachter: Prof. Dr. Ivo Bischoff, Universität Kassel

Tag der Disputation: 07.07.2020

On the Cooperation of Groups in Social Dilemmas

Doctoral thesis to fulfill the requirement for the degree of

Doktor der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.)

Submitted to the Department of Economics at the University of Kassel

By Sven Christens

Acknowledgments

This dissertation would not have been possible without the support of a number of people who I would like to thank.

First and foremost, I would like to thank Astrid Dannenberg for the supervision of this dissertation, the scientific advice and guidance over the past years and the numerous possibilities she provided for me to learn and to develop my ideas.

I am particularly thankful to my office mate Sonja Zitzelsberger and my colleagues Corina Haita-Falah, Elina Khachatryan, Marcel Lumkowsky and Philipp Händel for the fruitful discussions and the willingness to help whenever it was needed. I would also like to thank Heike Wetzel and her chair, in particular Victor von Loessl, Jan Kallok and Jonas Bender. I am also thankful for Christiane Großberndt's support. Furthermore, I would like to thank Andreas Ziegler, Gunnar Gutsche, Elke Groh and Daniel Engler. I would also like to thank our student assistants, Arne Vilmar, Eva Weingärtner and Mario Sicenica, who bravely endured each cumbersome task I had for them. More generally, I am thankful to all members of FB07 and GS EBGo at the University of Kassel and in particular to Guido Bünstorf and Ivo Bischoff for agreeing to review this dissertation.

I would like to thank my co-author Florian Sachs for the fun time during the experimental sessions in Magdeburg.

I am also grateful to Aaron Wolf, Todd Jarvis, Melissa McCracken, Maureen Walschot and Gerald van Belle for welcoming me at Oregon State University and for their support and ideas. I would also like to thank Scott Barrett for helping to make the visit possible.

My research has also benefited from ideas and scientific input of Matthias Klein, Pablo Duarte and Simon Melch.

Lastly, I would like to thank my parents, sisters, and family. In particular, I am grateful to my son, Lasse Johann, for bringing such joy to our lives and to my wife, Svenja, for her patience and loving support.

Summary

Mankind faces numerous environmental problems such as climate change, air and water pollution, biodiversity loss and resource overuse. Common to all these problems is that they can be characterized as social dilemmas and can only be solved if a large number of people and groups cooperates. This dissertation provides an analysis of the cooperation of groups in four separate studies. It consists of two parts. Part I presents two laboratory experiments that examine the effect of identification on individuals and groups. Part II empirically examines the formation and the impact of treaties between states that share an international river basin.

In the first study, my coauthors and I examine the effect of identification on individuals and groups in a laboratory public goods experiment. Previous experimental studies have shown that revealing the identity of players and their decisions in public goods games increases contributions to the public good. We experimentally investigate whether this finding also holds true when the decision is made by groups. We use digital photos to identify players to each other. The groups communicate either face-to-face or via a computer chat. The results confirm the positive effect of identification on the contributions of individuals. For groups we find only a small and temporary effect. However, groups contribute relatively much to the public good even in the anonymous game compared to individuals. Chapter 3 presents a study that further investigates the effect of identification. Individuals and groups of three receive an endowment and subsequently are asked if they would like to make a donation from this endowment. Both individuals and groups, make the decision either anonymously or their pictures and donations are shown to observers. The results show that women and men differ in their behavior. Women make relatively high donations regardless of whether they decide alone or in teams or whether they are observed or not. When deciding alone, donations by men increase significantly when they are observed to similar levels as donations made by women. In contrast, men do not increase their donations when being identified as team members.

In Part II, I examine cooperation between states in international river basins. A large proportion of the world's population suffers from water shortages. Due to changes in the climate and a growing world population, this proportion will increase even further. Rivers play a vital role in providing humans with freshwater. Many of these rivers are shared by two or more countries. Conflicts between countries are in particular common where rivers flow from one country to another (as opposed to rivers that form the border between countries). Treaties that regulate the allocation of water from upstream riparian states to downstream riparian states are important tools to formalize cooperation in international river basins and to provide downstream states with greater security over their water supply. While downstream states are the main beneficiaries of these treaties, upstream states generate little gains from them. Still, numerous allocation treaties were signed. In Chapter 4, I examine what factors lead to cooperation in the form of allocation treaties between states in internationally shared river basins. I analyze in particular the effect of the downstream state's power relative to that of the upstream state. In addition, local basin data is used to control for the upstream and the downstream states dependency on the river's water and for the construction of dams that are a common cause for disputes. It is shown that the downstream state's relative power is not the driving factor in the formation of treaties. The downstream state's available water, its exposure to floods and droughts and dam constructions in the upstream country are far more decisive. In the final and fifth chapter, I test whether treaties that aim at reducing the pollution in international river basins are effective in doing so. The results indicate that states address increasing pollution in international rivers by signing treaties. When controlling for endogeneity in an IV approach, treaties are found to significantly reduce pollution.

Content

List of Tables

List of Figures

List of Abbreviations

BOD – Biological oxygen demand
DO – Dissolved oxygen
EKC – Environmental Kuznets Curve
ERC – European Research Council
EU – European Union
FAO – Food and Agriculture Organization (of the United Nations)
GDP – Gross domestic product
GDP p.c. – Gross domestic product per capita
GEMS – Global Environment Monitoring System
GF – General (water) framework
HUCO – Human Cooperation to Protect the Global Commons
IV – Instrumental variable
IWRM – Integrated Water Resource Management
MWW – Mann-Whitney-Wilcoxon
OLS – Ordinary least squares
RBO(s) – River Basin Organization(s)
TWAP – Transboundary Water Assessment Programme
UN – United Nations
US(A) – United States of America
WQA(s) – Water quality agreement(s)

1 Introduction

Climate change, air and water pollution, biodiversity loss and resource overuse are only some of the numerous environmental problems faced by mankind. All of these problems have in common that they can be characterized as social dilemmas. Social dilemmas are characterized by two properties: Firstly, choosing defection yields a higher individual payoff than cooperation and, secondly, individuals are better off if everybody cooperates (e.g. Dawes 1980). What makes solutions such as the sustainable exploitation of resources, environmental policies and the provision of public goods in particular challenging is the large number of people that needs to cooperate. The number of people involved increases the transaction costs of the organization and makes coordination more difficult (Ostrom 2009). What adds to the problem and points out the importance of this research field is that most decisions are made by groups, such as firms, households or entire countries, rather than by individuals. This makes cooperation even more difficult. Experimental studies have shown that groups behave significantly different from individuals. Groups tend to be less driven by social considerations and to focus stronger on their payoff (Charness and Sutter 2012). Therefore, it is important to evaluate whether solutions that are effective in improving social behavior of individuals and, thereby, the overall welfare, can also be applied to groups that make decisions.

The four studies presented in this dissertation are organized in two parts. The first part of this thesis analyzes if making the actions of decision makers observable to others can increase the pro-social behavior of groups. Chapters 2 and 3 present experimental studies that examine the effect of identification on individuals and groups. The study presented in Chapter 2 uses a public goods game in order to examine whether identification is an effective tool to overcome social dilemmas. Previous studies have shown that making the actions of individuals observable to others increases pro-social behavior (Andreoni and Petrie, 2004; Rege and Telle, 2004). Our experiment tests if this finding holds true when decisions are made by groups rather than by individuals. The third chapter presents a follow-up study in which we examine the role of identification on groups in the context of charitable donations. This setting allows us to ensure that the expectations of individuals and group members are consistent and to keep also the number of people observing the identities constant. Therefore, possible differences in the behavior should be purely driven by the protection that the group offers to the members.

The second part of this dissertation turns to the cooperation of states in international river basins. One of the major global challenges is the adequate provision of freshwater to the world's population. It is estimated that by 2050 around 52% of the world's population will be affected by water stress (Schlosser et al. 2014). The cooperation of people and states will be essential in order to secure the provision of freshwater. International rivers play a crucial role in the provision of water as their basins account for 80% of all global river flows and present a crucial source of freshwater (Food and Agriculture Organization of the United Nations 2019). A major global challenge is the allocation of water between countries that share a common river, especially in geographical constellations with an upstream and a downstream state. Upstream riparian states can almost unilaterally withdraw the river's water and, thereby, create severe externalities for the downstream riparian states. The use of water from international rivers is a major reason for conflicts between states (Yoffe, Wolf, and Giordano 2003). Still, most countries are able to solve these conflicts and to engage in some degree of cooperation (Yoffe et al. 2004; Bernauer and Böhmelt 2020). This cooperation often takes the form of international treaties that regulate the allocation of water between states. In Chapter 4, I examine the factors that lead to the formation of such treaties. Another major problem in international rivers is pollution. In the final and fifth chapter of this dissertation, I examine whether international agreements that address water pollution in international rivers are effective in reducing this pollution. Table 1.1 provides an overview of the studies presented in this dissertation. The four studies are presented in Chapters 2 to 5 after this introductory chapter.

Table 1.1 Schematic overview of the dissertation

Part	Chapter	Co-Authors	Method
I. Identification of individuals and groups	2) Identification of individuals and groups in a public goods experiment	Astrid Dannenberg; Florian Sachs	Laboratory experiment
	3) The effects of identification on prosocial behavior of individuals and teams	Astrid Dannenberg; Florian Sachs	Laboratory experiment
II. Cooperation in international river basins	4) What Drives States to Sign Water Allocation Treaties in International River Basins?	-	Empirical study
	5) Do international agreements improve water quality in rivers?	-	Empirical study

1.1 Groups and cooperation

In recent decades, behavioral economists examined the validity of many assumptions made in standard economics. Experimental studies have demonstrated that decision makers rarely act in accordance with game theoretic predictions. The field revealed many aspects of the limits of rationality, behavioral biases and cognitive limitations of humans. To a large extent human behavior is not just driven by limitations and biases but also by distributional and fairness preferences. Behavioral economics has confirmed the positive impact on cooperative behavior of numerous factors such as communication or punishment. However, as in standard economic theory that often considered groups of people simply as uniform actors, behavioral economists have only relatively recently started to examine the decision-making of groups (Charness and Sutter 2012).

Laboratory experiments have demonstrated that groups differ decisively from individuals when making decisions. Groups tend to behave closer to the standard economic predications than individuals (Charness and Sutter 2012). This means that their behavior has a stronger focus on monetary gains than on other considerations. Groups are also found to be less influenced by behavioral biases and fairness considerations. Groups transfer, for example, less money in dictator games than individuals (Luhan, Kocher, and Sutter 2009), make more selfish decisions in prisoners' dilemmas (Wildschut et al. 2003), send less money in trust games (Kugler et al. 2007) and accept lower offers in ultimatum games (Bornstein and Yaniv 1998). Groups also tend to use less costly punishment in public goods games (Auerswald et al. 2018). At the same time, groups tend to make cognitively more sophisticated decisions and show better reasoning. For instance, groups learn quicker in the beauty contest game (Kocher and Sutter 2005). Social psychology refers to the phenomenon of groups being more competitive and less social than individuals as the individual-group discontinuity (Wildschut et al. 2003).

Despite the tendency of groups to behave more selfishly than individuals in many situations, important mechanisms and decision structures exist that offset this tendency. It has been shown that concerns for the long-term outcome can reduce groups' tendency to behave more competitively and less selfishly (Insko et al. 1998). Groups are, for example, more willing to cooperate under uncertainty (Rockenbach, Sadrieh, and Mathauschek 2007). Groups generate also higher gains in a Stackelberg market game because they show higher levels of cooperation (Müller and Tan 2013). This behavior might explain groups' higher contributions in public goods games that are played for several rounds (Auerswald et al. 2018).

Because of this complex, sometimes ambiguous behavior of groups, it is important to carefully examine if mechanisms, which are able to increase the cooperative behavior of individuals, can also increase the social behavior of groups. For instance, communication between individuals has been shown to significantly increase cooperative decisions by individuals. However, communication between groups does not increase cooperation (Insko et al. 1993). Another mechanism that has been proven to increase the cooperative behavior of individuals is revealing the players' identities (Andreoni and Petrie 2004; Rege and Telle 2004). We examine in Chapter 2 and Chapter 3 if this finding also holds true for groups.

1.2 Cooperation in international river basins

Environmental problems present a field in which particularly large groups need to cooperate in order to arrive at socially optimal solutions. No supranational organization sets environmental standards or rules. States rely on bilateral and multilateral agreements in order to define rules on the emission of pollutants or the use of resources. Cooperation is usually codified as environmental agreements. In particular, since 1945 an increasing number of such agreements has been signed. The reason for the strong increase in the last decades can be found in the increased demand for institutions that address international environmental problems (Barrett

2003, 136). The most severe and complex challenge in this field is certainly global warming. Virtually all countries in the world are required to coordinate and to cooperate in order to stop it.

International river basins present another critical field in which groups and countries need to cooperate. Cooperation over internationally shared rivers is in particular interesting and different from other environmental challenges. The interdependence of states sharing a river is very high. While states can compensate emissions created by non-complying states in other international emission scenarios like greenhouse gas emissions or ocean pollution, this is not possible in rivers. A single defector can easily destroy the effort of other riparian states to mutually and sustainably manage a river. In addition, the relation between upstream and downstream riparian states is highly asymmetric. The overconsumption of water and pollution spillovers in the upstream countries create severe negative externalities for downstream countries (Toset, Gleditsch, and Hegre 2000). Upstream countries, in contrast, are largely unaffected by the actions of downstream states.

Against the background of these challenges and an increasing number of people and countries suffering from water scarcity, the former United Nations Secretary Boutros-Ghali warned in 1985 that "[…] the next war in the Middle East will be fought over water, not politics".[1] Indeed, the allocation of water and infrastructure projects are common causes for conflicts (Yoffe et al. 2004). However, the warnings of "water wars" have not become reality so far (Brochmann and Gleditsch 2012b). While ongoing tensions and conflicts exist, especially in regions like the Middle East and Central Asia, cooperation in international river basins is the norm (Yoffe et al. 2004; Bernauer and Böhmelt 2020). This cooperation is typically codified in treaties and agreements. Hundreds of treaties over international river basins have been signed already (Giordano et al. 2014).

Previous studies identified a number of factors that make the signature of treaties in international river basins more likely. Factors such as common trade, power imbalances, political factors, water scarcity and the geographical constellation have been proposed and examined (Espey and Towfique 2004; Song and Whittington 2004; Tir and Ackerman 2009; Dinar, Dinar, and Kurukulasuriya 2011). In Chapter 4, I will add to this literature by examining whether power is a relevant factor for the signature of water allocation treaties when set in relation to the geographical positions of the riparian states. Moreover, I introduce a number of local factors, such as dam constructions and the exposure to floods and droughts, that have not been tested before.

While several studies helped to understand what drives the signature of treaties, the literature on the effectiveness and impact of treaties is still very scarce. Previous studies focused mostly on the question of whether treaties are able to reduce conflicts between states in river basins. Treaties are found to be an effective way to reduce tensions if they are designed well

[1] http://news.bbc.co.uk/2/hi/talking_point/2951028.stm (accessed: March 2020)

(Brochmann 2012; Dinar et al. 2019). It has also been shown that agreements between upstream and downstream countries reduce potentially harmful dam constructions in the upstream basin (Olmstead and Sigman 2015). However, the question if treaties improve the environmental conditions of the rivers still presents a large gap in the literature. Whether treaties are an effective tool to improve water quality, to increase water flow or to smoothen the flow is largely unknown. One of the few studies on the subject is presented by Köppel and Sprinz (2019) who find that treaties mitigate pollution from upstream states to downstream states in Europe when they are legally binding. Bernauer and Kuhn (2010) find that a higher number of environmental treaties that a country signed is related to better water quality in European rivers. However, no effect of water quality agreements is found. The lack of studies in the field might be explained by two main challenges that arise. There is a lack of data on water quality and flow volumes and it is difficult to address the possible endogeneity problems resulting from the self-selection of countries and river basins into treaties. In the final and fifth chapter, I will try to close this gap by examining the effect of treaties on water quality globally.

1.3 Methodology and main results

This work comprises of four studies. All studies are part of the ERC Starting Grant project *Human Cooperation to Protect the Global Commons* (HUCO). Two of the studies are laboratory experiments that examine the effect of identification on the behavior of individuals and groups. These two studies were written in collaboration with Prof. Astrid Dannenberg and Florian Sachs. The objective of these studies is to examine the difference between individuals and groups in their reaction to identification. The subsequent two studies examine the formation and effect of agreements over international rivers. These two studies are single-authored. In the following, I will present the research questions of the studies and the main findings.

In the first study, presented in Chapter 2, we examine the effect of identification on individuals and groups in a public goods experiment. The game is played either by individuals or by groups of four that decide unanimously on their contribution to the public good. The same players play the game for ten rounds. Teams communicate either face-to-face or anonymously via computer chat. We first examine contributions of individuals and groups when no identifiable information about players is provided. In the identified treatments, we enable the participants to observe the identities of the other players and their contributions. Following the approach of Andreoni and Petrie (2004) and Samek and Sheremeta (2014, 2016), we use digital photos to identify players. At the end of each round, the contributions of the players are shown along with the photos of the individuals or teams. Our results confirm the finding that identification increases pro-social behavior of individuals (Andreoni and Petrie 2004). The anticipation of approval or disapproval is sufficient to increase the contributions of individuals. In particular, individuals who strongly care about their image increase their contributions when they can be identified. Groups, in contrast, react significantly less sensitive to the revelation of their identities than individuals.

Their contributions do not increase significantly when the photos are displayed. However, irrespective of the type of communication, groups contribute more than individuals in the anonymous treatments. Especially at the beginning of the game when no information on the behavior of other players is available, groups are more willing to make higher contributions than individuals. The difference in the reaction to identification is in particular pronounced between individuals and teams that communicate face-to-face. The effect of the pictures on groups is not significantly enhanced when the team's members are concerned about their image. Being part of a team appears to help subjects to overcome their image concerns when they can be identified by others.

In the second study, we examine the effect of identification on groups more closely in another laboratory experiment. Individuals and groups of three are offered the possibility to make a donation to the Red Cross from an endowment that they receive. Group members do not meet but submit proposals through the computer and then vote on the proposals until a unanimous decision is reached. This procedure allows us to observe the initial proposal of each team member. As this proposal is made without any information regarding the preferences of other participants, the proposals can be considered as the subject's individual preference. As in the first study, the individuals and groups make their decision either anonymously or their donation and identity can be observed by others. Again, we use digital photos to identify individuals and teams in the identified treatments. After the donation decision is made, four observers see the donations along with the photos of the players. This design allows us to keep the number of participants who see the donations and the pictures constant between individual and team treatments. In addition, we provide individuals and teams with the same reference point of what uninvolved persons consider to be an appropriate donation. This procedure aims at aligning the moral standards of individuals and teams. We find that women are relatively generous regardless of whether they decide as individuals or in a group and whether they are observed or not. Men donate significantly less than women as individuals in the anonymous treatment. They become significantly more generous when they are observed but only when deciding individually. This change is mostly driven by a strong decrease in the number of men who donate zero. In the team treatments, identification does not increase donations made by men. Men seem to be motivated by image concerns to a larger extent. However, identification is only effective when men can claim full responsibility for their actions.

The third study empirically analyzes what factors lead to the cooperation of states over international rivers. More specifically, I test which factors lead to the signature of allocation treaties that regulate the allocation of water from the upstream state's basin to downstream states. These treaties are interesting because upstream and downstream states are highly asymmetric in their ability to consume water. Upstream states can almost unilaterally withdraw the river's water. Yet, upstream states agreed in numerous cases to guarantee downstream state's fixed or flexible amounts of water. In the study, I test the role of the downstream state's power relative to that of the upstream state. I also introduce local variables that control for the available water, the exposure to droughts and floods and the construction of dams. The results show that the economic and military power of the downstream state, who is the main beneficiary of a treaty, is not the decisive factor in the formation of such treaties. However, in water scarce regions, the downstream state's relative military power is a relevant factor. Far more important than power is the downstream state's dependency on the water. Allocation treaties become more likely when the downstream state suffers from water scarcity and a high exposure to floods and droughts. Moreover, dam constructions in the upstream country are a common reason for the signature of allocation treaties.

In the fourth and final study, I empirically examine the effectiveness of transboundary agreements. I focus on water quality agreements that aim at reducing pollution in international river basins. Treaties tend to be signed in situations when the pollution in rivers increases. This demonstrates that states consider them an adequate tool to deal with the problem. Testing for strict exogeneity reveals that simultaneity between water quality and the signature of agreements exists. Therefore, it is important to control for endogeneity in order to estimate a causal effect of agreements. I use the prior signature of border and navigation treaties as an instrument in an IV approach in order to address the problem. Water quality in international river basins is found to significantly improve after the signature of water quality agreements.

1.4 General lesson

The research questions examined in this dissertation are relatively wide-ranging. Are there still general lessons that can be drawn from the different chapters? Yes, to some degree that is the case. Throughout the studies presented in this dissertation a more optimistic picture of human cooperation in social dilemmas can be found than standard economic theory would predict. This tendency of humans to cooperate and the existence of social norms and compliance with them is well documented (for example by Ostrom 1990; Fehr and Schurtenberger 2018). However, past research focused mostly on the behavior of individuals. The behavior of groups is less studied and there is good reason to assume that groups behave less cooperative than individuals in many situations (Charness and Sutter 2012; Kugler, Kausel, and Kocher 2012). Nevertheless, the results of this dissertation suggest that also groups and even countries are willing to cooperate under the right circumstances. Individuals and groups make contributions to the public goods well above the Nash equilibrium and, thereby, generate higher profits. Groups show even relatively high levels of contributions when they make their decision anonymously. One explanation for this behavior is the aim to make initial investments that signal the willingness to cooperate in repeated interactions. Even in the donation experiment, individuals and groups donate a considerable amount of their endowment although this reduces their payoff and generates no long-term gains.

Cooperation in international river basins is also common, despite alarming warnings of water wars as a result of an increasing water scarcity. The results of Chapter 4 suggest that upstream states acknowledge the dependency and the vulnerability of downstream states and are willing to sign allocation treaties, even if downstream states are not powerful. The reason for this cooperation might also be found in the advantages and the long-run benefits that can be realized with cooperation. Furthermore, states are also willing to sign treaties over the quality of water when the pollution in international rivers increases. The estimations suggest that treaties are effective in bringing down pollution.

2 Identification of Individuals and Groups in a Public Goods Experiment

Sven Christens[a], Astrid Dannenberg[a], and Florian Sachs[b]

[a] Department of Economics, University of Kassel, 34109 Kassel, Germany

[b] Faculty of Economics and Management, University of Magdeburg, 39106 Magdeburg, Germany

This chapter has been published as:

Christens, Sven, Astrid Dannenberg, and Florian Sachs. 2019. "Identification of individuals and groups in a public goods experiment." *Journal of Behavioral and Experimental Economics* 82.

Abstract: Revealing the identities of contributors has been shown to increase cooperation in public goods games. In this paper we experimentally investigate whether this finding holds true when decisions are made by groups rather than individuals. We distinguish between groups in which members can discuss face-to-face to reach a decision and groups in which members communicate via computer chat. The results confirm the positive effect of identification on cooperation among individuals. For groups, however, we only find a small and temporary effect of identification, irrespective of the type of communication.

2.1 Introduction

People often behave more socially oriented when their actions are observable by others. For example, visitors to a national park contribute more to the maintenance of the park when the solicitor can see the contribution (Alpizar, Carlsson, and Johansson-Stenman 2008). Energy consumers are more likely to take part in a blackout prevention program when their decision can be observed by their neighbors (Yoeli et al. 2013). Church goers donate more when the offering can be observed by the persons sitting next to them (Soetevent 2005). Citizens are more likely to vote when there is a high chance that this decision is observed by others (Funk 2010). Experiments have shown that cooperation in public goods games increases when the players' identities and their actions are revealed to the other players (Andreoni and Petrie 2004; Rege and Telle 2004).

Given people's sensitivity to identification and observance by others, increasing transparency can be an effective way to improve cooperation. For instance, Finland publishes information about its citizens' taxable income in order to reduce tax avoidance.[2] But does this result also hold for group behavior? Are groups, like individuals, sensitive to the observance and potential approval or disapproval by others? To the best of our knowledge, these questions have not been answered yet, although many important economic, financial, and political decisions are in fact made by groups rather than individuals. Examples include decisions by households, firms, governments, delegation teams, nongovernmental organizations, or unions. It has long been known in social psychology that groups behave differently than individuals. The concept of "deindividuation" describes how anonymity and diffusion of responsibility within groups lead to less restrained and more impulsive and aggressive behavior (Festinger, Pepitone, and Newcomb 1952; Zimbardo 1969). Groups show more competitive behavior in social dilemma situations than individuals which has been labeled the "interindividual-intergroup discontinuity effect" (Insko and Schopler 1998; Wildschut et al. 2003). A growing behavioral economics literature shows that groups learn more quickly, make more sophisticated and payoff-oriented decisions, and are less influenced by cognitive limitations, behavioral biases, and social considerations (for reviews see Charness and Sutter 2012; Kugler, Kausel, and Kocher 2012).

Despite the growing interest in group behavior, little is known about the effects of identification on groups. Previous research has shown that, while identification of individuals intensifies emotions and moral reactions towards them (Kogut and Ritov 2005a; Small and Loewenstein 2003, 2005), identification of groups does not have the same effects (Kogut and Ritov 2005b). However, this research only shows how observers respond but not how groups themselves respond to their identification. Shepherd, Spears, and Manstead (2013) show that members of a group feel less shame for a questionable group decision the more they identify with the group.

[2] Barry, E., 2018. Happy 'National Jealousy Day'! Finland Bares Its Citizens' Taxes. The New York Times online. https://www.nytimes.com/2018/11/01/world/europe/finland-national-jealousy-day.html (accessed November 2, 2018).

This study, however, neither provides a comparison between individuals and groups nor a comparison between identified and unidentified actions.

In this paper, we investigate the effects of identification on cooperation among individuals and groups in a controlled experimental setting. This approach allows us to create clear counterfactual situations without identification and to compare the behavioral responses of individuals and groups. We first compare the willingness to cooperate of individuals and groups in a finitely repeated public goods game in which no identifiable information about players is displayed. The public goods game is played either by four individuals or by four groups consisting of four individuals each and acting as a unitary player. In half of the groups, members communicate face-to-face to reach a decision while in the other half of the groups, members communicate anonymously via computer chat. All groups are required to discuss the problem and make consensus decisions. We then increase the transparency in the game by revealing the players' identities and actions. Following the approach of Andreoni and Petrie (2004) and Samek and Sheremeta (2014, 2016), we use digital photos to identify individuals and teams. At the end of every round, photos of the individuals or teams are displayed along with their contributions to the public good. Subjects know about this procedure in advance and can adjust their contributions if they wish. The two types of communication among group members, face-to-face and computer chat, represent two different group decision processes. Face-to-face communication represents a process where members openly discuss the available strategies and jointly make a decision. The members of the group know each other and what each of them contributes to the final decision. Computer chat communication represents a process where the members of a group decide jointly but the individual members' input remains anonymous. Therefore, they do not have much more information than outsiders.

We predict that groups react less sensitively to the disclosure of identities than individuals do. As members of a group share the responsibility and accountability for a decision, they do not feel singled out for doing something inappropriate. They can support and convince each other that they have made an appropriate decision for which there is no need to be concerned about others' disapproval. This opportunity does not exist for an individual decision maker. Furthermore, groups may expect less (unspoken) disapproval from other groups than individuals expect from other individuals. It is well known that individuals dislike being the "sucker" and that they get angry and frustrated when they have been exploited by others (Ahn et al. 2001; Kurzban et al. 2001). Individuals can thus be expected to strongly disapprove of free-riders. The feelings of frustration and anger may be less intense for groups because the members share the fate.

Our experimental results largely confirm the prediction. Revealing identities significantly improves cooperation among individuals while the effect for groups is relatively small and does not last, irrespective of the type of communication. Without identification, groups contribute more to the public good than individuals. In particular at the beginning of the game when it is unclear what the other players will do, groups are more willing to risk a high contribution.

However, groups increase their contributions only slightly and temporarily when their identities are revealed to the other groups. Individuals, by contrast, make significantly higher contributions when their identities are revealed and the increase persists. This confirms that revealing identities and the mere suspicion that others may disapprove of one's behavior constitute strong incentives for individuals to behave more socially oriented. The novel insight is that this effect is smaller for groups.

The remainder of the paper proceeds as follows: Section 2 provides the background for our study, summarizing previous findings from social psychology and behavioral economics. Section 3 explains the experimental design. Section 4 presents the results, Section 5 discusses the results and concludes.

2.2 Background

As background for our study, this section will summarize previous findings on the influence of identification and disapproval on social behavior and the differences between individuals and groups.

One of the most robust results of research on human cooperation is that cooperation improves when actions are observable. Making actions observable has significant positive effects in settings as diverse as blood donation, blackout prevention, support for national parks, church offerings, or voting in small communities (for reviews, see Rand, Yoeli, and Hoffman 2014; Kraft-Todd et al. 2015).

Also laboratory experiments have shown that cooperation in public goods games improves significantly when the players have to convey their contributions to the other players after the game (Rege and Telle 2004) or when a photo of them is shown along with their contributions (Andreoni and Petrie 2004). The photos have a much smaller effect when they are published without the contribution decisions (a similar result was obtained by Brosig, Weimann, and Ockenfels 2003). Building on these findings, Samek and Sheremeta (2014) show that the positive effect remains when only the two lowest contributors are shown, but disappears when only the two highest contributors are shown, indicating that shame associated with having given less than others is a stronger motivation than prestige which can be gained by contributing more than others. Similarly, allowing subjects to communicate their disapproval increases pro-social behavior. Subjects behave more cooperatively when they can send disapproval points (Masclet et al. 2003) or judgmental messages to each other after the game (López-Pérez and Vorsatz 2010; Peeters and Vorsatz 2013), even when the feedback has no direct effect on payoffs. The opportunity to give feedback also increases transfers in dictator games (Ellingsen and Johannesson 2008; Andreoni and Rao 2011). Taken together, this research suggests that, given individuals' sensitivity to observance and potential disapproval by others, increasing transparency can be an effective way of improving cooperation. It is not yet clear, however, if this is also the case for groups.

Recent reviews of the experimental literature have concluded that group behavior tends to be closer to standard game theoretical predictions than individual behavior (Charness and Sutter 2012; Kugler, Kausel, and Kocher 2012). For instance, groups have been shown to send less money in the trust game (Kugler et al. 2007), to make and accept smaller offers in the ultimatum game (Bornstein and Yaniv 1998), to behave more rationally in a bargaining game (Vollstädt and Böhm 2019), and to give less in the dictator game (Luhan, Kocher, and Sutter 2009). Groups have also been shown to be less cooperative in prisoners' dilemma games (Insko and Schopler 1998; Wildschut et al. 2003) or common-pool resource games (Gillet, Schram, and Sonnemans 2009). The lower cooperativeness has been explained by the ability of groups to justify selfish decisions (social support of shared self-interest hypothesis), to create a shield of anonymity and diffuse responsibility (identifiability hypothesis), and to anticipate the selfishness of other groups (schema-based distrust hypothesis). But there are also some reasons to expect groups to be more cooperative than individuals. It is well known that the fear to be exploited by others is an important barrier for individuals to cooperate (Ahn et al. 2001; Kurzban et al. 2001). Many people are conditional cooperators, meaning that they are willing to cooperate only if others do so, too. Thus, when it is unclear how the other players will act, cooperation is a risky decision. Groups have been shown to be better at handling risk than individuals (Rockenbach, Sadrieh, and Mathauschek 2007) and they may be more prepared to cooperate under strategic uncertainty. Also, as mentioned before, the feeling of being the "sucker" may be less disturbing for groups as it is shared among the members. Another possible reason why groups may be more cooperative than individuals is provided by the social comparison theory. According to this theory, people are motivated to present themselves in a more favorable way than they expect others to be (Cason and Mui 1997). An individual who chooses to free ride when deciding alone may be reluctant to recommend this action when discussing within a group. Finally, groups might be better able to reason through a repeated game, anticipate other players' behavior, and choose a strategy that gives a higher overall payoff. Müller and Tan (2013), for example, find that groups are more cooperative and earn higher payoffs in a repeated Stackelberg market game than individuals. So far, only two studies compared individuals and groups in a repeated public goods game. Auerswald et al. (2018) find that groups contribute more to the public good than individuals, whereas Huber, Model, and Städter (2019) do not find a significant difference between individuals and groups. The difference between the two studies may be explained by the different group size (3 versus 2). Both studies find that groups punish less and earn higher payoffs when the game includes a punishment mechanism. In short, although most studies point to more self-interest in groups, many aspects of group behavior are still not fully understood. This is clearly the case for group behavior in finitely repeated public goods games where only little research has been done so far.

Another relevant difference between individuals and groups pertains to how people perceive and react to their identification. Identified individuals generally evoke stronger emotions and moral reactions than non-identified individuals. This can lead to more generous behavior towards identified victims or more punitive behavior towards identified wrongdoers (Kogut and Ritov 2005a; Small and Loewenstein 2003, 2005). These effects of identification have not been found for groups (Kogut and Ritov 2005b). These findings support the conjecture mentioned above that there may be weaker and less emotional disapproval among identified groups than among identified individuals.

To the best of our knowledge, there is no study that has looked into the effects of revealing the identities of unitary groups. A few studies have explored related questions. Using a prisoners' dilemma, Insko et al. (1987) show that groups behave more cooperatively when, prior to the game, all members from both groups meet and discuss than when only two representatives meet. Identification might play a role for this positive effect of social contact but it is impossible to distinguish it from the other aspects of social contact such as communication or familiarity. In a related study, Schopler et al. (1995) find that groups cooperate more when they can hear not only the names and decisions from the members of their own group but also from the members of the opposing group. The difference to our study is that, instead of revealing the identity of the whole group as a unitary decision maker, the identities and decisions of the individual members are revealed. Another difference is that identification is done through voice and not a picture. Hauge and Rogeberg (2015) show that representatives who act on behalf of groups contribute more to a public good when there is a chance that they will have to make their decision public. This effect is stronger for men than for women. The difference to our study is that individuals do not make a decision within a group but on behalf of a group. This is an important difference because these decisions are still individual decisions and not group decisions.

2.3 Experimental design

We consider an n-player linear public goods game. In each round of the game (there is a finite number of repetitions), n symmetric players who are endowed with y tokens each may contribute to the production of a public good. Each player's contribution costs are assumed to depend only on the own contribution level while the benefits depend on the total provision of the public good. The payoff function for player i is given by

(1) $\pi_i = y - g_i + a \sum_{j=1}^{n} g_j$

where g_i is i's contribution to the public good with $0 \leq g_i \leq y$ and a denotes the constant marginal per capita return from contributing to the public good with $0 < a < 1 < na$. The full cooperative public goods contribution level that maximizes social welfare is given by $g_i^{FC} = y \; \forall i$. However, under the standard economics assumption of rational payoff-maximizing agents, the only subgame perfect Nash equilibrium in the finitely repeated game is given by $g_i^{NC} = 0 \; \forall i$. The Nash equilibrium involves dominant strategies such that each player's choice does not depend on the contribution levels chosen by the remaining players.

In all of our experimental treatments, $n = 4$ players played the public goods game for ten rounds with $y = 100$ and $a = 0.4$. Depending on the treatment, a player was represented either by an individual or a unitary group of four persons. We chose relatively large groups in order to give the members a real chance to hide within the group, even when identities are revealed. The experimental sessions were held in a computer lab (MaXLab) at the University of Magdeburg, Germany, using undergraduate students recruited from the general student population. The experiment was organized and recruited with the software hroot (Bock, Baetge, and Nicklisch 2014).

Table 2.1 Treatments

Treatment	Picture	Decision making	Communication within teams	Number of subjects	Number of observations
Indi-NoPic	No	Individual	-	40	10
Indi-Pic	Yes	Individual	-	40	10
F-Team-NoPic	No	Team	Face-to-face	160	10
F-Team-Pic	Yes	Team	Face-to-face	160	10
C-Team-NoPic	No	Team	Computer chat	160	10
C-Team-Pic	Yes	Team	Computer chat	160	10

Overall, 720 students participated in the experiment, whereby each student took part in one treatment only. We conducted six treatments which are summarized in Table 2.1: (1) a treatment in which players decided individually and no information about players was revealed (*Indi-NoPic*), (2) a treatment in which players decided individually and information about each individual's identity was revealed to all players (*Indi-Pic*), (3) a treatment in which players decided as a four-person team with face-to-face communication and no information about the teams was revealed (*F-Team-NoPic*), (4) a treatment in which players decided as a four-person team with face-to-face communication and information about each team was revealed to all players (*F-Team-Pic*), (5) a treatment in which players decided as a four-person team with computer chat communication and no information about the teams was revealed (*C-Team-NoPic*), (6) a treatment in which players decided as a four-person team with computer chat communication and information about each team was revealed to all players (*C-Team-Pic*).
Following the design of Andreoni and Petrie (2004) and Samek and Sheremeta (2014), we used digital photos to identify individuals and teams to one another. Digital photos show the appearance but do not allow for communication between players, which may confound the effects of identification alone. In addition to the photo, first names were included as part of the identification of players. Upon arriving at the lab, each subject got a printed name card with his or her first name and hold up the name card while the photo was taken. In the individual treatments and the team treatments with computer chat, we took a photo of each individual separately because players in the same group and members of the same team were not supposed

to meet each other. Team members in the treatments with face-to-face communication, on the other hand, were supposed to meet each other, so in these cases we took a photo of the whole team. Care was taken that the faces displayed on all photos had about the same size, so it was not the case that the individual photos showed subjects more prominently than the team photos (see Appendix for samples).

Participants in the individual treatments were randomly assigned into groups of four players to play the game and they stayed together for the ten rounds of play. Similarly, in the team treatments, teams of four persons were formed randomly and then four teams were randomly assigned into a meta-group to play the public goods game. The four persons within a team and the four teams within the meta-group stayed together throughout the game. In all treatments, contribution decisions in each round were made simultaneously. After all players made their contribution decisions, the total amount of the public good was displayed as well as the contribution made by each player or team, sorted from the largest to the smallest amount. In the treatments *Indi-NoPic*, *F-Team-NoPic*, and *C-Team-NoPic*, no additional information about the players was revealed (not even an ID number). In the treatments *Indi-Pic*, *F-Team-Pic*, and *C-Team-Pic* the names and photos of every individual or team were displayed next to their contribution. This way, each individual or team was recognized and also ranked according to their contribution to the public good from the largest to the smallest amount. In *C-Team-Pic*, the four individual photos were shown next to each other, jointly forming a team photo.

During the game, earnings were presented in tokens. In the individual treatments, 100 tokens converted to €1. In the team treatments, 100 tokens converted to €4 and earnings were distributed equally among team members. In each session, subjects were seated at linked computers to play the game (software z-Tree; Fischbacher 2007). In the team treatments with face-to-face communication, each team had its own room where the members could openly talk face-to-face. Each team member had his or her own computer. In the team treatments with computer chat, team members also had their own computer but they had no visible or other contact with each other, except of the anonymous computer chat which was open throughout the game.[3] In all team treatments, members of a team were asked to discuss the contribution decision in a civilized way (without using threats or insults) and make a decision within five minutes. In the team treatments with computer chat communication, members were also told that they must not identify themselves, and they adhered to this rule. To ensure consensus decisions during the game, each team member had to enter the same contribution for the computer to accept the team decision. If any one member deviated, the computer did not accept the decision and all team members had to start anew.[4] Note that this feature makes our design particularly conservative. It ensures that teams made consensus decisions where each member

[3] In the computer chat, subjects were denoted by numbers which could not be linked with the photos. The chat was open in every stage of the game and closed between the stages. When a member of the team left the stage in order to proceed to the next stage, all remaining team members were informed that one member has left the stage.

[4] All teams were able to reach a common decision within the time limit.

had to agree. Allowing for majority voting where members can be overruled should increase the difference between teams and individuals because it further obscures responsibility among team members.

The experiment included two short questionnaires, one before subjects knew about the game and another one after they had played the game. In the ex-ante questionnaire, subjects were asked about their personal background and some attitudinal characteristics, including gender, trust, and beliefs about others' selfishness. An important question was how much they care about what other people think about them which they could answer on a scale from 1 being "not at all" to 10 being "very much." This question was included to elicit subjects' concerns about their image. After this questionnaire, a set of written instructions was handed out which explained the game and included several numerical examples and control questions (see Appendix for instructions). The control questions tested subjects' understanding of the payoff function given in (1) to ensure that they were aware of the payoff-maximizing strategy and the dilemma situation. The game only began after all subjects read the instructions and answered the control questions correctly. After the game, subjects were asked to complete a second questionnaire which asked about their motivations and emotions during the game. While the teams with face-to-face communication were allowed to talk during the game, they were requested to read the instructions and complete the control questions as well as the two questionnaires individually and in silence, which they did. After the final questionnaire was completed, the subjects were paid their earnings in cash. Care was taken that individuals and teams left the lab one by one so that they did not meet or see each other.

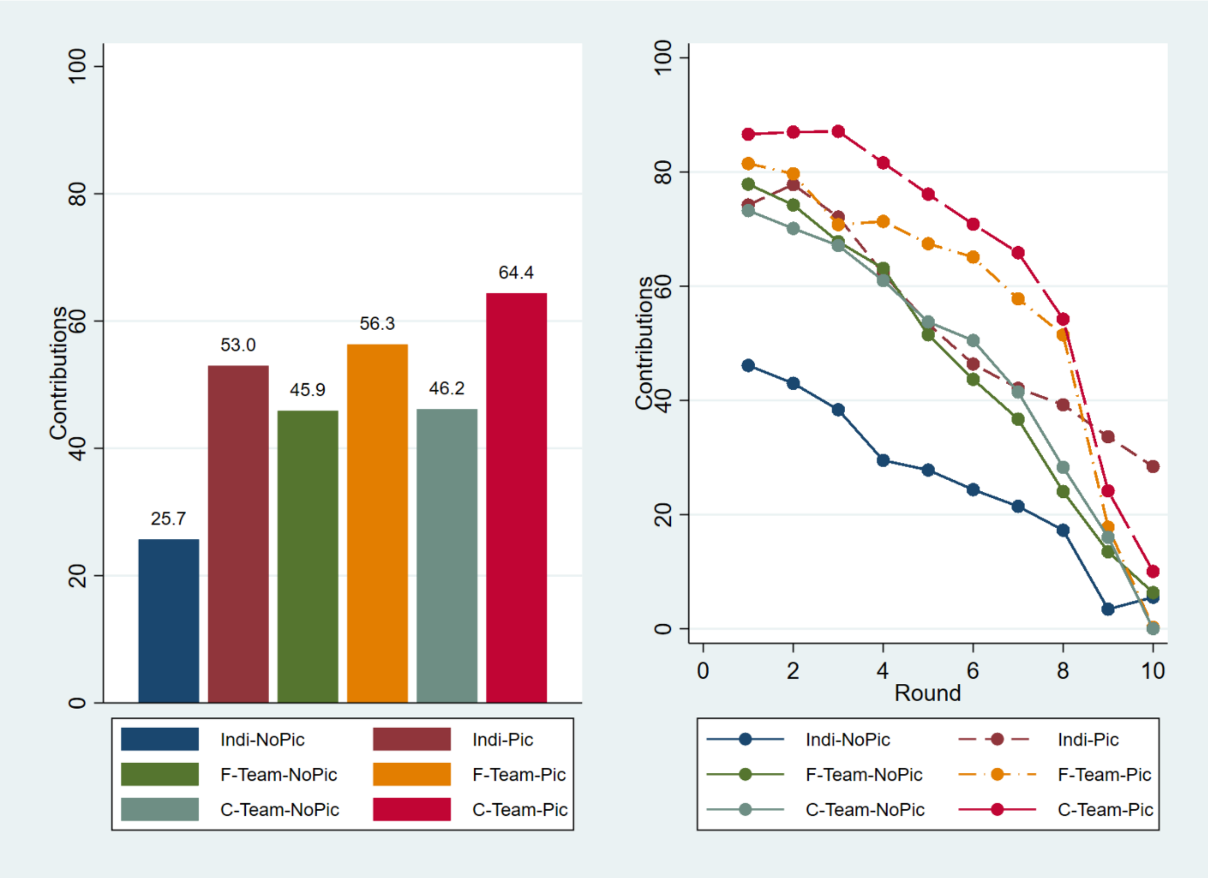
2.4 Results

2.4.1 Contributions to the public good

The left panel in Figure 2.1 shows the average contributions across rounds by treatment. The *Indi-NoPic* treatment shows by far the lowest contributions with 25.7 tokens on average (out of 100 tokens). Individuals in the *Indi-Pic* treatment contributed more than twice as much, namely 53 tokens on average. A Mann-Whitney-Wilcoxon (MWW) test shows that the difference between the two treatments is statistically significant ($N = 20$, $P = 0.0257$).[5] This result confirms the findings from the previous literature that revealing contributors' identities significantly increases cooperation among individuals (Andreoni and Petrie 2004; Samek and Sheremeta 2014).

[5] Unless stated otherwise, we use the meta-group as unit of observation in the statistical tests. That means, four individuals constitute an observation in *Indi-NoPic* and *Indi-Pic* and four teams (16 individuals) constitute an observation in *F-Team-NoPic*, *F-Team-Pic*, *C-Team-NoPic*, and *C-Team-Pic*. We use two-sided tests throughout the paper.

Figure 2.1 Average contributions across rounds and over time by treatment



Teams in the *F-Team-NoPic* treatment allocated on average 45.9 tokens to the public good and teams in *C-Team-NoPic* contributed 46.2 tokens on average. Compared to *Indi-NoPic* this is an increase of approximately 80 percent, and the differences are at least weakly significant (N = 20, P < 0.10 each). Thus, irrespective of the type of communication, teams contribute more than individuals when identification is not possible which has also been observed by Auerswald et al. (2018).

When identities were revealed, teams in *F-Team-Pic* contributed 56.3 tokens on average which is an increase of 23 percent compared to *F-Team-NoPic*. Teams in *C-Team-Pic* contributed 64.4 tokens on average, 39 percent more than the teams in *C-Team-NoPic*. The differences in average contributions due to the revelation of identities are much smaller for teams than for individuals (the increase for individuals is 106 percent) and they lack statistical significance (N = 20, P > 0.10 each). We find no significant differences between the teams with face-to-face communication and the teams with computer chat communication, neither when identities are kept private nor when identities are revealed (N = 20, P > 0.10 each). Thus, whether the discussion takes place face-to-face or via computer chat appears to matter little for cooperation. This result is in contrast to Kocher and Sutter (2007) who found more generous behavior with face-to-face communication than with communication through the computer, but in their experiment the computer communication was not only anonymous but also restricted to proposals and votes. As there are no significant differences between the teams with different types of communication, we can pool *F-Team-NoPic* and *C-Team-NoPic* into *Team-NoPic* and pool *F-Team-Pic* and *C-Team-Pic* into *Team-Pic*, and test again if identification has an effect. With the larger data set, we find a weakly significant difference between *Team-NoPic* and *Team-Pic* (N = 40, P = 0.0834). This shows that identification has an effect on teams, too, but the size of the effect is much smaller and the statistical significance is weaker than for individuals.

As teams contribute more than individuals in the anonymous game, the smaller effect of identification on teams may be due to a mere ceiling effect. In the following, we will show that this may be true at the beginning of the game but not towards the end. The right panel in Figure 2.1 shows how average contributions in the different treatments develop over time. As has been observed in many other public goods experiments, average contributions decrease over time in all six treatments. However, the initial contribution level and the slope of the downward trend differ. The first round is particularly interesting because players had to choose their contributions without any information about what the other players might do. Therefore, we can consider each player and each team as an independent observation in the statistical tests for the first round. Subjects in *Indi-NoPic* started the game carefully with relatively low contributions in the first round (46.1 tokens on average), arguably to avoid the risk of being exploited by others. Subjects in *Indi-Pic*, by contrast, started the game at a much higher contribution level (74.3 tokens), which represents an increase of 61 percent. A Mann-Whitney-Wilcoxon test at the player level reveals that the difference in the first round is highly significant (N = 80, P = 0.0005). The difference between the two treatments remains relatively stable until the end of the game. Interestingly, *Indi-Pic* is the only treatment in which subjects managed to stay well above zero contributions in the last round (28.4 tokens). Thus, the disclosure of identities has an immediate and lasting effect on individual contribution decisions.

Without identification, teams contributed more than individuals, especially at the beginning of the game. In the first round, teams in *F-Team-NoPic* contributed 77.9 tokens and teams in *C-Team-NoPic* contributed 73.3 tokens on average. Teams appeared more willing to risk a high contribution at the start of the game when the contributions by the other players were not yet known. However, the difference between teams and individuals decreased over time and vanished by the end. In the last round, teams in *F-Team-NoPic* contributed only 6.3 tokens on average and the teams in *C-Team-NoPic* contributed almost zero.

Teams in *F-Team-Pic* and *C-Team-Pic* also started at a high level (81.5 tokens and 86.6 tokens, respectively). The increase in first round contributions due to identification amounts to 5 percent for the F-Teams (N = 80, P = 0.4726) and 18 percent for the C-Teams (N = 80, P = 0.0341). Also in the middle part of the game, contributions in *F-Team-Pic* and *C-Team-Pic* exceed the contributions in their counterparts without picture, but then drop sharply in the last three rounds. In the last round, the differences are very small and contributions in *F-Team-Pic* are even lower than in *F-Team-NoPic*. Hence, for teams, the disclosure of identities only has a temporary effect.

It is also interesting to look at the extreme decisions, that is, contributing either all or nothing to the public good. Table 2.2 shows the proportions of zero contributions and full contributions for the first round, the last round, and all rounds together. It shows that, in the first round of *Indi-NoPic*, 22.5 percent of individuals contributed the full amount to the public good. The share of full contributions is substantially higher in the other five treatments (45 – 70 percent). This confirms that individuals in *Indi-NoPic* started the game rather carefully and tried to avoid the risk of being exploited by others. This concern appears to be less important in the other treatments. This is especially remarkable for *F-Team-NoPic* and *C-Team-NoPic* in which identities were kept private. For *F-Team-NoPic*, one could argue that concerns about others' opinions when showing selfish behavior are triggered within the team and so lead to higher contributions. But this argument cannot explain the high contributions in *C-Team-NoPic* where the members of a team remained anonymous. Thus, being in a team alone appears to reduce the fear of being exploited by others and increase the willingness to risk a high contribution in the first round. However, teams were unable to keep cooperation up and experienced a sharp reduction in contributions towards the end of the game.

Table 2.2 Percentage of zero and full contributions by treatment

	First round		Last round		All rounds	
	Zero	Full	Zero	Full	Zero	Full
Indi-NoPic	22.5	22.5	90.0	5.0	49.3	10.3
Indi-Pic	2.5	45.0	55.0	22.5	25.8	36.5
F-Team-NoPic	5.0	47.5	85.0	5.0	39.5	24.5
F-Team-Pic	2.5	60.0	85.0	0.0	26.5	45.3
C-Team-NoPic	5.0	47.5	97.5	0.0	29.3	19
C-Team-Pic	0.0	70.0	82.5	7.5	20.8	48.8

Turning to the other extreme, *Indi-NoPic* has a much higher percentage of zero contributions in the first round (22.5 percent) than the other treatments. The share of zero contributions is very low in *Indi-Pic* (2.5 percent), arguably because individuals did not want to be identified as a free-rider. Zero contributions in the first round are also rare in the team treatments (0 – 5 percent). A plausible explanation for this is that groups are better at anticipating the negative effects that such a strategy may have on the other players and overall payoffs. Indeed, the analysis of the chat protocols (see Appendix) suggests that maximizing payoffs and keeping the others' contribution level up were the most important motivations for the groups, whereas fairness or concerns to be exploited were mentioned only rarely. One point becomes clear when we compare the individual and the team treatments: Teams did not just average over what the members would have done individually. If they did, we would observe a similar average contribution level and fewer extreme decisions at both ends, that is fewer zero contributions and fewer full contributions. But this is not the case.

Let's now look at the extreme decisions in the last round. Here, the *Indi-Pic* treatment turns out to be the outlier. In *Indi-Pic* there are more full contributions (22.5 percent) than in the other treatments (0 – 7.5 percent). Likewise, there are fewer zero contributions (55 percent) than in the other treatments (82.5 – 97.5 percent). This confirms that the effect of the identification on individual behavior is still at play in the last round, whereas the differences for the teams are much smaller.

2.4.2 Differences between individuals and teams

To compare the effects of the pictures between individuals and teams in greater detail we employ a series of regression models. To this end, we pool the data of all treatments. The dependent variable is a player's or a team's contribution per round. The first two specifications (1) and (2) in Table 2.3 show results from linear regressions on contributions in the first round that are still independent of the behavior of the other players. In addition to the variables shown in the table, column (2) includes control variables controlling for subjects' gender, image concerns, and beliefs about others' trustworthiness and helpfulness. These variables were elicited prior to the game. The dummy variable *Picture* quantifies the effect of revealing the identities. The dummy variables *F-Team* and *C-Team* indicate whether decisions were made by individuals, teams with face-to-face communication, or teams with computer chat communication. The regressions include two interaction terms to show if the teams reacted significantly different to the pictures than the individuals. *Picture*F-Team* takes the value one for the subjects in *F-Team-Pic* and the value zero otherwise. Likewise, *Picture*C-Team* takes the value one for the subjects in *C-Team-Pic* and the value zero otherwise. The regression results confirm that the pictures lead to a significant increase in first round contributions for individuals. The increase amounts to almost 30 tokens and thus almost one-third of the endowment. Teams contribute significantly more in the first round than individuals when identities are not revealed. The interaction terms show that the teams react significantly less sensitively to the pictures than the individuals. The difference is smaller for the C-Teams than the F-Teams and it is only weakly significant when the control variables are included.

Columns (3) and (4) in Table 2.3 show random effect regressions for the contributions decisions made in rounds 1 to 10. In these regressions, the number of the current round is additionally included as an explanatory variable. The variable *Round* accounts for the downward trend of contributions over time. It is negative and significant in both specifications. The significant dummies *F-Team* and *C-Team* indicate that contributions in *F-Team-NoPic* and *C-Team-NoPic* are significantly higher than in *Indi-NoPic*. The interaction terms are no longer significant, though both coefficients point in the same direction as before. The effect of the pictures on individuals is still significant. Finally, columns (5) and (6) show regressions results on the contributions decisions made in the last round. We observe that the interaction term *Picture***F-Team* is significantly negative while *Picture***C-Team* is negative but not significant. In summary, the regression analyses confirm that the pictures have an immediate positive effect on contributions of individuals. Teams react significantly less sensitively to the pictures. The difference is particularly pronounced for the teams that use face-to-face communication and less pronounced for the teams than use computer chat communication.

Table 2.3 Results from linear regressions on contribution decisions

	OLS First round		Random effects Round 1-10		OLS Last round	
	(1)	(2)	(3)	(4)	(5)	(6)
Picture (d)	28.12***	29.49***	27.28**	27.60**	22.90**	22.75**
	(7.264)	(7.231)	(11.36)	(10.93)	(11.22)	(11.11)
F-Team (d)	31.75***	29.17***	20.20*	17.66*	0.800	-0.256
	(7.394)	(7.771)	(10.38)	(10.17)	(6.819)	(5.918)
Picture (d) * F-Team (d)	-24.50**	-24.97**	-16.84	-16.25	-28.98**	-28.20**
	(10.25)	(10.27)	(16.58)	(16.07)	(12.75)	(12.20)
C-Team (d)	27.13***	27.62***	20.47**	20.20**	-5.500*	-5.448*
	(7.506)	(7.612)	(8.903)	(8.632)	(3.174)	(3.081)
Picture (d) * C-Team (d)	-14.75	-17.41*	-9.063	-10.68	-12.88	-13.99
	(9.418)	(8.964)	(14.67)	(14.10)	(14.02)	(14.10)
Round			-7.119***	-7.119***		
			(0.414)	(0.414)		
Constant	46.12***	41.46***	64.84***	59.58***	5.525*	-0.859
	(6.221)	(7.676)	(7.440)	(7.903)	(3.174)	(4.080)
Control variables	No	Yes	No	Yes	No	Yes
Observations	240	240	2400	2400	240	240

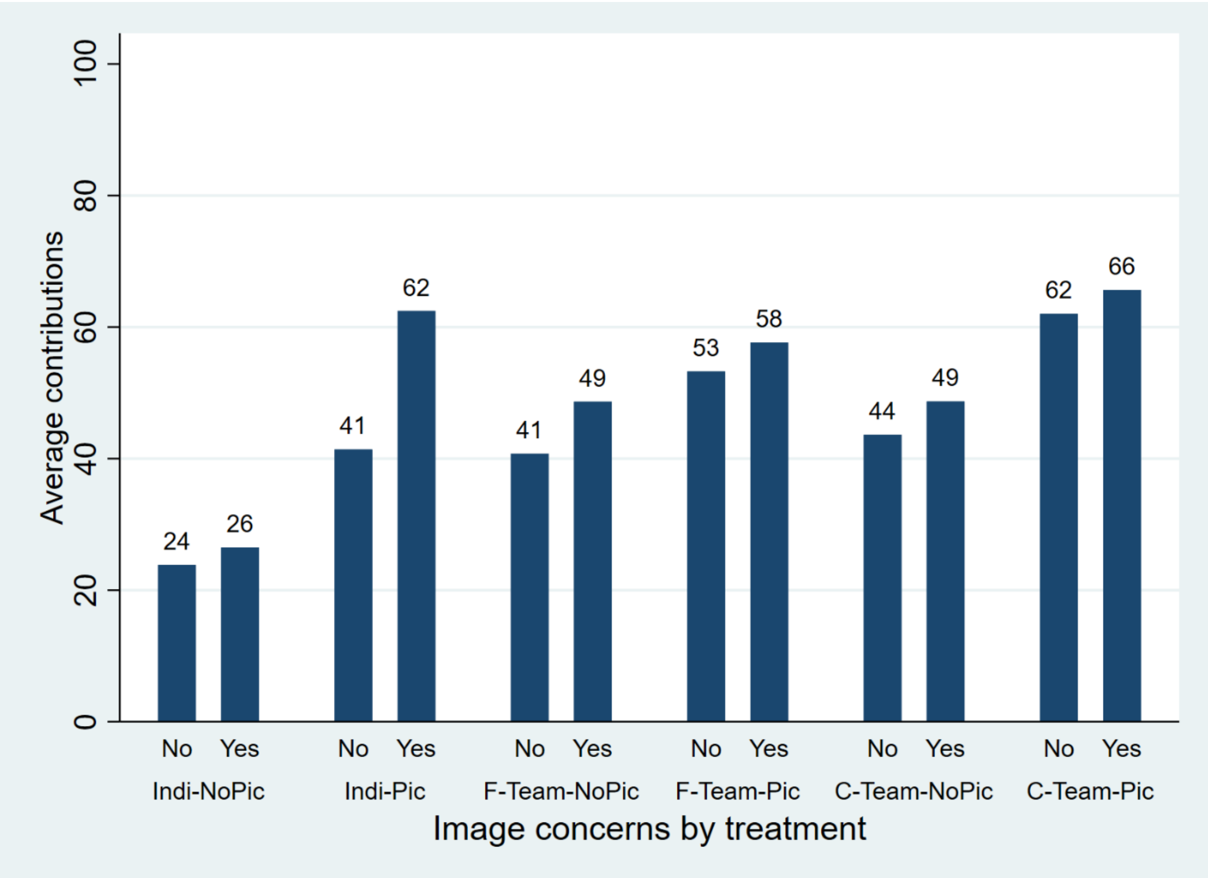
Ordinary least squares (OLS) regression and random effects panel regression with clustering of standard errors at the meta-group level. Numbers are marginal effects or discrete effects in case of dummy variables; standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Depended variable is an individual's (or team's) contribution per round. (d) indicates dummy variable. Definition of variables: *Picture* = 1 if identities are revealed, 0 otherwise; *F-Team* = 1 if face-to-face treatment, 0 otherwise; *C-Team* = 1 if chat treatment, 0 otherwise; *Picture*F-Team* = Interaction dummy of *Picture* and *F-Team*; *Picture*C-Team* = Interaction dummy of *Picture* and *C-Team*; *Round* = number of round. Control variables: *Female* = 1 if individual is female (or team consists of at least three female members), 0 otherwise; *Image concerns* = 1 if individual (or team on average) cares about image, 0 otherwise; *Trust* = 1 if individual (or team on average) considers others as trustworthy, 0 otherwise; *Others helpfulness* = 1 if individual (or team on average) considers others as helpful, 0 otherwise.

2.4.3 Image concerns

In our ex-ante questionnaire, we asked participants to state how much they care about what other people think about them on a scale from 1 being "not at all" to 10 being "very much." Figure 2.2 compares the average contributions of subjects who care strongly about their image ("Yes," answer categories 6-10) and subjects who do not care much about their image ("No," answer categories 1-5). For the team treatments, we consider the average answer to this question to distinguish between teams that care strongly (”Yes,” average answer is 6 or higher) and teams that care less (”No,” average answer is below 6). In all treatments, individuals and teams who care about their image contribute more to the public good than those who do not care (difference between “Yes” and “No” within treatments). Remarkably, the by far largest difference can be found in the *Indi-Pic* treatment; individuals who care about their image contribute about 50 percent more than those who do not care. In the other treatments, this difference is less than 20 percent.

Furthermore, we see that the disclosure of identities increases contributions by both players who care strongly about their image (difference between “Yes” with and without picture) and players who care less (difference between “No” with and without picture). However, the disclosure of identities has by far the largest impact on individuals who care strongly; their contributions are substantially higher in *Indi-Pic* than in *Indi-NoPic* (138 percent). This difference is much larger than for the individuals who do not care much (71 percent) and any type of team (18-41 percent).

Figure 2.2 Average contributions conditional on image concerns



We run additional regressions to study the effects of image concerns in greater detail. The results are provided in the Appendix. They confirm that the pictures have a significantly stronger effect on individuals who are concerned about their image than on individuals who care only little about this. We do not find significant differences between teams consisting of members who are concerned about their image and teams whose members care only little. While teams tend to contribute more to the public good when their members are concerned about their image (in some cases significantly), the effect of the pictures is not enhanced by image concerns. These findings suggest that being in a team helps sensitive individuals to overcome their image concerns when identities are revealed. While intuitive, our data do not allow us to examine this hypothesis further. Future research may test the suitability of alternative psychological scales to elicit image concerns and whether the lower sensitivity within teams is due to the diffusion of responsibility or merely because sensitive subjects are matched with insensitive subjects.

2.4.4 Acquaintances in the treatments with pictures

In the ex-post questionnaire, we asked subjects in all treatments in which identities were revealed if they knew another player or a person in another team. Two persons (5 percent) in *Indi-Pic*, 36 persons (22.5 percent) in *F-Team-Pic*, and 40 persons (25 percent) in *C-Team-Pic* answered this question positively. The difference between individuals and teams is simply caused by the fact that the subjects in the team treatments got to see 12 persons in the other three teams while the individuals only saw three other persons.[6] Regression analysis (see Table 2.6 in the Appendix) shows that knowing another player or someone in the other teams had no significant effect in *Indi-Pic* and *F-Team-Pic*. A higher number of team members who knew a person in another team increased contributions in *C-Team-Pic*. If anything, however, this makes our main result only stronger since removal of those teams would lower contributions in *C-Team-Pic* and move it even closer to the contribution level in the *C-Team-NoPic* treatment.

We also asked a number of other questions in the ex-post questionnaire in order to elicit subjects' perceptions after having played the game. The results are shown in the Appendix. They show, for example, that the participants in the team treatments were generally satisfied with their team, they felt involved in the decision making process and agreed with the final decision. The most interesting finding is the difference between individuals and teams with respect to their appreciation of the pictures. The individuals in *Indi-Pic* appear to perceive the pictures as more useful and influential than the teams in *F-Team-Pic* and *C-Team-Pic* which is consistent with actual behavior. However, when the participants in the anonymous treatments were asked whether they would have preferred to play the game with pictures, high contributing teams in *F-Team-NoPic* and *C-Team-NoPic* supported the idea of removing anonymity much more than low contributing teams, while the support among high contributors and low contributors in *Indi-NoPic* was equally low. This raises interesting questions about the

[6] This question referred only to acquaintances in the other teams and not to acquaintances in the own team. In *C-Team-Pic*, we also asked about acquaintances in the own team and nine persons (5.6 percent) answered this question positively. Regression analysis shows that knowing someone in the own team had no significant effect on contributions.

willingness of individuals and teams to employ a "naming and shaming" mechanism which go beyond the scope of this study.

2.5 Discussion and conclusion

Increasing transparency, and thereby exploiting the human tendency to behave more socially oriented under supervision, has been suggested as an effective way to improve cooperation (Rand, Yoeli, and Hoffman 2014; Kraft-Todd et al. 2015). This can even have positive side effects beyond the interpersonal relations, for example, when a change of personal eating or commuting habits due to social pressure has positive effects on the global climate (Nyborg et al. 2016). Our results confirm previous findings that identification and the mere suspicion of others' approval or disapproval is an incentive for individuals to behave more cooperatively (Andreoni and Petrie 2004; Rege and Telle 2004; Samek and Sheremeta 2014). The effect of revealing individuals' identities on cooperation is immediate, sizable, and permanent. A more detailed analysis shows that in particular individuals who care about their image make higher contributions to the public good when identities are revealed.

The novel result of our experiment is that the disclosure of identities only has a relatively small and temporary effect on cooperation among groups. Groups react significantly less sensitively to the pictures than the individuals. While teams consisting of members who care about their image tend to contribute more to the public good than teams whose members care only little, the effect of the pictures is not enhanced by image concerns. Thus, being in a team appears to help subjects to overcome their image concerns when identities are revealed.

We furthermore find that the differences between individuals and teams are particularly pronounced when the teams communicate face-to-face. The differences are smaller and less significant when the teams use a computer chat to communicate. A plausible reason why the difference is more pronounced for the F-Teams is that they become acquainted more easily. This facilitates social support within the team and enables a sense of group identity. Previous research has shown that members of groups feel less shame for a questionable decision the more they identify with the group (Shepherd, Spears, and Manstead 2013).

In conclusion, decision makers who want to use transparency to improve social outcomes, as for example policy makers or fundraisers, should try to target individuals rather than groups as increasing transparency among groups may only have small effects on their behavior. Group interactions seem to require stronger regulations at least when responsibility for decisions is diffused and members can hide within the group. We believe that this is the case for most group decisions.

Our study also adds to the relatively small literature on the differences between individuals and groups in the anonymous public goods game. We find that, irrespective of the type of communication, groups contribute more than individuals which is in line with Auerswald et al. (2018). Especially at the beginning of the game when it is not yet clear what the other players will do, groups appear to be more willing to risk a high contribution. They are also less likely than individuals to start the game with contributing nothing, perhaps because groups are better able to anticipate the negative effect this strategy may have on the other players' willingness to cooperate. The analysis of the chat protocols shows that payoff maximization and keeping others' contributions high were the main motivations for the teams (see Appendix). These motivations can also help to explain why groups' contributions decrease quickly over time and come close to zero by the end of the game. It is important to note that, while the potential for the pictures to make a difference is limited at the beginning of the game when groups contribute a lot anyway, there is great potential at the end of the game when contributions are very low. But, unlike in the case of the individuals, the potential is not used.

Finally, our study shows that contributions in the anonymous individual public goods game, the "workhorse" to study human cooperation, is the outlier among our treatments and produces the by far lowest contributions to the public good. All other treatments which include more real-life elements like identification, communication, and interaction produce higher contributions, at least in the first half of the game. The low contributions in the standard public goods game should perhaps not be interpreted as lack of cooperativeness but instead as a lack of real-life elements. Allowing subjects to communicate and decide together with others or to see more information about each other increases cooperation significantly without any changes in the incentives or the rules of the game.

It would be interesting to test if our results hold under different group decision making rules, for instance, when a majority rule is used or when one member decides as a group representative. As a majority rule further obscures responsibility and accountability within groups we would not expect a greater effect of identification under this rule. The decision by a group representative would be more interesting as it combines elements of both individual and group decision making (Hauge and Rogeberg, 2015). Likewise, revealing the input of each single member, rather than the final decision only, may lead to different results as this would make responsibility more transparent (Schopler et al., 1995). By forming groups according to certain preferences instead of random formation, for example by subjects' image concerns, one could further investigate if subjects become less sensitive in groups or if the matching of heterogeneous subjects is the more important factor.

Acknowledgements: We are grateful to Scott Barrett for a fruitful discussion on the topic that inspired this research and to Corina Haita-Falah, Elina Khachatryan, and Sonja Zitzelsberger for comments on an earlier version of this paper. The work was financially supported by the European Union (EU) Horizon 2020 program, action ERC-2014-STG, Project HUCO, grant number 636746.

2.6 Appendix

2.6.1 Samples of the pictures

The pictures below show how the pictures used in the experiment looked like. The people on the pictures did not take part in the experiment. They were informed, and they agreed, that the pictures would be used for research purposes. The participants in the experiment were informed that the pictures would be used only during the experiment and deleted afterwards.

F-Team-Pic

C-Team-Pic

Indi-Pic

2.6.2 Experimental Instructions

(The instructions below are for the *F-Team-Pic* treatment, translated from German. The instructions for the other treatments are similar.)
Welcome to this experiment!

General information

You can earn money in this experiment. How much you earn depends on the game play, or more precisely, on the decisions you and your fellow co-players will make. Please remain at your seat for the entire experiment. We will inform you when the experiment is finished and you are allowed to leave your seat.

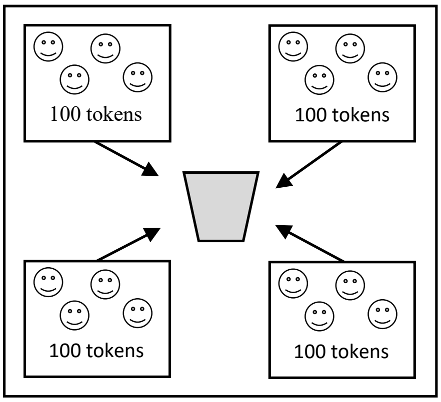
Now, read the following rules of the game carefully. Subsequently, answer the control questions below. All participants should read the instructions and answer the control questions independently from each other. Therefore, please do not talk with each other during this phase. In case you have a question, give us a hand signal. We will come to you and answer the question.

Game rules

There are four teams participating in this game, meaning your team and three other teams. Each team consists of four persons. So, in total, 16 persons participate in the game. Every team is confronted with the same decision task.

Each team receives 100 tokens. The teams decide if they keep their tokens or contribute them to a common project. The tokens that a team keeps benefit only the team. The tokens that a team contributes to the common project benefit all teams. The tokens that are contributed to the common project will be multiplied by 1.6 and then divided equally among the four teams. So, every team benefits from tokens contributed to the common project regardless of how much it contributed itself. The team's profit is the sum of the tokens kept and the tokens that it receives from the common project.

Example: If all four teams keep their tokens and do not contribute to the common project, every team receives 100 tokens (= 100 + 1.6*0 / 4). If every team contributes 100 tokens to the common project, each team receives 160 tokens (= 0 + 1.6 * 400 / 4). If three teams contribute 80 tokens each to the common project and one team contributes nothing, the former teams receive 116 tokens each (= 20 + 1.6*240 / 4) and the latter team receives 196 tokens (= 100 + 1.6*240 / 4).



All teams decide simultaneously how much they contribute to the common project. Any amount between 0 and 100 tokens is possible. Every team is supposed to discuss the decision about the contribution and to make the decision within 5 minutes. The discussion is to be held civilized (no threats, insults etc.). When the team has agreed on a contribution, all members have to enter the decision into the computer. Only then the decision is valid.

After all teams have chosen their contributions to the common project, the contributions of all teams will be shown on the screen. The computer will sort the contributions from the highest to the lowest. Next to the contribution, the photo of the respective team will be shown. Note that the photos will not be published outside of the experiment and not be passed to third parties. After the experiment, all photos will be deleted.

The game will be played for 10 identical rounds. The four teams will remain the same for all rounds. Each team will receive 100 tokens in every round which can be kept or contributed to the common project. After every round, contributions and the respective photos will be shown. The profit of a team is the sum of all tokens from all 10 rounds. Each token is converted into 0.04 Euros (25 tokens = 1 Euro). That means, for example, if a team earns 1500 tokens in the 10 rounds, it receives 60 Euros (= 0.04 * 1500). The profit of a team is equally distributed among the four members. In the example above, every team member would receive 15 Euros. Before and after the game, you will be asked a few questions. All participants are supposed to answer these questions independently from each other and in silence. This applies also to the following control questions.

Control questions

a. Right or wrong? Four teams participate in the game and every team consists of four members.

O Right O Wrong

b. Right or wrong? The game will be played for 10 identical rounds. The teams remain the same for all rounds.

O Right O Wrong

c. Assume that all teams have contributed a total of 120 tokens to the common project. Your team has contributed 30 tokens. What is the profit of your team in this round (in tokens)?

O 68 O 84 O 118 O 136 O 148 O 196

d. Assume that all teams have contributed a total of 120 tokens to the common project. Your team has contributed 0 tokens. What is the profit of your team in this round (in tokens)?

O 68 O 84 O 118 O 136 O 148 O 196

e. Assume that all teams have contributed a total 240 tokens to the common project. Your team has contributed 60 tokens. What is the profit of your team in this round (in tokens)?

O 68 O 84 O 118 O 136 O 148 O 196

f. Assume that all teams have contributed a total of 240 tokens to the common project. Your team has contributed 0 tokens. What is the profit of your team in this round (in tokens)?

O 68 O 84 O 118 O 136 O 148 O 196

When you have answered all control questions, give us a hand signal. We will come to you and check the answers. When we have checked the answers of all participants and there are no more questions, the game will start. Good luck!

2.6.3 Supplementary regression analyses

Tables 2.4 and 2.5 present additional regression results to show the effects of subjects' image concerns in greater detail. In these regressions, we pool the two individual treatments and we pool the two team treatments with the same type of communication. Like the regressions shown in the main paper, we include the variable *Image concerns* which measures individuals' and teams' concerns about their image. Before the game, all subjects were asked how much they care about what other people think about them which they could answer on a scale from 1 being "not at all" to 10 being "very much." For the individual treatments, the dummy variable *Image concerns* takes the value one if an individual's answer to this question is 6 or higher and zero otherwise.[7] For the team treatments, the dummy variable is one if team members answer this

[7] The underlying scale (from 1 to 10) provides only an ordinal measure. Following Wooldridge (2002, 223) we coded the variable as a dummy because we cannot be sure that the successive categories are perceived as equally spaced across the full scale. For instance, it is not clear that respondents perceived the difference between

question with 6 or higher on average and zero otherwise. We additionally include the variable *Image concerns*Picture* which is an interaction dummy of *Image concerns* and *Picture*. Thus, this variable is one if subjects care about their image and their identity was revealed, and zero otherwise.

Table 2.4 Regression results with interaction of pictures and image concerns

	(1)	(2)	(3)	(4)	(5)	(6)
	Indi-NoPic & Indi-Pic		*F-Team-NoPic & F-Team-Pic*		*C-Team-NoPic & C-Team-Pic*	
Round	-5.239***	-5.239***	-8.119***	-8.119***	-7.999***	-7.999***
	(0.746)	(0.749)	(0.653)	(0.655)	(0.577)	(0.579)
Picture (d)	27.28**	13.67	10.44	8.960	18.21*	18.16**
	(11.56)	(10.33)	(12.28)	(12.75)	(9.438)	(8.731)
Image concerns (d)		-0.722		13.61**		4.321
		(6.006)		(6.492)		(5.499)
Image concerns (d) * Picture (d)		22.44**		-5.068		-1.387
		(11.42)		(12.66)		(6.737)
Female (d)		-9.795		10.16		-7.203
		(6.293)		(6.268)		(6.856)
Trust (d)		24.30***		-2.875		8.564
		(7.301)		(8.242)		(5.669)
Others helpfulness (d)		-7.896		-18.36**		-2.550
		(6.113)		(7.928)		(4.185)
Constant	54.50***	48.66***	90.54***	91.04***	90.15***	85.73***
	(8.980)	(9.351)	(8.480)	(10.02)	(7.028)	(8.301)
Observations	800	800	800	800	800	800

Random effects panel regression with clustering of standard errors at the meta-group level. Numbers are marginal effects or discrete effects in case of dummy variables; standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Depended variable is an individual's (or team's) contribution per round. (d) indicates dummy variable. Definition of variables: *Round* = number of round; *Picture* = 1 if identities are revealed, 0 otherwise; *Image concerns* = 1 if individual (or team on average) cares about image, 0 otherwise; *Image concerns*Picture* = Interaction dummy of *Image concerns* and *Picture*. Control variables: *Female* = 1 if individual is female (or team consists of at least female members), 0 otherwise; *Trust* = 1 if individual (or team on average) considers others as trustworthy, 0 otherwise; *Others helpfulness* = 1 if individual (or team on average) considers others as helpful, 0 otherwise.

The regression results in Table 2.4 show that image concerns moderate the effect of the pictures on individuals. For the individuals, the variable *Picture* does no longer have a significant effect while the interaction dummy *Image concerns*Picture* has a significantly positive effect on contributions. All else being equal, individuals who care about their image and whose identities were revealed contributed on average 36.1 tokens (13.7 + 22.4 tokens) more than individuals who also care about their image but whose identities were kept private. Likewise, they contributed on average 21.7 tokens more than individuals whose identities were revealed but who do not care much about their image. The variable *Picture* has no significant effect on

categories 2 and 3 the same as the difference between categories 7 and 8, or that they interpreted the category 8 as twice as much as 4.

contributions of the F-Teams but it has a significantly positive effect on contributions of the C-Teams. As shown in Figure 2.2 in the main paper, subjects' image concerns tend to increase contributions to the public good in all treatments. The effect is significant in *F-Team-NoPic* but not in other team treatments. The interaction term *Image concerns*Picture* is not significant, neither for the F-Teams nor the C-Teams. This shows that the effect of the pictures is not enhanced by image concerns when decisions are made by teams.

That the effect of the pictures does not interact with image concerns when decisions are made in teams may partly be explained by the matching of heterogeneous subjects. Highly sensitive subjects are often matched with less sensitive subjects so that the team on average has a medium sensitivity. Nevertheless, it is unlikely that the matching of heterogeneous subjects explains everything. As long as there is enough variation among teams, we may still expect a larger effect of the pictures in teams whose members are concerned about their image than in teams whose members care less. Note also that, even when the differences in average image concerns among teams are relatively small, they are based on more subjects. A 1-point difference means that *four* subjects in a team gave an answer that was one category higher on average than *four* subjects in another team. The average reported values range from 4 (the team that cares the least) to 8.5 (the team that cares the most) and yet there is no significant difference in the effect of the pictures between teams caring little about their image and teams caring a lot.

Table 2.5 provides additional regression analyses where we use the minimum or maximum value of the team, instead of the average, in order to measure image concerns at the team level. These regressions examine if the member of the team with the highest or the lowest image concerns has a significant effect on contributions. None of the regressions shows a significant effect. Taken together, these findings suggest that, even when the members of a team care about their image, this concern does not enhance the effect of the pictures. Being in a team appears to help subjects to overcome their image concerns when identities are revealed.

Table 2.5 Regression results with alternative aggregation of image concerns within teams

	(1)	(2)	(3)	(4)
	F-Team-NoPic & F-Team-Pic	*C-Team-NoPic & C-Team-Pic*	*F-Team-NoPic & F-Team-Pic*	*C-Team-NoPic & C-Team-Pic*
Round	-8.119***	-7.999***	-8.119***	-7.999***
	(0.655)	(0.579)	(0.655)	(0.579)
Picture (d)	6.352	19.33	2.772	17.07*
	(13.80)	(10.21)	(12.34)	(8.629)
Trust (d)	-1.268	8.789	-0.397	9.505
	(7.863)	(5.343)	(7.134)	(5.491)
Others helpfulness (d)	-16.70*	-1.970	-15.37*	-1.025
	(8.483)	(3.926)	(7.219)	(4.045)
Female (d)	9.421	-6.099	9.326	-6.163
	(6.313)	(6.944)	(6.745)	(6.904)
Minimum image concerns (d)	2.696	5.510		
	(7.845)	(7.118)		
Minimum image concerns (d) * Picture (d)	1.305	-4.052		
	(11.54)	(7.172)		
Maximum image concerns (d)			-10.79	-3.541
			(10.14)	(7.207)
Maximum image concerns (d) * Picture (d)			11.72	2.395
			(17.83)	(8.183)
Constant	96.59***	84.90***	100.8***	87.56***
	(9.421)	(9.147)	(8.857)	(7.838)
Observations	800	800	800	800

Random effects panel regression with clustering of standard errors at the meta-group level. Numbers are marginal effects or discrete effects in case of dummy variables, standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Depended variable is an individual's (or team's) contribution per round. (d) indicates dummy variable. Definition of variables: *Round* = number of round; *Picture* = 1 if identities are revealed, 0 otherwise; *Minimum image concerns* = 1 if the lowest value within a team is above the median of 3, 0 otherwise; *Minimum image concerns*Picture* = Interaction dummy of *Minimum image concerns* and *Picture*; *Maximum image concerns* = 1 if the highest value within a team is above the median of 8, 0 otherwise; *Maximum image concerns*Picture* = Interaction dummy of *Maximum image concerns* and *Picture*. Control variables: *Female* = 1 if individual is female (or team consists of at least female members), 0 otherwise; *Trust* = 1 if individual (or team on average) considers others as trustworthy, 0 otherwise; *Others helpfulness* = 1 if individual (or team on average) considers others as helpful, 0 otherwise.

Table 2.6 Regression results on the effect of acquaintances

	(1)	(2)	(3)	(4)	(5)
	Indi-Pic	*F-Team-Pic*	*F-Team-Pic*	*C-Team-Pic*	*C-Team-Pic*
Round	-5.778***	-7.904***	-7.904***	-8.158***	-8.158***
	(1.335)	(0.692)	(0.692)	(1.029)	(1.029)
Acquaintance (d)	-17.33	-0.589		1.456	
	(12.80)	(12.41)		(7.683)	
Number acquaintances			0.680		7.590*
			(5.529)		(3.653)
Female (d)	-11.73	12.01	11.96	-8.427	-10.38
	(8.092)	(8.335)	(8.298)	(11.33)	(10.71)
Trust (d)	25.19**	-10.98	-11.01	17.31*	16.37*
	(7.669)	(11.22)	(11.14)	(7.154)	(6.813)
Image concerns (d)	22.39*	12.09	12.54	0.579	-2.928
	(8.997)	(9.373)	(9.514)	(7.394)	(7.245)
Others helpfulness (d)	-1.626	-20.26	-20.07	-1.381	-0.969
	(10.31)	(12.69)	(12.75)	(7.620)	(6.879)
Constant	63.52***	101.3***	99.99***	99.05***	95.87***
	(10.78)	(15.16)	(14.56)	(9.522)	(8.730)
Observations	400	400	400	400	400

Random effects panel regression with clustering of standard errors at the meta-group level. Numbers are marginal effects or discrete effects in case of dummy variables, standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Depended variable is an individual's (or team's) contribution per round. Definition of variables: *Round* = number of round; *Acquaintance* = 1 if individual (or at least one team member) knows another player (a person in another team), 0 otherwise; *Number acquaintances* = Number of team members who know a person in another team. Control variables: *Female* = 1 if individual is female (or team consists of at least female members), 0 otherwise; *Trust* = 1 if individual (or team on average) considers others as trustworthy, 0 otherwise; *Image concerns* = 1 if individual (or team on average) cares about image, 0 otherwise; *Others helpfulness* = 1 if individual (or team on average) considers others as helpful, 0 otherwise.

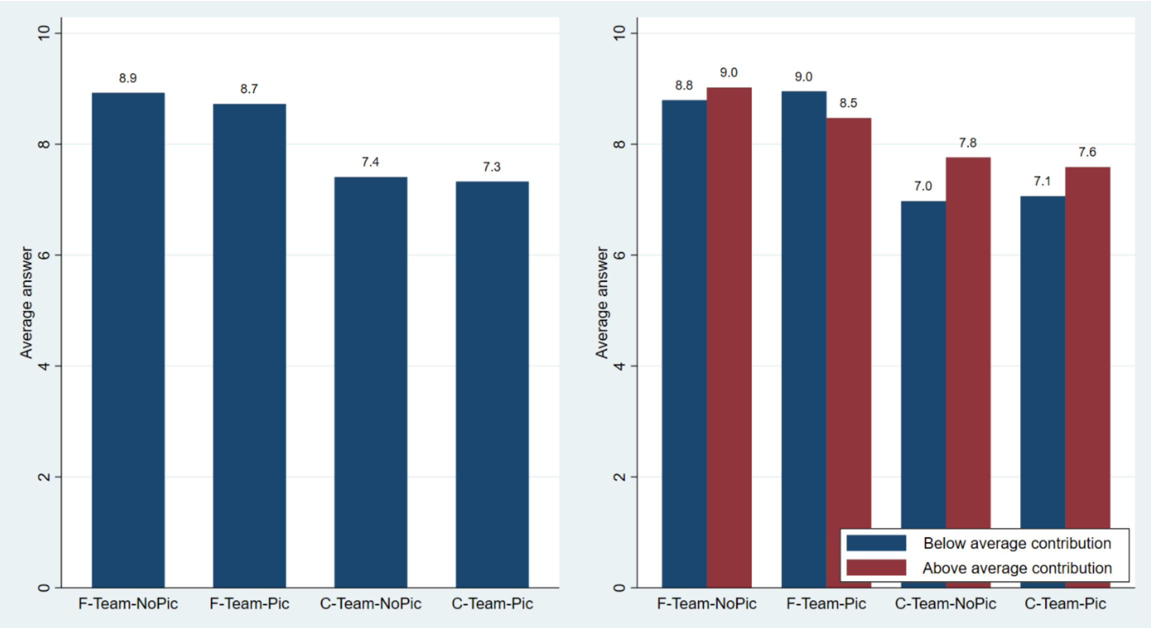
2.6.4 Results from the ex-post questionnaire

In our ex post questionnaire, we asked subjects about their emotions and motivations during the game. In particular, we wanted to know if the team decisions were consensus decisions and if members were satisfied with their team and the final decisions. The answers show that members on average were satisfied with their team, they felt involved in the team decision making process and agreed with the final decisions (Figure 2.3 and Figure 2.4). The majority of subjects stated that they would have made the same decision if they had decided alone (Figure 2.5). Not surprisingly, subjects felt more obliged towards the own team than towards the other teams (Figure 2.6 and 2.7).

In another set of questions, we wanted to know if subjects were satisfied with the version of the game they had played. Participants in the treatments with pictures were asked in how far they found the disclosure of the photos along with the contributions helpful and if the photos influenced their decisions. The individuals appear to perceive the pictures as more useful and influential than the teams (Figure 2.8 and Figure 2.9). We also asked if the participants would

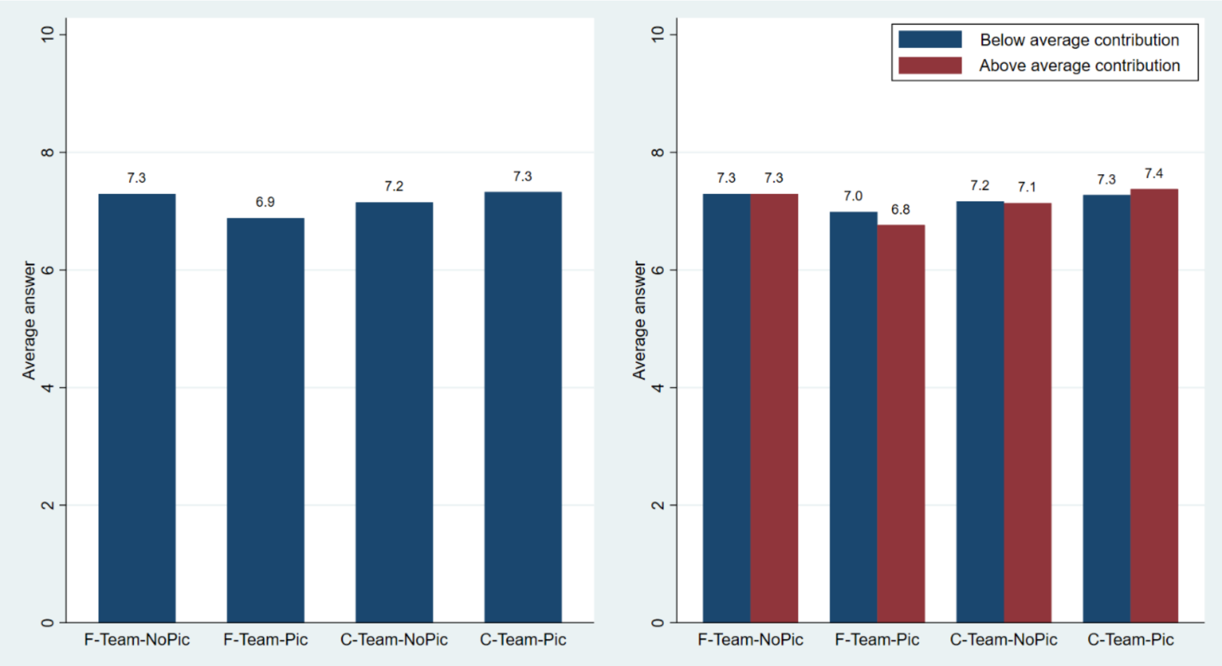
have preferred to play the game without disclosure of the photos. The support for the anonymous game was limited as can be seen in Figure 2.10.
Likewise, we asked participants in the anonymous treatments if they would have preferred to play the game with the disclosure of photos. The alternative version of the game received only limited support on average (Figure 2.11). It received more support from high contributing teams than from low contributing teams while there is no difference between high contributors and low contributors among the individuals. This lack of support for the disclosure of identities in *Indi-NoPic* is surprising since they had experienced a very low level of cooperation in the anonymous game. This raises an interesting question for future research: How would subjects with no experience choose between the two versions of the game, with or without disclosure of identities, if they had the choice?

Figure 2.3 Satisfaction with the own team



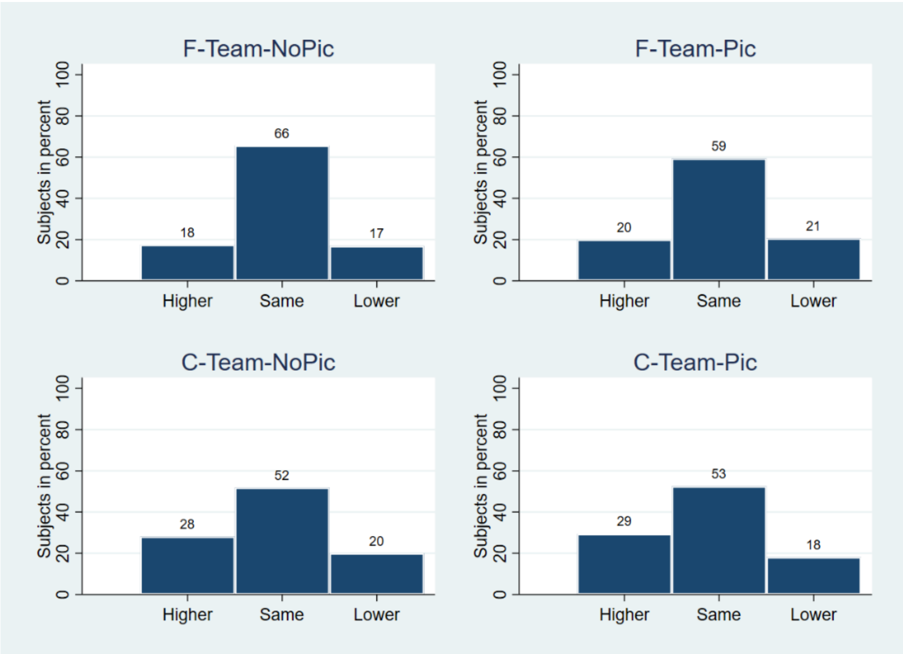
Average answers to the question: "Were you generally satisfied with the behavior of the other team members in your team? Mark a number on a scale from 1 (not at all) to 10 (very much)." "Below average contribution" in the right panel indicates that the team's average contribution across all ten rounds is below or equal to the average contribution of all four teams. "Above average contribution" indicates that the team's average contribution across all ten rounds is above the average of all four teams.

Figure 2.4 Involvement in the team decision process



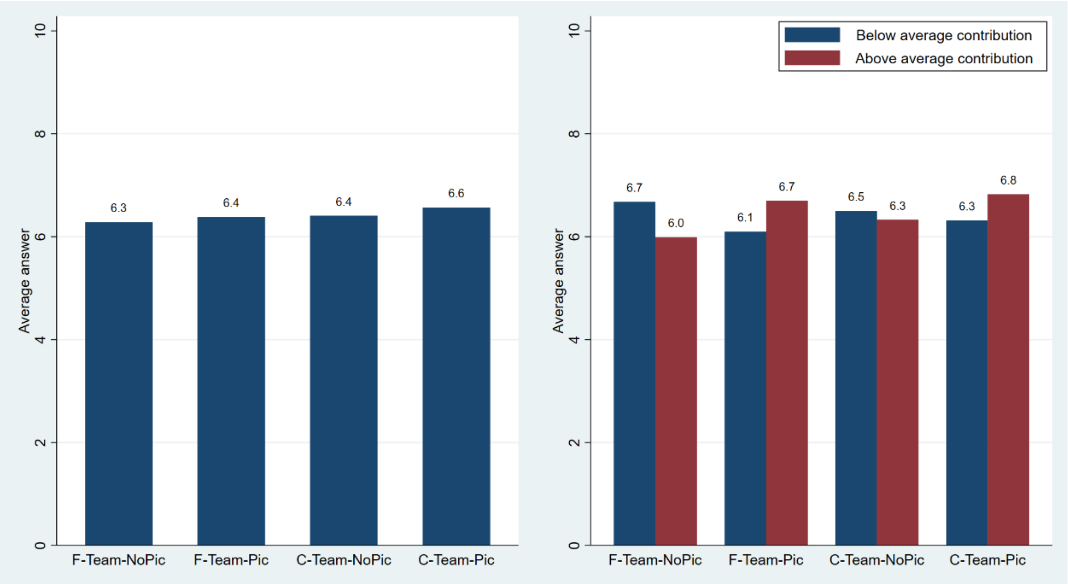
Average answers to the question: "How strongly were you involved in your team's decisions? Mark a number on a scale from 1 (not at all) to 10 (very much)." "Below average contribution" in the right panel indicates that the team's average contribution across all ten rounds is below or equal to the average contribution of all four teams. "Above average contribution" indicates that the team's average contribution across all ten rounds is above the average of all four teams.

Figure 2.5 Agreement with the team decisions



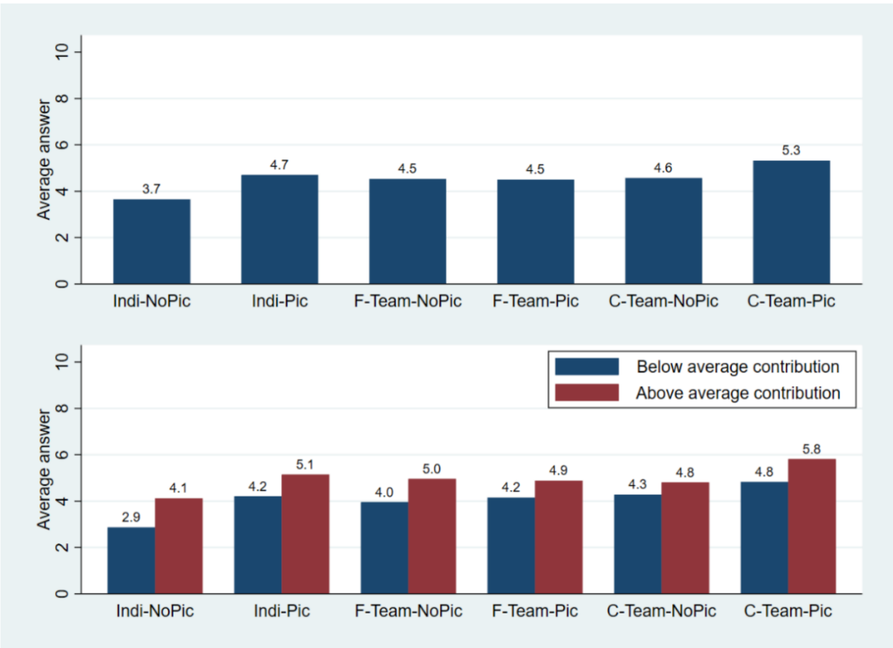
Answers to the question: "If you could have made the decisions alone, would you have chosen higher contributions, the same contributions, or lower contributions than your team?"

Figure 2.6 Feeling an obligation to the own team



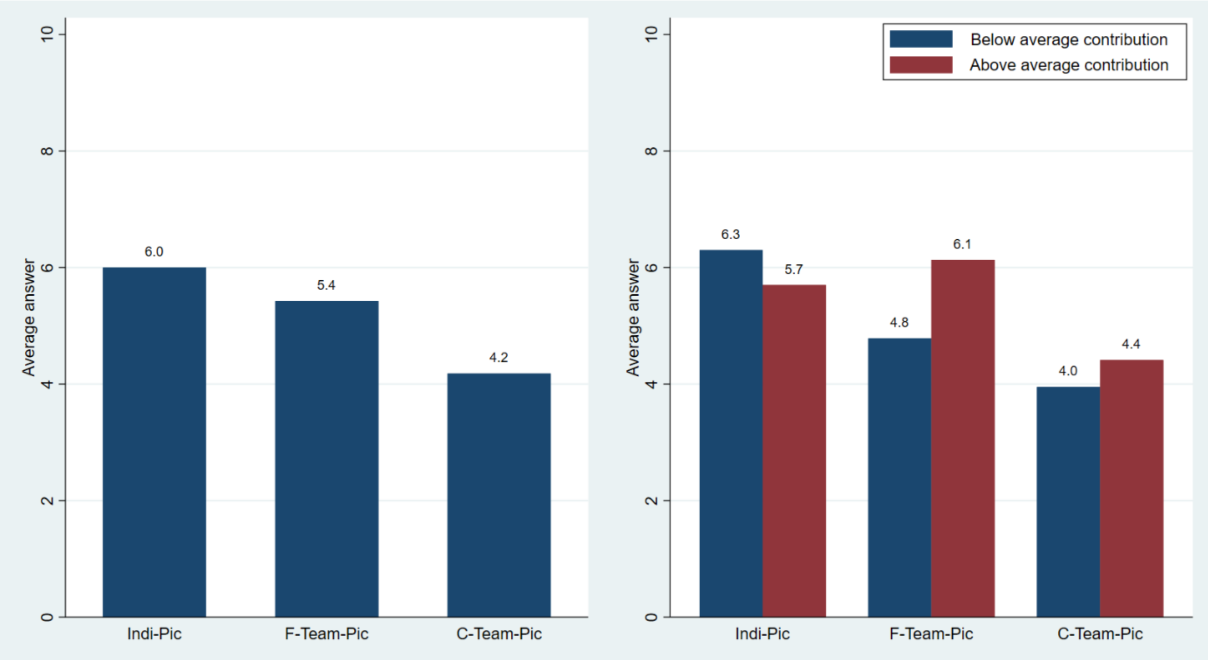
Average answers to the question: "Did you feel an obligation to your team in the game? Mark a number on a scale from 1 (not at all) to 10 (very much)." "Below average contribution" in the right panel indicates that the team's average contribution across all ten rounds is below or equal to the average contribution of all four teams. "Above average contribution" indicates that the team's average contribution across all ten rounds is above the average of all four teams.

Figure 2.7 Feeling an obligation to the other players / other teams



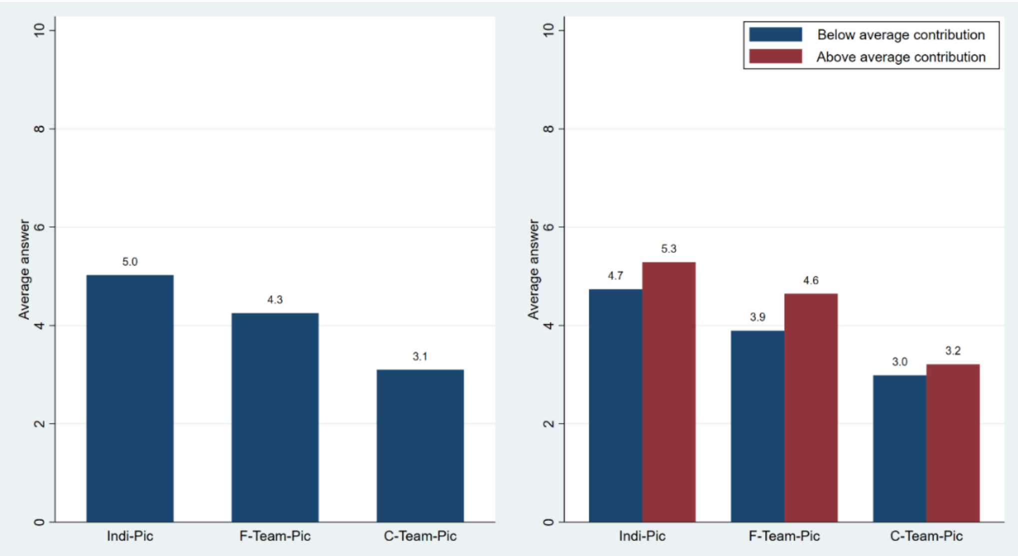
Average answers to the question: "Did you feel an obligation to the other players (to the other teams)? Mark a number on a scale from 1 (not at all) to 10 (very much)." "Below average contribution" in the lower panel indicates that the individual's (team's) average contribution across all ten rounds is below or equal to the average contribution of all four players (teams). "Above average contribution" indicates that the individual's (team's) average contribution across all ten rounds is above the average of all four players (teams).

Figure 2.8 Usefulness of the photos



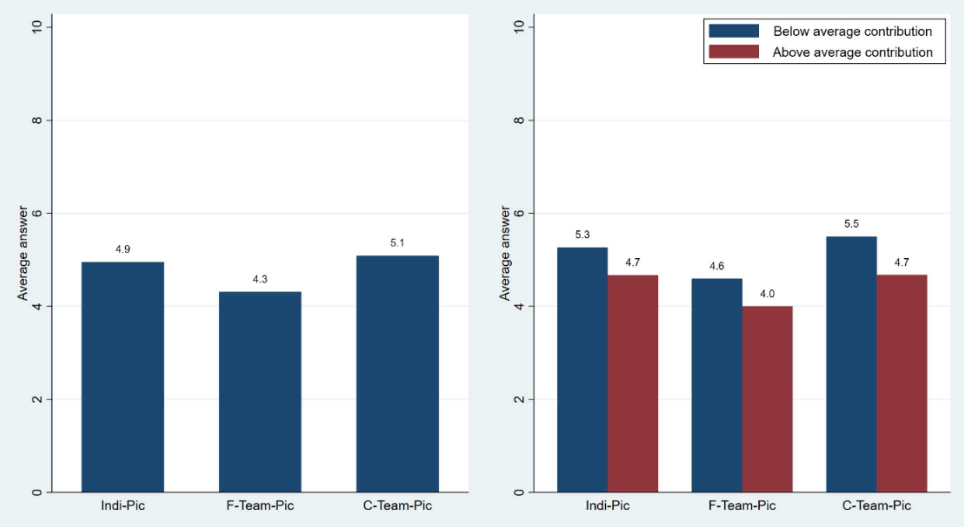
Average answers to the question: "Do you agree with the statement that the display of the photos along with the contributions was helpful? Mark a number on a scale from 1 (not at all) to 10 (very much)." "Below average contribution" in the right panel indicates that the individual's (team's) average contribution across all ten rounds is below or equal to the average contribution of all four players (teams). "Above average contribution" indicates that the individual's (team's) average contribution across all ten rounds is above the average of all four players (teams).

Figure 2.9 Influence of the photos



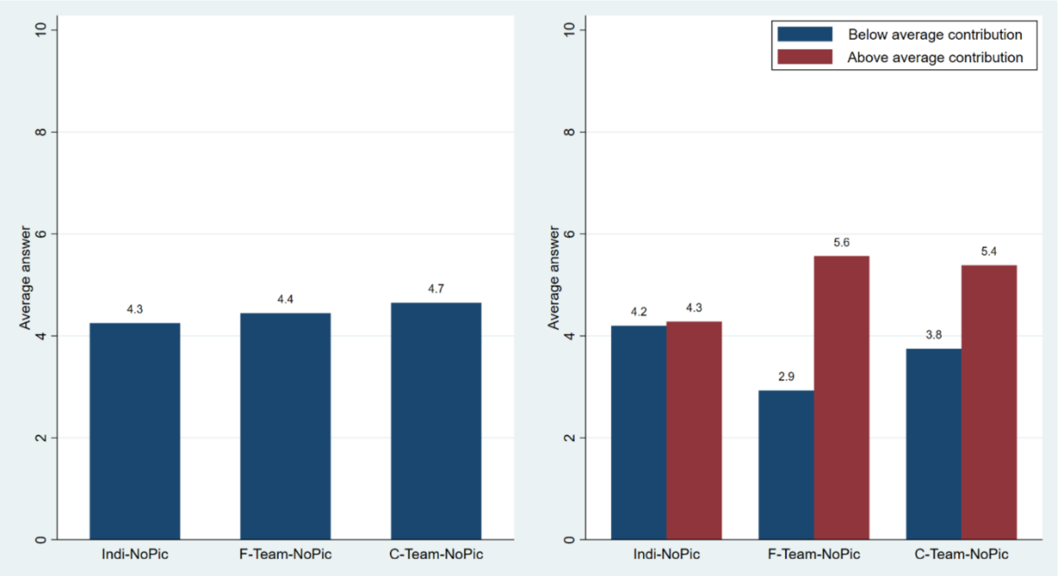
Average answers to the question: "Did the fact that the photos were displayed along with the contributions influence your decision (the decision of your team)? Mark a number on a scale from 1 (not at all) to 10 (very much)." "Below average contribution" in the right panel indicates that the individual's (team's) average contribution across all ten rounds is below or equal to the average contribution of all four players (teams). "Above average contribution" indicates that the individual's (team's) average contribution across all ten rounds is above the average of all four players (teams).

Figure 2.10 Preference of subjects in the treatments with pictures for anonymity



Average answers to the question: "In another version of the game, contributions of the players (teams) were shown anonymously without photos. Would you have preferred this anonymous version to the version you played? Mark a number on a scale from 1 (not at all) to 10 (very much)." "Below average contribution" in the right panel indicates that the individual's (team's) average contribution across all ten rounds is below or equal to the average contribution of all four players (teams). "Above average contribution" indicates that the individual's (team's) average contribution across all ten rounds is above the average of all four players (teams).

Figure 2.11 Preference of subjects in the treatments without pictures for removal of anonymity



Average answers to the question: "In another version of the game, contributions were shown along with photos of the players (teams). Would you have preferred this version with photos to the anonymous version you played? Mark a number on a scale from 1 (not at all) to 10 (very much)." "Below average contribution" in the right panel indicates that the individual's (team's) average contribution across all ten rounds is below or equal to the average contribution of all four players (teams). "Above average contribution" indicates that the individual's (team's) average contribution across all ten rounds is above the average of all four players (teams).

2.6.5 Analysis of the chats

In the two treatments with computer chat communication, *C-Team-NoPic* and *C-Team-Pic*, we were able to analyze how the teams discussed the problem and reached a decision on how much to contribute to the public good. Most decisions (56%) were made unanimously without much disagreement or discussion among the members. In 34% of all cases the decision was a compromise of the initial proposals of the team members. Around 5% of all contributions were determined by majority voting. In around 4% of cases one player could persuade all others to accept his or her proposal. These proportions were almost identical between the two chat treatments.

In total, subjects in these two treatments made 755 comments regarding their motivations. These comments convey the aim to earn as much money as possible (37%), to motivate the other teams to contribute more (25%), to treat the other teams fairly (25%), or not to be exploited by the other teams (12%). Again, these proportions were very similar in the two treatments, except that the concern to be exploited by others was expressed more often in *C-Team-NoPic* than in *C-Team-Pic*. Many comments expressed dissatisfaction or satisfaction with the behavior of the other teams and, in both treatments, subjects expressed more often dissatisfaction (459) than satisfaction (288). Interestingly, the photos were not discussed much in *C-Team-Pic*. Only 23 comments concerned the photos. Most of them speculated about the effect that the photos may have on the contributions.

3 The Effects of Identification on Prosocial Behavior of Individuals and Teams

Sven Christens[a], Astrid Dannenberg[a], and Florian Sachs[b]

[a] Department of Economics, University of Kassel, 34109 Kassel, Germany

[b] Faculty of Economics and Management, University of Magdeburg, 39106 Magdeburg, Germany

Abstract: It has been shown that groups often make more self-interested decisions than individuals, but it is not clear why. Are moral standards different for groups or do members of groups simply care less about moral standards? In this paper, we study if people care less about moral standards when they act in a group. Using a lab experiment where participants can donate money to a charity, we compare how individuals and groups react to the same moral standard when they are observed by others and when they are not observed. Overall, identification has little effect on behavior but a more detailed analysis shows interesting differences between women and men. Women are relatively generous irrespective of whether they decide alone or in a team and whether they are observed by others or not. Men in contrast become significantly more generous when they are observed by others but only when they decide alone and not when they decide in a group.

3.1 Introduction

People's willingness to volunteer and share money with others has received a lot of attention in economics and other disciplines. A robust result of this research is that people are particularly generous when they are observed by others. This tendency has been observed in many different field settings, such as blood donations, voting, energy saving programs, or church offerings (Rand, Yoeli, and Hoffman 2014; Kraft-Todd et al. 2015). Also lab experiments have shown that cooperation in public goods games is higher when the players' identities and their actions are revealed to the other players (Andreoni and Petrie 2004; Rege and Telle 2004). Given people's sensitivity to identification and observation by others, increasing transparency can be an effective way to support socially oriented behavior. Indeed, it has been shown that an increased degree of transparency can be more effective than monetary incentives (Rand, Yoeli, and Hoffman 2014; Kraft-Todd et al. 2015) and it is often used to improve social outcomes.[8] Fundraisers routinely offer public recognition to their generous donors to increase donations (Harbaugh 1998a; 1998b; Karlan and McConnell 2014).

When people act in groups, however, they often make more self-interested decisions than when they decide alone. Anonymity and diffusion of responsibility within groups can lead to more impulsive and aggressive behavior (Festinger, Pepitone, and Newcomb 1952; Zimbardo 1969). Groups behave more competitively and less cooperatively in social dilemma situations than individuals (Insko and Schopler 1998; Wildschut et al. 2003). They tend to make more payoff-oriented decisions and are less influenced by social considerations than individuals (Charness and Sutter 2012; Kugler et al. 2007). There is also evidence that observance by others does not significantly change the behavior of groups (Christens, Dannenberg, and Sachs 2019). This begs the questions whether moral standards are different for groups or whether members of groups care less about moral standards even when they are observed.

In this paper, we are concerned with the second part of this question. We exogenously control for the moral standard and study the prosocial behavior of individuals and groups both when they are observed by others and when they are not observed. Prosociality is measured in our experiment by offering participants the possibility to make a donation to a well-known charity (the Red Cross). Participants decide either alone or within a three-person group. Within the group, members do not meet but only exchange proposals through the computer and then vote on the proposals until unanimity is reached in order to determine the common donation of the group. This procedure allows us to use the first proposals that the group members make as an indicator for individual preferences of the group members. Half of the subjects make the donation decisions anonymously while the other half is informed that their donation decisions

[8] For instance, Finland regularly publishes information about the taxable income of its citizens in order to reduce tax avoidance (www.nytimes.com/2018/11/01/world/europe/finland-national-jealousy-day.html, accessed in July 2019). The French government has announced plans to publish the names of citizens guilty of the worst cases of tax fraud (www.reuters.com/article/uk-france-tax/france-to-name-and-shame-worst-cases-of-tax-fraud-pm-idUSKBN1FK2OT, accessed in July 2019).

and their identity will be observed by other people. Following the approach of Andreoni and Petrie (2004) and Samek and Sheremeta (2014, 2016), we use digital photos to identify individuals and teams. After the donation decision has been made, photos of the individuals or group members are displayed along with their chosen donations to a board of observers. Subjects know about this procedure in advance and can adjust their donations if they wish. An important goal was to exogenously control for the moral standard to allow for a comparison of how individuals and groups react to the same standard. For this purpose, prior to this study, we asked a number of uninvolved students which donation amount they would find appropriate. The average answer was provided to all individuals and groups in the present study just before they were asked to decide about their own donation.

The experimental results show that overall identification has no significant effect on donations. A plausible explanation is that the provision of a moral standard, as described above, already makes people think about others' expectations and identification does not increase donations much further. However, the overall averages mask some interesting differences between women and men. Women are relatively generous irrespective of whether they are observed or not and whether they decide alone or in a team. Identification does not make a difference to them, neither when they decide alone nor when they decide in a team. Men, in contrast, behave more generously when they are observed than when they decide privately. This effect is mostly driven by subjects who donate zero when they are not observed and make a positive donation when they are observed. The positive effect of identification disappears when men decide within a team. A large proportion of men prefers to donate zero even when the decision of the team is observed. These results suggest that men are to a greater degree motivated by image concerns than women. These image concerns, however, are effective only for personal acts where full responsibility can be claimed, and not collective acts where responsibility is diffused.

The remainder of the paper proceeds as follows: Section 2 provides the background for our study, summarizing previous findings from social psychology and behavioral economics. Section 3 explains the experimental design. Section 4 presents the results, Section 5 discusses the results and concludes.

3.2 Background

As background for our study, this section will summarize previous findings from social psychology and behavioral economics on how identification and observation improve prosocial behavior and how individuals and groups differ in their economic decision-making.

One of the most robust results of research on prosocial behavior is that people behave more generously when their actions are observed by others. For example, blood donors in Italy donated more frequently when these donations were made public (Lacetera and Macis 2010). Citizens in Switzerland were more likely to fulfill their civic duty of voting when there was a high chance that others would notice the act (Funk 2010). Visitors to a national park in Costa

Rica contributed more to the maintenance of the park when the solicitor was able to see the contribution (Alpizar, Carlsson, and Johansson-Stenman 2008) or when they had company who could see the contribution (Alpízar and Martinsson 2013). Energy consumers in Northern California were more likely to take part in a blackout prevention program when their decision was observable to their neighbors (Yoeli et al. 2013). Church goers in the Netherlands donated more when the persons sitting next to them was able to see the offering (Soetevent 2005). The tendency to behave more prosocially when being watched appears so strong that even displaying a picture of a pair of eyes close to where the decision is made suffices to change behavior (Bateson, Nettle, and Roberts 2006; Ekström 2012; Ernest-Jones, Nettle, and Bateson 2011; Francey and Bergmüller 2012; Powell, Roberts, and Nettle 2012).

The effect of identification has also been shown in lab experiments. Players' cooperation rates in a public goods improved significantly when they had to convey their contributions to the other players after the game (Rege and Telle 2004) or when a photo of them was shown to all players along with their contributions (Andreoni and Petrie 2004). The photos had little effect when they were published without the contribution decisions. Building on these findings, Samek and Sheremeta (2014) found that publishing the pictures of only the lowest contributors sufficed to improve overall cooperation while publishing the pictures of only the highest contributors did not lead to significant differences. This finding suggests that feelings of guilt or shame associated with low contributions are a stronger motivation than pride or prestige which may be felt after contributing more than others. Using a meta-study on dictator games, Engel (2011) found that identification makes dictators less likely to give nothing but also less likely to give more than half of the pie. Taken together, this research suggests that people in many different contexts and cultures are sensitive to surveillance and potential approval or disapproval by others. Increasing transparency can thus be an effective way to regulate social processes.

Previous research on group behavior has shown that groups tend to be more payoff-oriented and less influenced by social considerations than individuals (Charness and Sutter 2012; Kugler, Kausel, and Kocher 2012). Experimental studies found, for instance, that groups sent less money in the trust game (Kugler et al. 2007), they made and accepted smaller offers in the ultimatum game (Bornstein and Yaniv 1998), they made lower transfers in the dictator game (Luhan, Kocher, and Sutter 2009), and they were less likely to keep a promise when this was costly (Nielsen et al. 2019). Groups have also been found to be less cooperative in prisoners' dilemma games (Insko et al. 1987; Schopler et al. 1995; Insko and Schopler 1998; Wildschut et al. 2003) or common-pool resource games (Gillet, Schram, and Sonnemans 2009). The lower cooperativeness has been explained by the ability of groups to justify selfish decisions (social support of shared self-interest hypothesis), to create a shield of anonymity and diffuse responsibility among the members (identifiability hypothesis), and to anticipate the selfishness of other groups (schema-based distrust hypothesis). In repeated social dilemma games, groups were often more cooperative than individuals and, if there was a punishment opportunity, they were less likely to punish, both factors leading to higher payoffs for groups (Müller and Tan 2013; Auerswald et al. 2018; Huber, Model, and Städter 2019). It should be noted, however, that groups are not always more payoff-oriented. For example, Cason and Mui (1997) found that groups gave more in a dictator game than individuals and Engel (2011) found no significant difference in giving behavior between individuals and groups in his meta-study on dictator games.

The studies that are closest to ours are Hauge and Rogeberg (2015) and Christens, Dannenberg, and Sachs (2019). Hauge and Rogeberg (2015) used a one-shot public goods game and a within-subject design to compare subjects' contribution decisions as an individual and as a representative of a group. Identification was introduced by a public condition in which subjects were told that there is a one-ninth probability that they will be asked at the end of the experiment to come forth to the blackboard and write down their contribution in front of all other participants in that session. They found that women contributed more when they acted on behalf of a group than when they acted alone. Women's individual contributions were higher in the public condition than in the anonymous condition. Men's choices as individuals and group representatives were very similar and contributions in both roles were higher in the public condition. The difference to our study is that individuals did not make a decision within a group but on behalf of a group. This is an important difference because these decisions were still individual decisions and not group decisions. Hiding within the group therefore was not possible.

Christens, Dannenberg, and Sachs (2019) used a repeated public goods game to compare the effects of identification between individuals and groups. Identification was introduced by showing digital pictures of individuals or groups along with their contributions after each contributions round. They found that, when contributions were made anonymously, groups were more cooperative than individuals at the beginning of the game. Identification had a significantly positive effect on individuals but not on groups. However, it is not clear if the effect of identification on groups was insignificant because members could hide within their group, because groups played with groups rather than individuals, or because groups were already more cooperative in the anonymous case.

In this paper, we want to show if identification affects individuals and groups differently when they are confronted with the same receiver, the same kind of observation, and the same moral standard; in other words, if being in a group alone changes the effects of identification.

3.3 Experimental design

In our experiment, subjects were randomly assigned to the role of players and observers. In each session, we had 16 participants. Twelve participants were assigned to the role of a player and four to the role of an observer.[9] For the individual decisions each player received 15€ of which he or she could donate any amount between 0€ and 15€ to the German Red Cross. The players could keep the amount that they did not donate. In the team decisions, three players formed a team, so that we had four teams per session. Each team received an endowment of 45€ of which they could make a donation between 0€ and 45€. The amount that was not donated was equally distributed among the team members.

We hypothesized that the effect of identification would be larger for people who care about other people's opinions. To test this hypothesis, subjects were asked prior to the game to state, on a scale from 1 to 10, how important it is for them what other people think of them. In order to be able to test the hypothesis not only for individuals but also for teams, teams were formed according to subjects' answers to this question. The algorithm we used sorted individuals who care the most in the first group, individuals who care the second most in the second group, and so on. This sorting mechanism allows us to test if teams consisting of members who care about others' opinions react differently to identification than teams consisting of members who do not care much. Subjects were unaware of this sorting so that the information per se could not influence behavior.[10] The donation decision within the teams was made through a proposal and voting scheme. Each round the three team members submitted simultaneously a proposal on how much money should be donated. After the members made their proposals, all three proposals were presented to the team members. Subsequently, the team members voted on the proposals in random order. The players could either accept or decline the shown proposal. A decision was implemented when all three team members accepted the shown proposal. If no unanimous decision was made, another round of proposing and voting started. If no consensus decision was made after ten such rounds, a random number between 0 and 45 was drawn as a donation by the computer. Members of a team never met and, apart from the proposals and voting, there was no other communication between them.

In all treatments, just before the players made the donation decision, individually or in a team, they were provided with a moral standard. Specifically, they were given the following information: "In a short survey in which 38 uninvolved students were asked which amount should be donated in such a situation, the average answer was 64 percent." This standard (64 percent) was shown to all players and it was the same for all treatments. The standard was obtained prior to the experiment by describing the situation to a group of students at another university in Germany and asking them which proportion of a monetary endowment they think should be donated in such a situation. Importantly, we did not ask them to state what they

[9] The players made two donation decisions in the experiment. When they made the first decision they did not know what the second decision would be but they were informed that one of the two decisions would be randomly selected at the end of the experiment for payment. In this paper, we present the results from the first decision only.

[10] A weaker reaction to identification in randomly formed teams can have two reasons. First, subjects who sensitive to others' opinions are matched with subjects who are less sensitive or, second, sensitive subjects become less sensitive in teams (Christens, Dannenberg, and Sachs 2019). To avoid two potential reasons, we decided to form teams according to subjects' sensitivity to others' opinions.

themselves would donate but what *should* be donated, putting them in the shoes of an observer. The description of the situation explained that a donation decision was made by participants in a lab experiment, but it did not include the precise value of the endowment, whether the donation was made by individuals or teams, or whether the donation was made anonymously or not. This allowed us to use the same standard for all treatments without deception.[11] Our aim of this manipulation was to align players' beliefs about what an observer would consider an appropriate donation. This manipulation allowed us to exclude different moral standards as a potential influential factor for behavioral differences between individuals and teams. In order to check if the manipulation was successful we elicited both observers' expectations about the appropriate donation and players' second-order beliefs about the observers' expectations. When the players made the donation decision, observers were asked which amount they think should be donated. Like the survey participants whose answers provided the standard, they were not asked what they themselves would donate but what *should* be donated. Unlike the survey participants, however, they knew at this moment whether the players made the donation decision individually or in a team, and whether the decision was made anonymously or not. After the players made the donation decisions and before the decisions were shown to all participants, the players were informed that the observers had been asked to state what they think should be donated. Then the players were asked to guess the average amount that the observers had stated. The players' guesses were incentivized because here we are interested in their true beliefs about the observers' statements and not in their own normative opinions. We paid €5 for a correct guess, meaning the actual average amount stated by the observers plus or minus 1 percentage point. All expectations and second-order beliefs were elicited in percent of endowment.

After the players' beliefs had been elicited, the donation decisions were shown to all players and observers in the session. In the anonymous treatments (*"NoPic"*), an anonymous outcome table was shown to the players and observers. The donations of the four teams or the 12 individual players were listed from highest to lowest donation and the donations could not be linked to the players or team members. In the treatments with identification (*"Pic"*), the players saw the same anonymous outcome table. The observers, by contrast, saw photos of the individual players or the team members next to their donations. On the photo, each player held a sign that showed his or her given name. Note that in both cases, individual donations and team donations, the number of shown faces on the photos was the same for the observers (12 individuals or four teams consisting of three individuals each). The photos of all participants were taken prior to the experiment. Subjects were also informed that the photos would be used only during the experiment and deleted afterwards. When the chosen donations were shown on the screen, the observers were asked to calculate the sum of all donations and to enter the sum

[11] We chose a group of students from whom we expected a relatively high average response (participants in an environmental economics class) which indeed happened.

into the computer. This task was required to ensure that the observers spent some time inspecting the outcome. Subsequently, the observers witnessed the transfer of the donations to the German Red Cross by the experimenter.

Table 3.1 gives an overview of all treatments. The first treatment is *Indi-NoPic* where the players decided as individuals and no photos were shown to the observers. In *Indi-Pic,* the players decided as individuals and photos were shown to the observers so that they could link the donations to the players. In *Team-NoPic*, the players decided within a team and no photos were shown to the observers. Finally, in *Team-Pic*, the players decided within a team and photos were shown to the observers so that they could link the donations to the teams but not the individual members. All rules were carefully explained to the players and the observers before the experiment started.

We collected data in four sessions for Indi-NoPic and in three sessions for the other three treatments. Overall, 208 subjects participated in the experiment. All sessions were conducted at the computer lab MaxLab at the University of Magdeburg, Germany. All decisions were entered into computers (software z-Tree; Fischbacher 2007). Subjects were recruited from the general undergraduate student population using hroot (Bock, Baetge, and Nicklisch 2014) and each of them took part in one treatment only.[12]

Table 3.1 Treatments

Treatment	Picture	Number of subjects	Number of individual donations / proposals	Number of team donations
Indi-NoPic	No	48 players + 16 observers	48	
Indi-Pic	Yes	36 players + 12 observers	36	
Team-NoPic	No	36 players + 12 observers	36	12
Team-Pic	Yes	36 players + 12 observers	36	12

3.4 Results

In the following, we present what the observers considered an appropriate donation in the different treatments, what players and teams believed about the observers' expectations, how much players and teams donated to the charity, and if the donation decisions were influenced by the pictures.

3.4.1 Expectations

An important goal of our experimental design was to manipulate the players' second-order beliefs about the observers' expectations. In order to test if this manipulation was successful,

[12] Even though a within-subject design where players make both decisions with and without identification would have been advantageous in terms of statistical power, we felt that this design would tell subjects too much about the purpose of the study (effects of identification) and thus opted for a between-subject design.

we asked the observers in all treatments about what they think should be donated to the charity and we asked the players and team members to guess the observers' expectations.

On average, the observers in *Indi-NoPic* stated that 28.5% of the endowment should be donated and the observers in *Indi-Pic* stated 23.4%. The difference is not statistically significant (Mann-Whitney-Wilcoxon test (MWW), N = 28, p = 0.527).[13] Interestingly, the observers expected much less than the standard of 64% which probably results from subject pool differences between the participants in the experiment and the students who provided the standard. For the teams, the observers thought on average that 22.5% of the endowment should be donated in *Team-NoPic* and 17.5% in *Team-Pic* (N = 24, p = 0.521). Expectations for the team decisions were thus lower than for the individual decisions but the differences are not statistically significant (p > 0.1 each).

Even more important than the observers' expectations are the players' second-order beliefs about the observers' expectations.[14] Individuals in *Indi-NoPic* guessed on average that the observers would think that 38% of the endowment should be donated and individuals in *Indi-Pic* guessed on average 33.1%. This difference is not statistically significant (N = 72, p = 0.372). Teams had similar beliefs. Team members in *Team-NoPic* guessed on average that the observers would think that 32.9% of the endowment should be donated and team members in *Team-Pic* guessed on average 36.9% (N = 48, p = 0.252). The beliefs of the teams are not significantly different from the beliefs of the individuals (p > 0.1 each). Even when we pool the two individual treatments and the two team treatments, we do not find a significant difference in the beliefs between individuals and teams (N = 120, p = 0.757). Altogether, this suggests that our manipulation of beliefs was successful.

3.4.2 Donations

The upper part in Figure 3.1 presents an overview of average donations in all treatments as percentage of the endowment (15€ for individuals and 45€ for teams). Our focus lies on the question whether the pictures that were shown to the observers increased donations compared to the anonymous donations. Individuals in *Indi-NoPic* donated on average 28.5% of the endowment. Average donations in *Indi-Pic* were slightly higher with an average of 30%, but the difference is not statistically significant (MWW, N = 84, p = 0.708). Teams in *Team-NoPic* donated on average 25.3% of their endowment and teams in *Team-Pic* donated on average 20.9%. Again, there is no statistically significant difference (N = 21, p = 0.520).[15] We also find

---

[13] All non-parametric tests in this paper are two-sided.

[14] As explained in footnote 3, subjects made two donation decisions in the experiment and only one of them was randomly selected and implemented. Players' beliefs were only elicited for the decision that was implemented. Therefore, the numbers presented here are based only on cases in which the first decision was implemented.

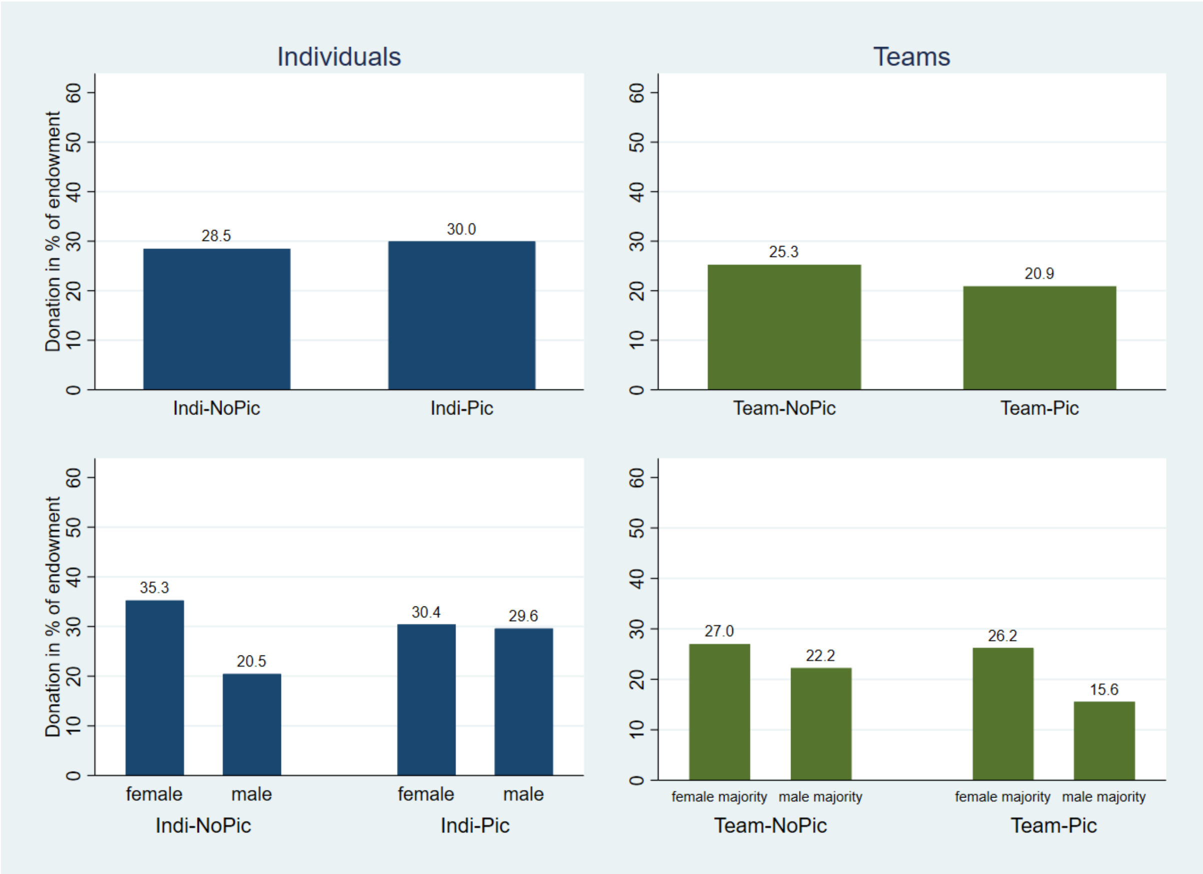
[15] The computer drew a random number for three teams that could not agree on a common donation within ten rounds. These observations were excluded from this analysis. The computer drew the donations randomly for the whole session. A donation of 25€ (56% of the endowment) was drawn for two teams that were in the same session and a donation of 27€ (60%) was drawn for the third team.

no significant differences between individuals and teams, neither for the anonymous treatments *Indi-NoPic* versus *Team-NoPic* (N = 59, p = 0.781) nor for the identified treatments *Indi-Pic* versus *Team-Pic* (N = 46, p = 0.256).

At first sight, it seems that showing pictures of the individual players and teams along with their donations to the observers does not increase donations. However, some differences arise when we consider men and women separately. The lower left part in Figure 3.1 shows the average donations of men and women when they decided as individuals. When deciding as individuals, women donated 35.3% of their endowment in *Indi-NoPic* and 30.4% in *Indi-Pic*. Women thus donated *less* when pictures were shown, but the difference is not statistically significant (MWW, N = 42, p = 0.545). Men, on the other hand, donated 20.5% in *Indi-NoPic* and 29.6% in *Indi-Pic*. This is an increase of 44% when pictures were shown and the difference is weakly significant (N = 42, p = 0.096). When we compare men and women, we find that women contribute significantly more in the anonymous treatment *Indi-NoPic* (N = 48, p = 0.010) while there is no significant difference in the identified treatments *Indi-Pic* (N = 36, p = 0.522). The differences between women and men become even more pronounced when we look at the proportion of subjects who make no donation and keep the entire endowment. In *Indi-NoPic*, 11.5% of women donated zero and a slightly larger proportion of 12.5% donated zero in *Indi-Pic* (Fisher's exact test, N = 42, p = 1.000). In contrast, while 40.9% of men did not donate anything in the anonymous treatment *Indi-NoPic*, this proportion decreased to only 10% in the *Indi-Pic* treatment when identities were revealed (Fisher's exact test, N = 42, p = 0.035). The lower proportion of zero donations by men in *Indi-Pic* increased the proportion of low and medium donations (smaller than or equal to one-third) compared to the anonymous treatment. The share of high donations (larger than one-third) is not affected. This behavior of men is in line with previous findings which showed that identification leads to more prosocial decisions especially by the least social (Samek and Sheremeta 2014). In short, the results show that women were relatively generous, irrespective of whether they were observed or not, while men were generous only when their pictures and donations were shown to the observers.

What can we say about the behavior of men and women when they act in teams? The lower right part of Figure 3.1 shows the average donations of teams which predominantly consist of women (at least two out of three team members) and teams which predominantly consist of men. Teams which predominantly consist of women made very similar average donations under both conditions (27% in *Team-NoPic* and 26.2% in *Team-Pic*, MWW, N = 12, p = 0.807). Teams that consisted mostly of men donated on average 22.2% in *Team-NoPic* and 15.6% in *Team-Pic* (N = 9, p = 0.602). The differences between more-male teams and more-female teams are not statistically different, neither in the anonymous treatments (N = 11, p = 0.702) nor in the identified treatments (N = 10, p = 0.242).

Figure 3.1 Average donations by treatment



This seem to suggest that the pictures had an effect on men when they decided as individuals but not when they decided in a team. However, the comparisons of team donations are not clean. Even when women were in the minority in the more-male teams, they had an influence on the final team decision. Likewise, men who were in the minority in the more-female teams still had an influence on the final team decision. This is why we also compare the proposals that men and women made initially when they had to decide in a team. These initial proposals are a good indicator for a person's preferred team donation and they are not yet influenced by the proposals made by other team members. Women made very similar proposals under both conditions. They proposed on average a donation of 30.7% of the endowment in *Team-NoPic* and 30.0% in *Team-Pic* (MWW, N = 39, p = 0.860). Men made *lower* proposals when the pictures were shown to the observers. On average, men proposed to donate 28.2% of the endowment in *Team-NoPic* and 18.3% in *Team-Pic.* The difference is not significant, though the p-value is not far from the 10% significance level (N = 33, p = 0.121). Comparing the proposals between men and women, we find no significant differences in the anonymous treatment *Team-NoPic* (N = 36, p = 0.700). In the identified treatment *Team-Pic* we find that the proposals made by women to be significantly higher than the proposals made by men (N = 36, p = 0.017).

The results so far suggest that the pictures had no significant effect on women irrespective of whether they decided as individuals or in a team while the pictures had a positive effect on men but only when they decided as individuals. A more formal test of these differences is provided in Table 3.2 where Tobit regression results on the donation decisions (for *Indi-NoPic* and *Indi-Pic*) and first proposals (for *Team-NoPic* and *Team-Pic*) are shown. Note that these regression analyses are merely used as a tool to show interaction effects between gender (men vs. women), anonymity (pictures vs. anonymous), and the type of decision (individual vs. team). They allow us to directly compare the effects of identification between individuals and teams as well as between women and men. The regressions contain all players and the shown coefficients are discrete effects (all explanatory variables are dummy variables) on the uncensored latent variable (McDonald and Moffitt 1980). The first two regressions in Table 3.2, (1) and (2), use male participants as baseline while the regressions (3) and (4) use female participants as baseline. We furthermore present regression results with and without additional control variables. The controls are field of study, self-reported trust, and self-reported importance of other people's opinions. They were all elicited prior to the game. Before the participants received the experimental instructions, they were asked about their field of study, how much they care about what other people think of them on a scale from 1 to 10, and how much they agree with the statement that in general people can be trusted on a scale from 1 to 10. To include these variables in the regressions, the answers to the two Likert-type questions are coded as dummy variables that take the value one if the answer is 6 or higher and zero otherwise. For the field of study, four main categories are formed of which three are included as dummy variables and the remaining one serves as baseline category. As can be seen in Table 3.2, the results of the regressions remain almost identical when these control variables are added. The variable *Pic* captures the effect of the pictures on the individual donations. It shows that women do not significantly increase their donations when pictures are shown. Men, in contrast, increase their donations (the latent uncensored donations) by almost 15 percentage points when they are identified. The variable *Team* displays the difference between the anonymous individual donations and the first proposals in the anonymous team decision. For both women and men these dummies are insignificant, indicating that, under anonymity, team proposals do not significantly differ from the individual decisions. The interaction term of *Team* and *Pic* is significant for men but not for women. Thus, while women do not react differently to identification when they decide individually or in a team, the effect of identification on men is significantly greater when they decide individually compared to team decisions. The dummy variables *Female* and *Male* demonstrate that women donate significantly more than men in *Indi-NoPic*. The interaction term between *Pic* and *Female* (*Male*) is not significant, though it is very close to the 10% significant level ($p = 0.102$ without control variables and $p = 0.112$ with control variables). If we run the same regression with teams as a baseline, the p-value of the interaction term between *Pic* and *Female* is well above the 10% significance level ($p = 0.309$ without control variables and $p = 0.370$ with control variables). Finally, the interaction term *Pic*Team* differs significantly between women and men, which is demonstrated by the significant three-way interaction of *Pic*Team*Female*. Thus, the different reactions of men to identification depending on whether they decide alone or in a team are significantly different from how women react to identification under the different conditions.

Now we turn to proportion of men and women who proposed to donate zero in their very first proposal. Men proposed to donate zero more often than women in both team treatments. The difference is not significant in *Team-NoPic* (13% vs. 23.1%, Fisher's exact test, N = 36, p = 0.645) but it is significant in *Team-Pic* (6.3% vs. 40%, N = 36, p = 0.026). Identification does not make a significant difference for the share of zero proposals in the team treatments, neither for women (N = 39, p = 0.631) nor for men (N = 33, p = 0.456).

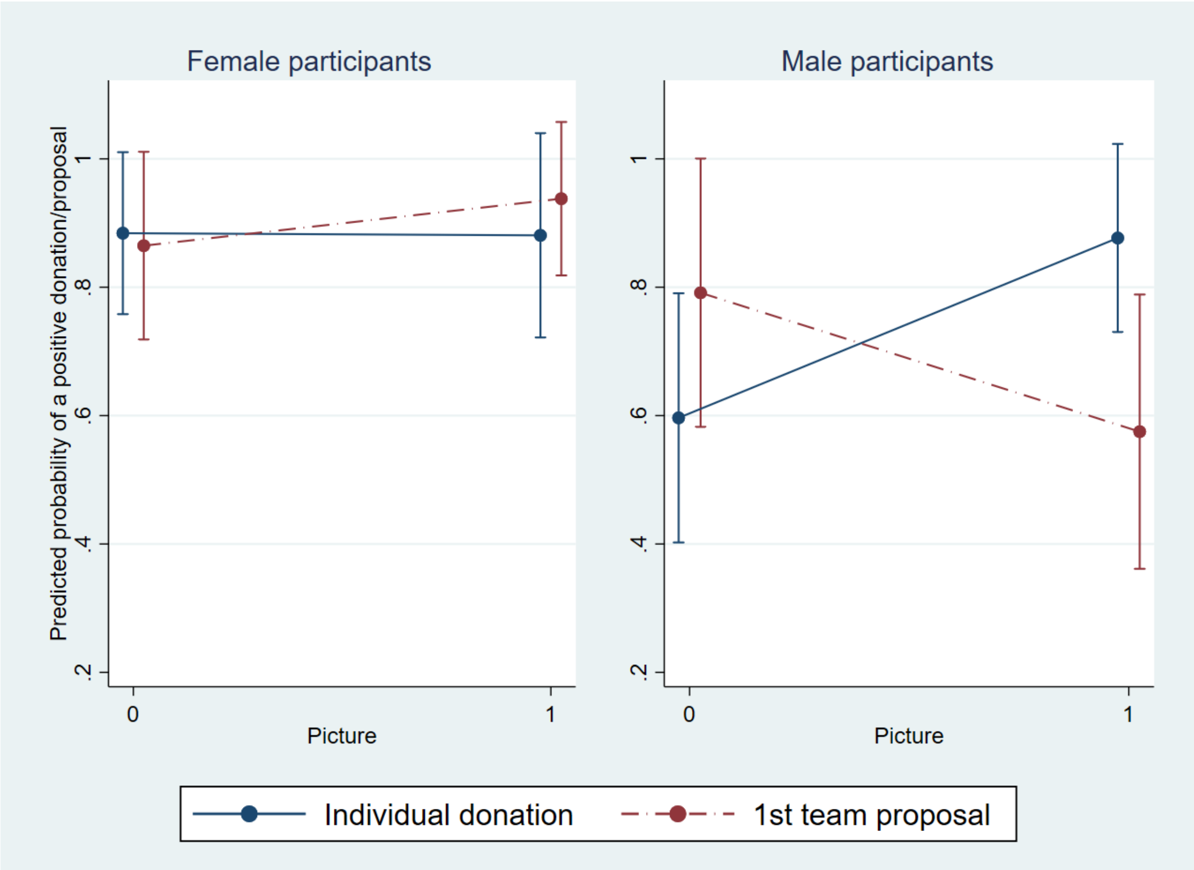
When we compare the proportion of zero proposals in the team treatments and zero donations in the individual treatments, we do not find significant differences for women, neither in the anonymous treatments (11.5% in *Indi-NoPic* vs. 13% in *Team-NoPic*, Fisher's exact test, N = 49, p = 1.000) nor in the identified treatments (12.5% in *Indi-Pic* vs. 6.25% *in Team-Pic*, N = 32, p = 1.000). For men, the proportion of zero donations in *Indi-NoPic* is higher than the proportion of zero proposals in *Team-NoPic*, though the difference is not statistically significant (40.9% vs. 23.1%, N = 35, p = 0.463). The order reverses in the identified treatments where 10% of men donated zero in *Indi-Pic* and 40% proposed to donate zero in *Team-Pic* (N = 40, p = 0.065).

Table 3.2 Tobit regression results on donations and team proposals including interactions of pictures with team and gender

	(1)	(2)	(3)	(4)
	Individual donation or first team proposal			
	Male participants are the baseline		Female participants are the baseline	
Pic (d)	14.75*	14.64*	-5.075	-4.249
	(8.501)	(8.298)	(8.524)	(8.423)
Team (d)	11.03	10.21	-4.834	-1.951
	(9.638)	(9.395)	(7.679)	(7.581)
Pic*Team (d)	-27.96**	-29.24**	5.266	1.095
	(13.01)	(12.65)	(12.19)	(12.04)
Female (d)	19.92**	19.04**		
	(7.987)	(7.829)		
Male (d)			-19.92**	-19.04**
			(7.987)	(7.829)
Pic*Female (d)	-19.83	-18.89		
	(12.04)	(11.81)		
Pic*Male (d)			19.83	18.89
			(12.04)	(11.81)
Team*Female (d)	-15.86	-12.16		
	(12.32)	(12.01)		
Team*Male (d)			15.86	12.16
			(12.32)	(12.01)
Pic*Team*Female (d)	33.22*	30.33*		
	(17.83)	(17.34)		
Pic*Team*Male (d)			-33.22*	-30.33*
			(17.83)	(17.34)
Constant	13.89**	15.59**	33.80***	34.63***
	(6.028)	(7.380)	(5.256)	(6.646)
Control variables	No	Yes	No	Yes
Observations	156	156	156	156

Tobit regression: left-censoring limit: 0, right-censoring limit: 100. Coefficients are discrete effects on the latent variable. Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Depended variable is the individual donation in *Indi-NoPic* and *Indi-Pic* and the proposal in the first round in *Team-NoPic* and *Team-Pic* as percentage of endowment. (d) indicates dummy variable. Definition of variables: *Pic* = 1 if identities are revealed, 0 otherwise; *Team* = 1 if decision is made in a team, 0 otherwise; *Pic*Team* = Interaction dummy of *Pic* and *Team*; *Female* = 1 if female participant, 0 otherwise; *Male* = 1 if male participant, 0 otherwise; *Pic*Female (Male)* = Interaction dummy of *Pic* and *Female* (*Male*); *Team*Female* (*Male*) = Interaction dummy of *Team* and *Female* (*Male*); *Pic*Team*Female* (*Male*) = Interaction dummy of *Pic*, *Team* and *Female* (*Male*). Control variables include field of study, trust, and in how far subjects care about others' opinions.

Figure 3.2 Likelihood of making a strictly positive donation or proposal



A graphical illustration of the different reactions to identification depending on the type of decision and gender is provided in Figure 3.2. The figure is based on a series of Probit models on the decision to make or propose a zero donation instead of a positive donation. The corresponding regression results are shown in the Appendix. Table 3.3 in the Appendix presents the results of the regressions and Table 3.4 shows the marginal discrete effect of identification for women and men when deciding alone or as team members. Figure 3.2 shows the interaction terms graphically as suggested by Greene (2010) for interaction terms in non-linear models. The left part of the figure shows that women do not change their behavior much when pictures are shown, neither for individual decisions nor for team decisions. By contrast, men react very differently to identification depending on whether they decide alone or in a team. The pictures have a positive effect on men's willingness to make a positive donation when they decide alone ($p = 0.025$) while they have a negative though insignificant effect when they decide in a team ($p = 0.157$). As can be seen in Table 3.3, the effect of identification is significantly different for men depending on whether they decide alone or within a team.

Finally, we want to emphasize that the differences in men's behavior depending on whether they are observed or not and whether they decide alone or in a team cannot be explained by different beliefs about the observers' expectations. Men have similar second-order beliefs in *Indi-NoPic* and *Indi-Pic* (34.5 vs. 30.3, MWW $p > 0.1$) as well as in *Team-NoPic* and *Team-*

Pic (34.0 vs. 34.9, $p > 0.1$). This suggests that the differences in behavior are driven by the diffusion of responsibility within the team.

3.5 Discussion and conclusion

The situation that we created in this experiment was arguably artificial for the participants. It is rarely the case in real life that we make a donation decision under these rather odd circumstances. On the other hand, these circumstances allow us to make clean comparisons and identify causal effects which is not possible if we opt for a more natural decision environment. Our goal in this experiment was to exogenously control for the moral standard and test if individuals and groups react differently to this standard when they are observed by others and when they are not observed. With this, we want to contribute to the discussion of why groups behave differently than individuals. While the previous literature has shown that groups tend to make more self-interested decisions, it is not clear if this happens because there are different moral standards for groups or if members of groups care less about moral standards.

Our first important result is that, overall, identification has no significant effect on donations in this setting. A plausible explanation for this is that participants in the experiment were explicitly made to think about the appropriate donation amount and that this limited the additional effects of identification. Nevertheless, a more detailed analysis reveals some differences between women and men that are not visible in the overall averages.

Women's donations are relatively stable across the different conditions. They make relatively generous donations regardless of whether they are observed or not. Men, by contrast, make higher donations when they are observed by others. The increase is mostly driven by the share of men who make a positive donation only when they know that their decision will be observed. These results suggest that women are more intrinsically motivated while men are more motivated by image concerns which is in line with the findings on gender differences in helping behavior (Eagly and Crowley 1986); Eagly 1987; Eagly and Wood 1991). Interestingly, when we asked our participants prior to the experiment about how much they care about what other people think of them, women reported to care more than men (6.5 vs. 5.5, MWW, $N = 156$, $p = 0.001$). The self-assessments thus are not consistent with the behavior in the experiment. We also find no evidence that individuals who say that they care a lot about others' opinions react stronger to the pictures than others. This finding can serve as a reminder that we should handle hypothetical answers with care.

An important goal of our study was to test if and how the effect of identification changes if subjects do not decide alone but in a team. Women do not react strongly to identification, neither when they decide alone nor when they decide in a team. Men, by contrast, react strongly to identification when they decide alone but not when they decide in a team. If anything, identification leads to lower rather than higher proposals within teams, though the effect is not statistically significant. Men thus react only to identification when they can claim full responsibility for the prosocial act. Importantly, the different reactions to identification when men decide alone or in a team cannot be explained by different standards for groups. After our manipulation of expectations, men have on average very similar beliefs about what the observers will consider an appropriate donation in all conditions. The different reactions to identification have to be explained therefore with the diffused responsibility within the group.

A straightforward implication of our research is that fundraisers and managers who rely on voluntary contributions should consider different measures for men and women. The provision of status and prestige seems to work especially well for men who can claim full responsibility for the decision. Increasing transparency and the provision of prestige and status may not be able to change the behavior of groups. Regulating the behavior of groups may require stronger measures, especially when responsibility within the group is diffused.

Acknowledgements: The work was financially supported by the European Union (EU) Horizon 2020 program, action ERC-2014-STG, Project HUCO, grant number 636746.

3.6 Appendix

3.6.1 Experimental Instructions for *Indi-Pic* (translated from German)

In this experiment you can earn money. You will receive your payoff in cash at the end of this experiment. For a successful experiment, you are not allowed to speak to the other participants during the experiment. Please read the following instructions carefully. If you have any question, please raise your hand. We'll come to your place and answer your questions.
There are two different roles in this experiment: players and observers. The players actively make a decision that determines their payoff. The observers make no decision and receive a fixed payment. You will be randomly assigned to one role. You will be informed about your role at a later stage. At this stage the observers will also be informed about their fixed payoff.
The players will make two decisions. Note that you will only be paid for <u>one</u> of the two decisions (decision 1 <u>or</u> decision 2). It will be randomly determined which decision will be relevant for the payoff. Both decisions are chosen with the same probability. Therefore, the players should make both decisions as if they were selected.

Decision 1
Every player receives 15 Euro. The player decides, whether he/she keeps the 15 Euro or donates part of it to the charitable organization German Red Cross. Any amount between 0 and 15 Euro is possible. The player will keep any money that has <u>not</u> been donated. Each player makes his/her decision on his/her own and independent of the other players.
If decision 1 is implemented at the end of the experiment, the donations of all players will be shown on the screens of the observers. The computer will order the donations from highest to lowest. Next to each donation, a <u>picture</u> of the player, who has chosen the donation, will be shown. The players will only see the donations of all players, but not their pictures. Please note that the pictures will not be published outside of this experiment and will not be given to third parties. After the experiment all pictures will be deleted.
The observers add up the donations of all players and enter the sum into the computer. Every observer does this on his/her own and independent of all other observers.
One of the experimenters will transfer the sum of all donations to the charitable organization German Red Cross. The observers will observe the transaction. The player will receive the share of the money that they didn't donate. The observers will get their fixed payoff.
If decision 1 is not implemented, the donations will not be shown.

Decision page of the players

Decision 1
You are a player and you will receive 15 Euro. Please state now, how much of this money you want to donate to the charitable organization German Red Cross. Every amount between 0 and 15 Euro is possible.
If decision 1 is implemented at the end of the experiment, the donations of all players will be shown on the screens of the observers together with the pictures of the respective players. The observers will add up the sum of all donations. The players will only see the donations of all players, but not their pictures.
<u>For your information</u>: In a short survey in which 38 uninvolved students were asked which amount should be donated in such a situation, the average answer was 64 percent.
My donation: _________ Euro.

Decision page of the observers

Decision 1
You are an observer. You will receive a fixed amount of 10 Euro as your payoff and you will not make any decisions. Please wait until all players have made their decision.
Which donation (in %), do you think, is appropriate in this situation? _________%

3.6.2 Supplementary regression analyses

Table 3.3 presents a probit regression on the likelihood to make a positive donation in the individual treatments or a positive first proposal in the team treatments. Women and men are examined in separate regressions in order to make the results more comprehensible. The coefficient of *Pic* in regression (1), which includes only female participants, demonstrates that the pictures have no significant effect on the likelihood to make a positive donation when deciding on the individual donations. Positive proposals in *Team-NoPic* are also not more likely than positive donations in *Indi-NoPic* for women, as indicated by the insignificant coefficient of *Team*. Likewise, the interaction term *Pic*Team* remains insignificant. Identification and deciding in a team have also no significant impact on whether women are in favor of donating or not when control variables are added, as can be seen in regression (2). Men, in contrast, are more likely to make a positive donation in *Indi-Pic* than in *Indi-NoPic*, as indicated by the significant coefficient of *Pic* in the regressions (3) and (4). Deciding on the team proposal instead on the individual donation has no significant impact on the likelihood of choosing a positive donation. The significant interaction term indicates that the reaction to identification differs significantly depending on whether men decide individually or in a team.

Table 3.3 Probit regressions on the decision to make a positive donation or proposal versus zero donation

	(1)	(2)	(3)	(4)
	Individual donation or first proposal			
	Female participants		Male participants	
Pic (d)	-0.0480	-0.0175	1.052**	1.014**
	(0.515)	(0.541)	(0.468)	(0.491)
Team (d)	-0.0740	-0.0962	0.506	0.629
	(0.462)	(0.491)	(0.469)	(0.502)
Pic*Team (d)	0.458	0.461	-1.535**	-1.704**
	(0.785)	(0.830)	(0.668)	(0.713)
Constant	1.198***	0.916*	0.230	0.319
	(0.322)	(0.474)	(0.270)	(0.477)
	No	Yes	No	Yes
Observations	81	81	75	75

Probit regression. Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Depended variable is the likelihood to make a positive donation in *Indi-NoPic* and *Indi-Pic* and a positive proposal in the first round in *Team-NoPic* and *Team-Pic*. (d) indicates dummy variable. Definition of variables: *Pic* = 1 if identities are revealed, 0 otherwise; *Team* = 1 if decision is made in a team, 0 otherwise; *Pic*Team* = Interaction dummy of *Pic* and *Team*; Control variables include field of study, trust, and in how far subjects care about others' opinions.

Table 3.4 presents the estimated discrete effect of identification on the likelihood to make a positive donation in the individual decision and a positive first proposal in the team decision, separated for women and men. The two regressions with control variables in Table 3.3 were taken as the basis for this estimation.

Table 3.4 Marginal discrete effects of identification

	Female participants	Male participants
Individuals	-0.003	0.280**
	(0.105)	(0.125)
Teams	0.073	-0.216
	(0.098)	(0.153)

Estimated marginal discrete effects of identification on the likelihood to make a positive donation or a positive first proposal. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5 presents regressions that include only those subjects who made a positive donation or positive first proposal. It turns out that neither the pictures nor deciding in a team has a significant effect on the decision of women or men. Thus, while we see that identification increases the share of men who make a positive donation, the level of the donation is not affected.

Table 3.6 shows regressions that use the final donation in the team treatments as the dependent variable instead of the first proposals. Therefore, the donation decision of the three team members provides only one observation, which reduces the number of observations in the team treatments by two-thirds. The dummy variable *Female* takes the value one if at least two of the three team members are women, the variable *Male* was created accordingly. Male participants in *Indi-NoPic* constitute the baseline in the regressions (1) and (2) and female participants in *Indi-NoPic* in the regressions (3) and (4). The results for the individuals basically reflect the results shown in the main paper. The variable *Team* is insignificant indicating that donations are not different between *Indi-NoPic* and *Team-NoPic* for both male and female participants. The interaction term *Pic*Team*, which is significant for men in Table 3.2, is now insignificant for both male and female participants. Thus, also more-male teams do not react significantly different to identification than male participants deciding individually when we compare final donation decisions. This finding can be explained by the lower number of observations in the team treatments but also by the fact that many more-male teams include also female participants which influences the final donation. We ran additional regressions that used more-female and more-male teams as a baseline (not shown). The results confirm that teams do not make significantly different donations when they are identified.

Table 3.5 OLS regressions using only participants who made or proposed a strictly positive donation

	(1)	(2)	(3)	(4)
	Individual donation or first proposal			
	Male participants are the baseline		Female participants are the baseline	
Pic (d)	-1.745	-1.082	-5.093	-4.564
	(7.236)	(7.150)	(6.739)	(6.812)
Team (d)	2.051	0.537	-4.522	-4.351
	(8.362)	(8.276)	(6.078)	(6.144)
Pic*Team (d)	-4.366	-3.811	1.760	0.874
	(11.17)	(11.07)	(9.566)	(9.655)
Female (d)	5.240	6.279		
	(6.898)	(6.865)		
Male (d)			-5.240	-6.279
			(6.898)	(6.865)
Pic*Female (d)	-3.348	-3.483		
	(9.888)	(9.825)		
Pic*Male (d)			3.348	3.483
			(9.888)	(9.825)
Team*Female (d)	-6.573	-4.888		
	(10.34)	(10.25)		
Team*Male (d)			6.573	4.888
			(10.34)	(10.25)
Pic*Team*Female (d)	6.126	4.686		
	(14.71)	(14.51)		
Pic*Team*Male (d)			-6.126	-4.686
			(14.71)	(14.51)
Constant	34.62***	38.94***	39.86***	45.22***
	(5.514)	(6.527)	(4.145)	(5.476)
Control variables	No	Yes	No	Yes
Observations	125	125	125	125

OLS regression. Coefficients are discrete effects on the dependent variable. Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Depended variable is the individual donation in *Indi-NoPic* and *Indi-Pic* and the proposal in the first round in *Team-NoPic* and *Team-Pic* as percentage of endowment. (d) indicates dummy variable. Definition of variables: *Pic* = 1 if identities are revealed, 0 otherwise; *Team* = 1 if decision is made in a team, 0 otherwise; *Pic*Team* = Interaction dummy of *Pic* and *Team*; *Female* = 1 if female participant, 0 otherwise; *Male* = 1 if male participant, 0 otherwise; *Pic*Female (Male)* = Interaction dummy of *Pic* and *Female* (*Male*); *Team*Female* (*Male*) = Interaction dummy of *Team* and *Female* (*Male*); *Pic*Team*Female* (*Male*) = Interaction dummy of *Pic*, *Team* and *Female* (*Male*). Control variables include field of study, trust, and in how far subjects care about others' opinions.

Table 3.6 Tobit regressions on individual and team donations

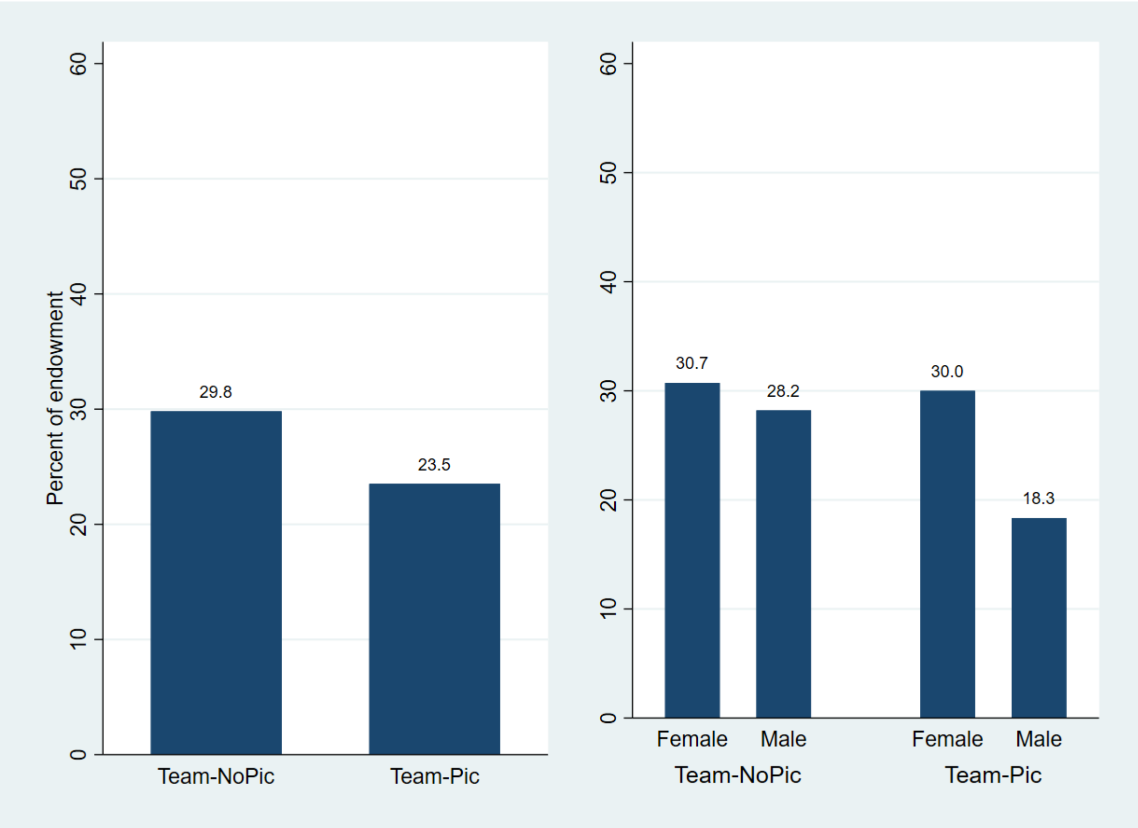
	(1)	(2)	(3)	(4)
	Individual donation or team donation			
	Male participants in *Indi-NoPic* are the baseline		Female participants in *Indi-NoPic* are the baseline	
Pic (d)	14.61*	14.87*	-5.070	-3.877
	(8.344)	(8.300)	(8.362)	(8.346)
Team (d)	4.337	2.537	-8.842	-7.125
	(14.65)	(14.61)	(11.24)	(11.25)
Pic*Team (d)	-24.65	-24.68	6.290	1.336
	(20.10)	(20.08)	(17.47)	(17.68)
Female (d)	19.79**	19.14**		
	(7.839)	(7.812)		
Male (d)			-19.79**	-19.14**
			(7.839)	(7.812)
Pic*Female (d)	-19.68*	-18.75		
	(11.81)	(11.78)		
Pic*Male (d)			19.68*	18.75
			(11.81)	(11.78)
Team*Female (d)	-13.18	-9.662		
	(18.47)	(18.59)		
Team*Male (d)			13.18	9.662
			(18.47)	(18.59)
Pic*Team*Female (d)	30.94	26.02		
	(26.64)	(26.75)		
Pic*Team*Male (d)			-30.94	-26.02
			(26.64)	(26.75)
Constant	14.05**	10.43	33.84***	29.57***
	(5.924)	(6.822)	(5.157)	(6.602)
Control variables	No	Yes	No	Yes
Observations	105	105	105	105

Tobit regression: left-censoring limit: 0, right-censoring limit: 100. Coefficients are discrete effects for the latent variable. Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Depended variable is the individual donation in *Indi-NoPic* and *Indi-Pic* and the final team donation in *Team-NoPic* and *Team-Pic* as percentage of endowment. (d) indicates dummy variable. Definition of variables: *Pic* = 1 if identities are revealed, 0 otherwise; *Team* = 1 if decision is made in a team, 0 otherwise; *Pic*Team* = Interaction dummy of *Pic* and *Team*; *Female* =1 if female participant, 0 otherwise; *Male* = 1 if male participants, 0 otherwise; *Pic*Female(Male)* = Interaction dummy of *Pic* and *Female* (*Male*); *Team*Female* (*Male*) = Interaction dummy of *Team* and *Female* (*Male*); *Pic*Team*Female* (*Male*) = Interaction dummy of *Pic*, *Team* and *Female* (*Male*). Control variables include field of study, trust, and in how far subjects care about others' opinions.

3.6.3 Supplementary graphics

3.6.3.1 Average first proposals

Figure 3.3 First proposals made in the team treatments



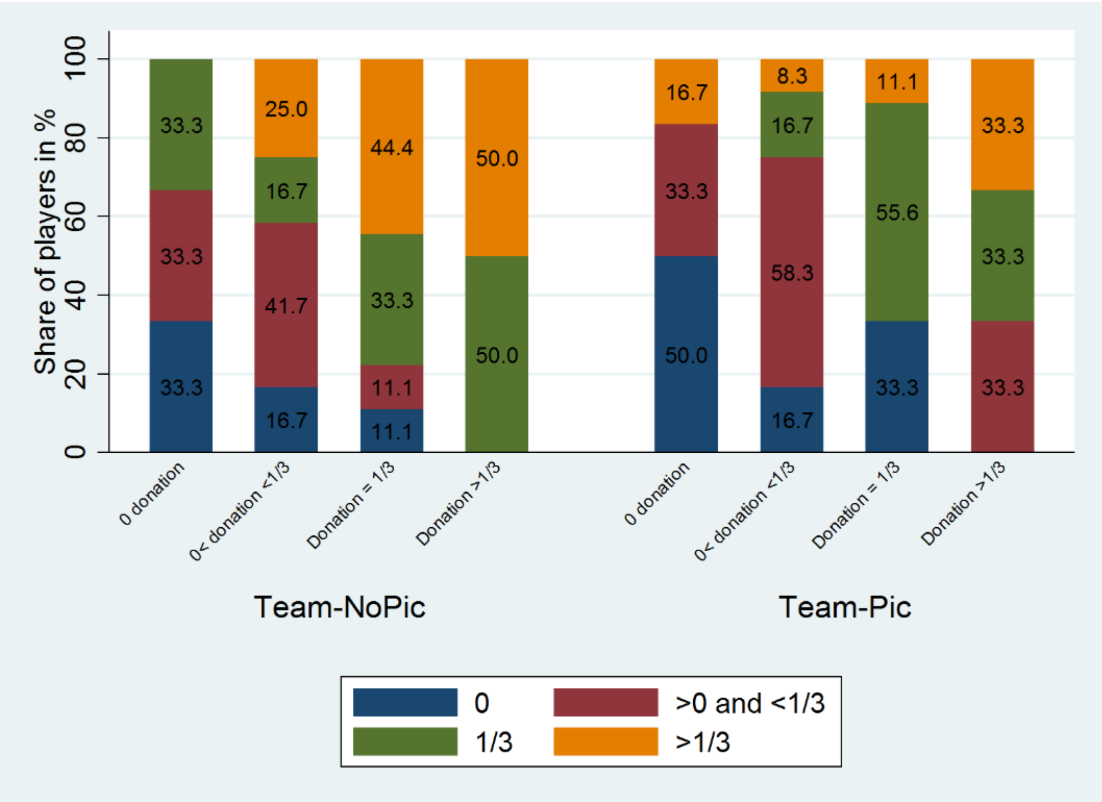
3.6.3.2 From first proposals to final team donations

Final donations of teams were often slightly lower than the initial proposals made by the team members. Participants in *Team-NoPic* proposed on average a donation of 29.8% of their endowment, while their teams donated 25.3%. Participants in *Team-Pic* proposed 23.5% and donated 20.9%.

Figure 3.4 presents the proportion of players who proposed zero, below one third, exactly one third, and above one third of the endowment in the first round of proposals separated by the final team donation. For *Team-NoPic* the picture is clear: The level of the team donations increases with a lower share in low proposals and a higher share in high proposals. One third of the participants in the teams that donated nothing proposed to do so. For positive donations below one third of the endowment and donations of one third of the endowment, donations are equal to the median proposal. For donations above one third, half of the subjects were in favor of a very high donation and half were in favor of donating one third. It seems that a high share of players in favor of a high donation in combination with an absence of players who proposed to donate nothing or very little is needed in order to end up with a high donation in the treatment. In order for their teams to make a zero donation a minority of players who proposed to so was sufficient.

In *Team-Pic*, donations below one third of the endowment and one third of the endowment were also equal to the median proposal. However, the proportion of players who proposed to donate nothing needed to be much higher (50% over all teams that made a zero donation) than in the anonymous team treatment. On the other hand, a minority of players in favor of a very high donation was sufficient in order to convince their teams to make such a donation. The fact that initial proposals were still higher than final donations in the treatment can be explained by those teams that couldn't find a solution. When these teams are excluded, the average first proposal in *Team-Pic* is only 20.7% (30% *Team-NoPic*). So, in *Team-Pic* less people in favor of a high donation were needed in order to end up with a high donation. At the same time initial proposals were much lower in the treatment, which resulted in lower donations in *Team-Pic* than in *Team-NoPic*.

Figure 3.4 Team donations and first proposals



4 What Drives States to Sign Water Allocation Treaties in International River Basins?

Sven Christens

Department of Economics, University of Kassel, 34109 Kassel, Germany

Abstract: Water is becoming scarcer in many of the world's river basins as a result of a growing population and a changing climate. The use of water in international rivers is, therefore, often the cause for tensions between countries. River basins with an upstream and a downstream country are in particular prone to conflict since their riparian states' potential to withdraw water from the river is highly asymmetric. Although upstream states have no direct incentive for cooperation, numerous treaties that define the allocation of water from upstream to downstream countries have been signed in the past. This paper runs a Cox proportional hazards model in order to examine the factors that lead to the formation of such treaties. I test in particular whether the power of the downstream state relative to the upstream state has an effect on the likelihood to sign a treaty. The results indicate that the relative economic and military power of the downstream state is not the main driver for the formation of allocation treaties. The downstream state's available water, its exposure to floods and droughts and dam constructions in the upstream country are far more decisive.

4.1 Motivation

Water is a crucial resource for human life but water is scarce in many regions of the world. Around 2 billion people live in watershed areas exposed to water scarcity (Gosling and Arnell 2016). These people depend heavily on rivers for nutrition, agriculture, industries, fisheries, hydropower and so on. Many of these rivers are shared by several countries. Around 52% of the world's population lives in basin areas of international rivers (McCracken and Wolf 2019). There are numerous ongoing tensions between states over these shared water sources. Conflicts are particularly common when states are in an upstream-downstream relation (also called through-border configuration), since this setting allows upstream states to exploit the river's resources and to create negative externalities for downstream riparian states (Toset, Gleditsch, and Hegre 2000).

Many international river treaties have emerged over the past decades in order to resolve existing and potential disputes. Treaties over the allocation of water can play a crucial role in mediating conflicts and promoting cooperation (Brochmann 2012; Dinar et al. 2015). These treaties reduce the risk for downstream states to receive lower or more varying volumes of water from the upstream riparian. In that sense, the allocation of water clearly benefits the downstream state, while it comes with potential costs to the upstream state and limits its ability to exploit the river's water. In recent years, a number of studies has analyzed factors that lead to the formation of treaties over international rivers. If these studies include power in the analysis, they usually test the effect of economic and military power asymmetries by taking the ratio of the more powerful to the less powerful state. The idea behind this approach is that hegemons can provide stability and, thus, enable formalized cooperation in the form of treaties. Most studies support the hypothesis that treaties become more likely with higher power imbalances (Song and Whittington 2004; Tir and Ackerman 2009; Dinar, Dinar, and Kurukulasuriya 2011).

However, the interaction between a country's location along the river and its willingness to engage in formal cooperation in the form of a treaty has often been neglected in the empirical literature. I argue that it is important to set the beneficiary's power in relation to the riparian state that would bear potential costs when an allocation treaty is signed. To this end, I test the role of the downstream state's power relative to that of the upstream state in an empirical analysis. This approach allows to test whether downstream states as main beneficiaries of allocation treaties use their power in order to spur cooperation. Do more powerful downstream states impose their interest for formal regulation of the water allocation upon weaker upstream states? Do relatively powerful upstream states prevent the formation of allocation treaties?

This study also introduces local data on country-basin level to control for water availability and exposure to floods and droughts in the basin area and to capture the role of dams. Previous research often relied on national averages, which does not allow to take into account the basin specific dynamics. While a state's power is determined nationally, its interest in a treaty stems from local conditions. Examining the local conditions provides important insides into the role of dependency on water resources, which is required for a deeper understanding of the interests driving cooperation or disputes between riparian states.

For the analysis, I use a Cox Proportional Hazard model that examines the formation of treaties between dyadic riparian states sharing an international river basin. The results show that the downstream state's power is not the driving force in the formation of treaties. The downstream state's dependency on the river and insecurity created by dam projects in the upstream state's basin area are far more important factors.

The remainder of this paper is organized as follows. The next section presents a literature overview on the factors that impact choices of states to sign treaties over international water basins. The subsequent section describes the underlying data of the estimations. Section 4 discusses the statistical approach. The results of the empirical analysis are presented in Section 5. The final section offers the conclusion.

4.2 Background

4.2.1 Water allocation in international river basins

International river basins account for 80% of all global river flows and are a crucial source of freshwater (Food and Agriculture Organization of the United Nations 2019). The location of a state along the river is decisive for the use of the river. Throughout all parts of rivers, irrigation is by far the dominant reason for the withdrawal of water (Munia et al. 2016). Domestic and industrial use can also substantially attribute to water consumption. Hydropower is in particular important in the upper parts of the river, while access to the sea is beneficial for lower parts of the river.

In general, the relationship between upstream and downstream riparian states is highly asymmetric, especially with regard to the ability to withdraw the river's water. The use of the river's resources can create a number of negative externalities for downstream states, such as reduced stream flows and water stress (Rogers 1993; Munia et al. 2016), while upstream states are largely unaffected by the downstream state's behavior (Barrett 2003, 70). Riparian states are similarly asymmetric with regard to pollution flowing from an upstream country to a downstream country (Mitchell and Keilbach 2001; Barrett 2003, 70). Because of this asymmetry and the upstream states' unlimited access to the water resources, upstream states have little incentives to guarantee fixed or flexible volumes of water to downstream states and, thereby, limit their use of the river's resources (Mitchell and Keilbach 2001). The only assets owned by downstream states are access to the sea (Warner and Zawahri 2012) and control over fish migration (Tir and Ackerman 2009).

Until today there are no binding rules concerning international river basins and no supranational organization enforcing laws. Nonetheless, states acknowledged the need for an equitable water use at the United Nations (UN) Stockholm conference in 1972 (United Nations 1973). The first UN Water Conference in Mar del Plata in 1977 (United Nations 1977) recognized people's right to adequate drinking water (Gupta, Ahlers, and Ahmed 2010). Another important meeting was the Dublin Statement on Water and Sustainable Development (International Conference on Water and the Environment 1992) which tried to establish the Integrated Water Resource Management (IWRM) as a guideline for the sustainable use of water resources. The concept aims at the international development and management of water resources. But the impact of these meetings remains marginal and their principles are hardly applied in international river basins (Biswas 2004). In 1997, the UN Convention on the Law of the Non-Navigational Uses of International Watercourses was the UN's attempt to create an international guideline for the sustainable use of freshwater resources (United Nations 1997). The convention provides a legal framework for claims of downstream countries to the resources of a watercourse by establishing the concept of equitable water sharing and joint management (Gupta 2016). So far only 36, mostly downstream, countries ratified the convention, which has led to its entry into force in 2014. However, without the support of upstream states this framework remains without implications for countries' behavior regarding international water resources.

Due to the lack of internationally binding rules, conflicts are common, in particular in downstream and upstream relations that create externalities from the upstream state to the downstream state (Brochmann and Gleditsch 2012b). Water quantity issues and related infrastructure projects are the main causes for conflicts over international rivers (Yoffe, Wolf, and Giordano 2003). Conflicts occur also more frequently in water scarce regions (Tir and Stinnett 2012; Devlin and Hendrix 2014). Riparian states have to rely on bilateral or multilateral negotiations in order to solve disputes and to agree on rules for the use of the river's water. In order to formalize their cooperation, states rely on treaties. These treaties became more complex over the past decades. Nowadays they often regulate multiple issues such as the common management and water quality. Still, allocation of water remains the most important subject of these treaties (Giordano et al. 2014). The mere existence of a treaty is not found to automatically reduce conflicts (Dinar et al. 2019). However, once a treaty is signed, cooperation between riparian states tends to increase (Brochmann 2012; Dinar et al. 2019). Treaties also reduce dam constructions of upstream states that come at the expense of downstream states (Olmstead and Sigman 2015). Allocation mechanisms are most effective in reducing conflicts between riparian states when they are both flexible in dealing with water variability and if they specify a concrete allocation mechanism (Dinar et al. 2015). Treaties that address water quantity (and hydropower) issues are usually the subject of bilateral treaties between states (Yoffe, Wolf, and Giordano 2003) even in multilateral basins (Zawahri, Dinar, and Nigatu 2016).

4.2.2 Power and cooperation in international rivers

Many studies that examine power in international river negotiations focus on the role of hegemons. The existence of a hegemon itself is assumed to enable cooperation by providing stability. In this understanding hegemons are able to provide mutually beneficial and stable outcomes (Milner 1992). They can also be necessary to initiate negotiations and to sustain cooperation. Sometimes a richer or more powerful state can provide infrastructure that is beneficial for both states.

Most empirical studies on the formation of treaties or conflict in international river basins test the effect of power imbalances by dividing the more powerful state's value by the weaker state's value. Commonly economic and military power is examined, although military force is rarely used in water disputes (Yoffe, Wolf, and Giordano 2003). It has been found that high imbalances in wealth are correlated with conflicts in basins, while imbalances in population density are associated with cooperation (Yoffe, Wolf, and Giordano 2003). At the same time, river treaties become more likely with higher imbalances in wealth (Dinar, Dinar, and Kurukulasuriya 2011) and military capacity (Tir and Ackerman 2009). High differences in GDP seem not to be conducive to the formation of treaties (Dinar, Dinar, and Kurukulasuriya 2011) or to even impede it (Dinar et al. 2010). Song and Whittington (2004) find that treaties are less likely when riparian states are similarly powerful. They argue that those states might not be in need of formal cooperation because they already managed to find a "balance of power". Thus, a majority of studies supports the idea that hegemony in the form of power differences is conducive for cooperation in the form of treaties. In the present study, I use a different approach by taking the ratio of the downstream state's power relative to that of the upstream state. This procedure allows to examine whether the power of the downstream state, which benefits most from cooperation, is an important factor. Scholars have argued that cooperation is more likely in the presence of a downstream hegemon as downstream states benefit most from the cooperation (Lowi 1993). To the best of my knowledge, this idea has not been empirically tested in a large scale study so far. However, Dinar (2009) presents a number of case studies that demonstrate that even when the weaker state is located downstream, cooperation can emerge through the creation of incentives like issue-linkage, reciprocity and side-payments. Zawahri and Mitchell (2011) examine the role of power on the formation of treaties by controlling for the upstream and the downstream state's power separately. Both, the upstream state's power and the downstream state's power, tend to be positively correlated with the occurrence of bilateral treaties but negatively with the formation of multilateral treaties.

While power seems to be an important factor, scholars have also pointed out that the degree to which power explains cooperation is limited. Petersen-Perlman and Fischhendler (2018) argue that hegemons are more vulnerable with regard to international rivers than it might seem at first glance. Especially when the weaker state's survival depends on the river, these states might accomplish concessions by linking water related issues with non-water issues. Also Warner and Zawahri (2012) argue that while bilateral treaties in multilateral basins become more likely in the presence of a downstream hegemon, weaker states are not powerless in negotiations over transboundary rivers with hegemons. There might be less margin for non-hegemonic, weaker states when dealing with hegemons, but cooperation is in many cases also in the interest of the hegemon. Transaction costs for the unilateral use of a basin are high and a lack of coordination results in additional costs. Both states can, for example, benefit from the construction of hydropower plants in the upstream areas. India and Bhutan provide an example of how this cooperation can look like in international rivers. India, the hegemon in the relation, constructed hydropower dams in Bhutan that provide both states with cheap energy and generate additional sources of income for Bhutan (Biswas 2011).

4.2.3 Additional factors that lead to the signature of treaties

Previous empirical studies also identified a number of other factors that are associated with the signature of treaties over international rivers. Geography is one important factor. Treaties appear less frequently in downstream-upstream relations (Song and Whittington 2004). The economic and political relation of states in the basin is another factor. Economic dependency, measured as the volume of common trade, is found to facilitate treaties (Tir and Ackerman 2009; Espey and Towfique 2004). Democracies are more likely to sign treaties (Tir and Ackerman 2009). The likelihood of signing a treaty increases also with the same religion, the same language, and with the basin size (Espey and Towfique 2004).

Dinar et al. (2010) examine the effect of runoff and precipitation variability on the likelihood of the occurrence of treaties. The authors find an inverted U-shaped relation between variability in runoff and precipitation and cooperation in the form of treaties. As the variability increases, treaties become more likely but after a peak is reached the likelihood decreases. Dinar, Dinar, and Kurukulasuriya (2011) find the same inverted U-shape for water scarcity. As water is becoming scarcer cooperation becomes more likely but in regions with very severe water shortage cooperation decreases again. The studies do not differentiate whether water scarcity and variability occur in the upstream basin or the downstream basin.

4.2.4 Side payments and issue-linkage

A common way to offset geographical asymmetries in international river basins is the introduction of side payments. Mahjouri and Ardestani (2010) show in a game theoretical model that the introduction of side payments can result in Pareto improvements. When benefits from hydropower, additional agricultural production and saved pumping costs are weighted against the costs of construction and operational and maintaining costs, side payments can motivate a regime to shift and improve welfare for both countries that are involved. As the Coase Theorem (Coase 1960) would predict, side payments are commonly used to reduce pollution spillovers from upstream states to downstream states (Barrett 1994; Warner and Zawahri 2012).

Dinar (2009) uses case studies to show that linking interests and other political issues with the negotiations over the allocation of water can be a reason for hegemonic upstream states to cooperate with weaker downstream states. This is for example the case when these upstream states share other rivers with their counterpart in which they are the downstream state. A hegemon might as well expect reciprocity as a result to benign behavior towards a weaker state.

While side payments are a common tool to reach cooperation in the form of treaties, side payments are rare in treaties that deal with the allocation of water (Dinar 2006).

4.3 Data

4.3.1 Dependent variable

The information on the international river basins was taken from the Register of International River Basins (McCracken and Wolf 2019). Out of the 310 rivers in the database, 257 possess at least one upstream-downstream relation between any of their riparian states. These rivers enter the analysis. Each river shared by a country dyad is considered separately in the analysis so that the units analyzed are river-country dyads. Country dyads are excluded when no clear upstream-downstream pattern exists, for example in the case of border-forming rivers. In total, 796 river-country dyad combinations existed in the years between 1907 and 2007 in the basin areas of these rivers. Brochmann and Gleditsch (2012a) provide information on the years a country dyad shared an international river as well as on the existence of an upstream-downstream relation between the two countries. For those rivers that are not included in their dataset, I added the information manually. Only sovereign countries are included in the analysis. This excludes colonial territories from the analysis.

The data on international treaties was derived from the International Freshwater Treaties Database (Giordano et al. 2014; Transboundary Freshwater Dispute Database 2018), which provides information on the signature of treaties on international waters and their content until the year 2007.[16] States are considered to have an allocation treaty when a primary document, an amendment to an existing treaty, a replacement of a treaty, or a protocol to a primary agreement was signed that addresses water quantity or specifies the allocation of water from one country to another.

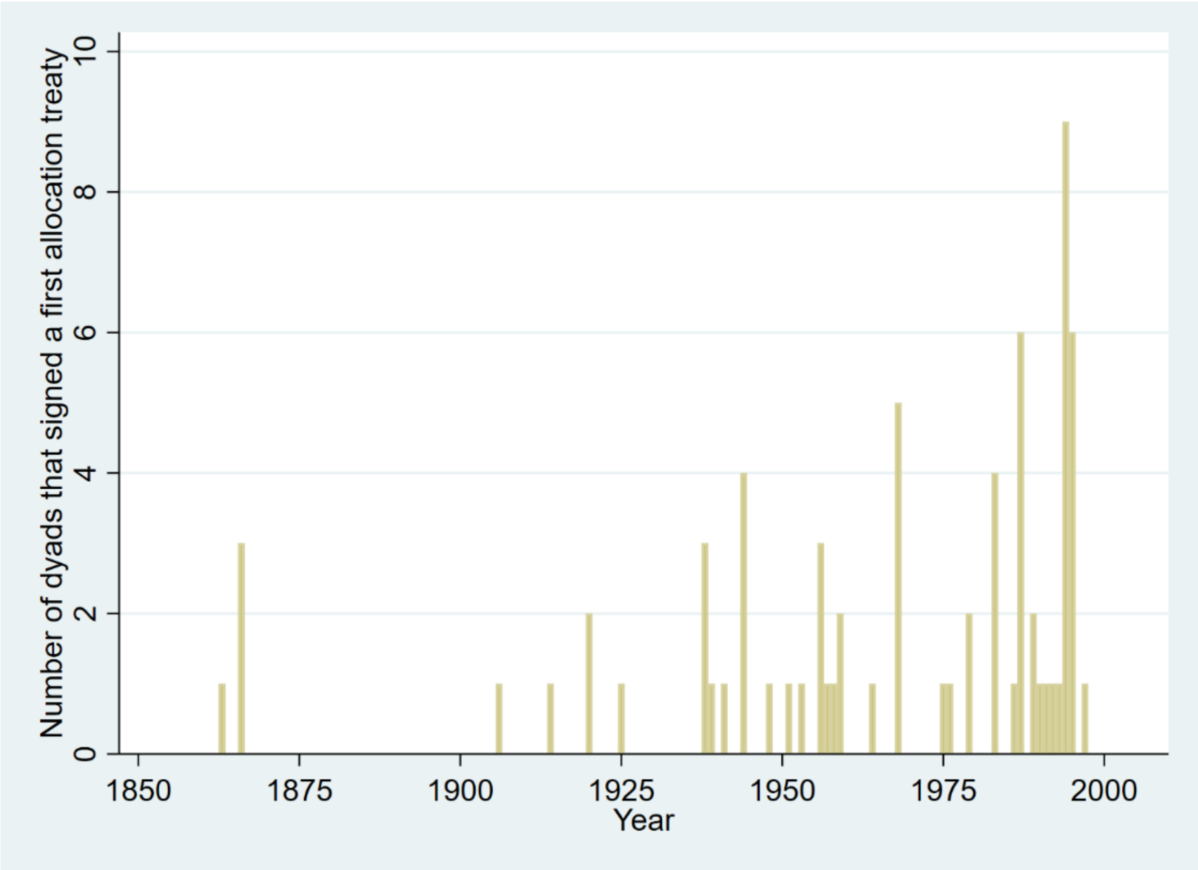
Figure 4.1 presents the number of river dyad units that signed a document on the allocation of water for each year. In the graph, the same documents can be accounted for multiple times if they specify the allocation of water between more than two countries that share the same river. Only the first document signed by a country dyad over a specific river is shown. Subsequent documents to this first document are not shown in the figure as they will not be considered in the empirical analysis.

The increase in dyads that signed treaties since 1970 reflects the increase in treaties but also the fact that many countries only gained their independence in the second half of the 20th century. The first treaty on the allocation of water was signed between Belgium and the Netherlands in

[16] Treaties providing a general framework for water-related issues like the 'International Convention Concerning the Regime of Navigable Waterways of International Concern', the 'African Convention on the Conservation of Nature and Natural Resources' or the 'Convention on Wetlands of International Importance especially as Waterfowl Habitat', as well as treaties that had no date or specific water content indicated in the data were not included in the analysis.

1863, followed by an allocation treaty signed in France and Spain (1866).[17] After these initial treaties, no documents were signed until the treaty between the USA and Mexico over the Rio Grande in 1906. In this treaty, Mexico waived any claims over the water of the Rio Grande. In return, the United States guaranteed to construct a dam that evens out the flow of the river and supplied a fixed quantity of river water to Mexico (Dinar et al. 2007, 201f.).

Figure 4.1 First documents signed over the allocation of water in a dyad



Since no allocation treaty was signed outside of Europe until 1906, I assume that signing allocation treaties was only an option for countries worldwide afterwards and hence I start the analysis in 1907. Those dyads that were formed after 1907, for example because of shifts in their territory or because they gained their independence, enter in the year the dyad was formed. Of the 796 basin-dyad combinations in the sample, 66 signed a treaty that addresses the allocation of water by the year 2007.[18] A river-dyad unit drops out from the analysis once the first document that specified the allocation of water is signed.

[17] These are the 'Treaty for the Regulation of Water Withdrawal from the Meuse' and the 'Treaty of Delimitation between France and Spain, signed at Bayonne', which regulates the allocation of water in the Bidasoa, the Ebra and the Garonne.

[18] Some treaties were joined by countries that were not located at the river. This happened mostly when colonial powers joined a treaty governing a river located in a colony. Since these countries are no riparian countries, treaties signed by them are not taken into account.

4.3.2 Explanatory variables

Table 4.1 provides an overview and a description of the explanatory variables used in the regressions. Section 4.7.1 in the Appendix offers a detailed discussion on the explanatory variables. Table 4.6 in the same section in the Appendix provides the summary statistics of the explanatory variables. The variables that test the role of power on the formation of treaties are total GDP, military expenditures and wealth measured as the GDP per capita. In order to set the power of the downstream state in relation to that of the upstream state, the value of the downstream state's power measure is divided by the upstream state's value. Subsequently, the logarithm of the ratio is taken. Since states are the central actors in the negotiations, power is determined on national level.

In the regressions, I also control for the capacity of the dams build in the ten years before and after the signature of treaties in the observed year. This time period was chosen because treaties might be signed in the anticipation or aftermath of a dam construction. The value of the variable is zero when no dam was built in the period. The location of the dams and the basin areas were spatially matched using google earth engine.

Another control variable is the available renewable water per capita in the country's basin area. The data was taken from the Transboundary Water Assessment Programme (United Nations Environment Programme 2015), which offers only the average over the period from 1971 to 2000. Thus, the values do not vary between years and present long-term averages. Values were also extended to the years before and after this period. Thus, the data does not allow to analyze the effect of single events that impacted the amount of water available, but it is likely that water diplomacy is driven not by single events but by long-term conditions (Dinar, Dinar, and Kurukulasuriya 2011). Previous studies examined the effect of water scarcity but mostly relied on national averages. It has also not been tested whether water scarcity in the upstream and the downstream basins have a different effect on cooperation. It seems intuitive that water scarce upstream states are less willing to sign an allocation treaty than water scarce downstream states.

I use two main setups of regression models. The first model controls for the relation of the countries in the dyad. The first of the factors included is being in a formal alliance. The second variable is a dummy that controls whether the countries in the dyad share a common border, because the allocation of water from upstream to downstream riparian states is most relevant if countries are direct neighbors with no country between them. Finally, I control for having the same regime. The two types of regime examined are democracies and autocracies.

Table 4.1 Explanatory variables

Name	Description	Source
GDP	Natural logarithm of the ratio of the downstream state's GDP divided by the upstream state's GDP.	Maddison Project database (Maddison 2010)
MilExp	Natural logarithm of the ratio of the downstream state's military expenditures divided by the upstream state's military expenditures.	National Material Capabilities database (version 5.0) (Singer, Bremer, and Stuckey 1972; Greig and Enterline 2017)
GDPpc	Natural logarithm of the ratio of the downstream state's GDP per capita divided by the upstream state's GDP per capita.	GDP: Maddison Project database (Maddison 2010); Population: National Material Capabilities database (version 5.0) (Singer, Bremer, and Stuckey 1972; Greig and Enterline 2017)
HydroDamUp *HydroDamDown*	Capacity of hydropower dams constructed in the ten proceeding and the ten preceding years to the observed year in the upstream and downstream state's basin area (in 10^{12} cubic meters).	GRanD dataset (Lehner et al. 2011, Beames, Lehner, and Anand 2019); Google earth engine was used for spatial matching
IrrigationDamUp *IrrigationDamDown*	Capacity of irrigation and water supply dams constructed in the ten proceeding and the ten preceding years to the observed year in the upstream and downstream state's basin area (in 10^{12} cubic meters).	GRanD dataset (Lehner et al. 2011, Beames, Lehner, and Anand 2019); Google earth engine was used for spatial matching
RenWaterSupplyUp *RenWaterSupplyDown*	Mean annual per capita discharge internally generated from 1971 to 2000 in the upstream and downstream state's basin area (in million m^3/person/year).	Transboundary Water Assessment Programme (United Nations Environment Programme 2015)
Allies	The dummy variable that takes the value 1 if an active defense, neutrality, nonaggression or entente agreement exists between the states in the dyad.	Gibler (2009)
Adjacent	The dummy variable that takes the value 1 if the states in the dyad share a common border.	Own codification
Common regime	The dummy variable that takes the value 1 if the two countries share a common regime (democratic or autocratic).	Polity IV Project (Marshall, Gurr, and Jaggers 2018)
ExpoDroughtUp *ExpoDroughtDown*	Mean variation in monthly river discharge from 1971 to 2000 in the upstream and downstream state's basin area.	Transboundary Water Assessment Programme (United Nations Environment Programme 2015)
ExpoFloodUp *ExpoFloodDown*	Economic exposure (in dollars) to flood hazards relative to the basin size from 1971 to 2000 in the upstream and downstream state's basin area.	Transboundary Water Assessment Programme (United Nations Environment Programme 2015)
Trade	Volume of trade between the two countries in the dyad divided by their combined GDP.	Trade: Correlates of War (Barbieri, Keshk, and Pollins 2009; Barbieri and Keshk 2016); GDP: Maddison Project database (Maddison 2010)

In a second regression specification, I test whether exposure to droughts or floods in the upstream state's basin area or the downstream state's basin area has an impact on the formation of treaties. These two variables represent the average exposure from 1971 to 2000 and were extended to preceding and proceeding years. Exposure to droughts is measured as the annual discharge variability. The exposure to flood measures the economic exposure to flood. Again I assume that exposure of the upstream basin makes the signature of treaties less likely, while exposure of the downstream basin is conducive. An additional set of regressions includes the common trade of the dyad as a control variable. The data on trade was taken from the Correlates of War project (Barbieri, Keshk, and Pollins 2009; Barbieri and Keshk 2016). The variable *Trade* was created by dividing the volume in trade in the dyad by its combined GDP.

4.4 Methodology

Following Tir and Ackerman (2009) and Bobekova (2015), I use a Cox regression (Cox 1972, Jenkins 2005) in order to estimate the hazard rate of the signature of a treaty over the allocation of the water of an international river basin. The 'failure', and thus the drop out from the analysis, is defined as the signature of a document that regulates the allocation of water between the two riparian states. In contrast to logit/probit models, the Cox model allows to examine the time-dependent factors that lead to the formation of treaties. Logit/probit panel models estimating this binary outcome suffer from serial correlation in the occurrence of a treaty. Moreover, the Cox model allows for right-censoring of the data.

Similar to Ovodenko (2016), one observation in the analysis is formed by a country dyad and an international river basin that the two countries share in a given year. With this proceeding the setup deviates from Tir and Ackerman (2009) and Bobekova (2015), who examine entire country dyads and do not differentiate between the different rivers of the dyad. The more detailed basin approach allows to account for the interests resulting from a state's position along the river.

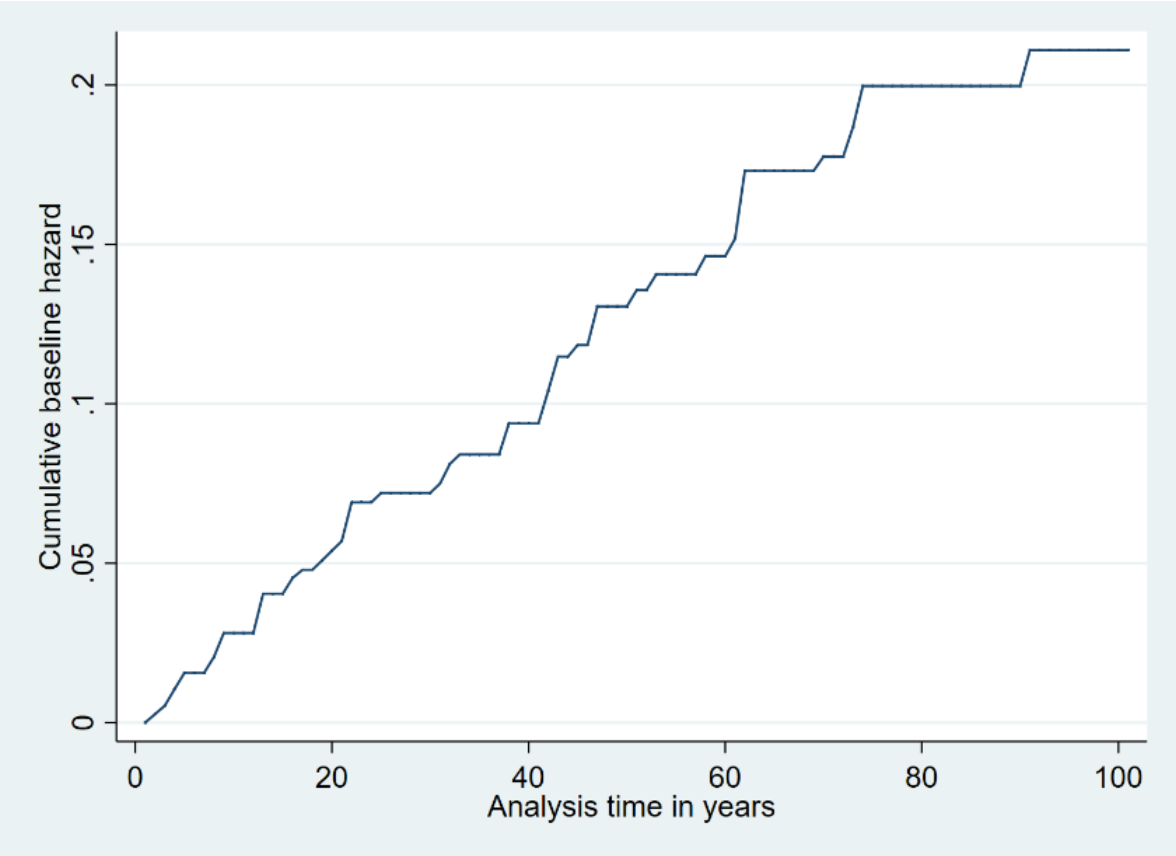
4.5 Results

Table 4.2 presents the regression results of the Cox regressions. The coefficients reported in the table indicate the change in the log hazard ratio as response to an increase of one unit in the independent variable, holding all other covariates constant. The three controls for power are not included together in the analysis since they are strongly correlated with each other. The first two columns test the effect of economic power between the downstream and the upstream state, the models (3) and (4) test the effect of military power and the last two columns test the effect of wealth. The columns (1), (3) and (5) include variables that control for the relation in the dyad, while the regressions models (2), (4) and (6) control for the exposure to drought and flood of the rivers' basins.

Figure 4.2 shows the estimated baseline hazard from the first regression model. The graph displays the estimated cumulated share of dyads that signed an allocation treaty by the number of years it exists. It demonstrates that the overall number of treaties signed is relatively stable over time. The estimated likelihood to sign a treaty, or 'hazard' in the language of the model, is only somewhat lower after 80 years. The estimated likelihood to have an allocation treaty after 100 years sharing a river is a little above 20%. Using Schoenfeld residuals confirms that the proportional-hazards assumption holds for all regression models.

The regression models (1) and (2) demonstrate that the ratio capturing the downstream state's relative economic power, *GDP*, is insignificant. Similarly, the ratio of military spending, *MilExp*, is insignificant in both models. The ratio of GDP per capita, *GDPpc*, remains also insignificant in both model setups. Hence, the results provide no evidence that treaties become more likely when the power of the downstream state increases.

Figure 4.2 Cumulative hazard of signing an allocation treaty



The capacity of hydropower dams constructed in the upstream basin around the observed year has a significant impact on the formation of treaties in all six model specifications. Treaties become more likely before and after these dams are built and increase with the dam's capacity. By contrast, hydropower dams built in the downstream region have no impact on the emergence of treaties. This finding reflects that the negative externalities of dams affect mostly the downstream areas. There is no need for states to sign treaties when downstream states construct dams. No correlation can be found in any regression for the capacity of irrigation and water supply dams constructed in the upstream country and the occurrence of treaties.

Table 4.2 Cox regressions

	(1)	(2)	(3)	(4)	(5)	(6)
GDP	0.0326	0.0475				
	(0.0819)	(0.0738)				
MilExp			0.0536	0.0613		
			(0.0598)	(0.0601)		
GDPpc					-0.276	-0.235
					(0.234)	(0.197)
HydroDamUp	24.75***	26.08***	26.14***	26.77***	24.18***	25.10***
	(4.864)	(4.348)	(4.297)	(3.935)	(3.929)	(3.479)
HydroDamDown	-3.205	-2.407	-5.135	-4.676	-2.127	-0.851
	(6.315)	(6.245)	(6.057)	(5.817)	(6.228)	(5.907)
IrrigationDamUp	27.51	9.490	26.68	11.26	22.47	2.916
	(16.99)	(18.74)	(17.72)	(18.60)	(15.91)	(17.82)
IrrigationDamDown	-3.067	-3.194	-3.804	-3.843	-1.547	-1.601
	(5.161)	(5.456)	(5.120)	(5.309)	(5.341)	(5.524)
RenWaterSupplyUp	-0.0400	-0.0102	-0.00198	0.0112	-0.0373	-0.0139
	(0.193)	(0.0708)	(0.0696)	(0.0561)	(0.144)	(0.0662)
RenWaterSupplyDown	-6.860**	-4.770*	-7.504**	-5.792*	-6.728**	-4.719*
	(3.025)	(2.547)	(3.256)	(3.006)	(2.947)	(2.580)
Allies (d)	-0.108		-0.104		-0.0793	
	(0.308)		(0.292)		(0.314)	
Adjacent (d)	0.531*		0.393		0.538**	
	(0.278)		(0.267)		(0.272)	
Common regime (d)	0.0275		0.00958		0.0579	
	(0.292)		(0.275)		(0.291)	
ExpoDroughtUp		0.365**		0.331*		0.328*
		(0.181)		(0.187)		(0.177)
ExpoDroughtDown		0.665***		0.612**		0.682***
		(0.244)		(0.246)		(0.248)
ExpoFloodUp		1.490		1.392		2.029**
		(1.051)		(1.030)		(1.024)
ExpoFloodDown		2.489***		2.491***		2.354***
		(0.692)		(0.675)		(0.704)
Observations	24307	24486	26959	27267	24307	24486
Number of dyads	558	550	592	584	558	550
Number of treaties	58	60	62	64	58	60

Cox Proportional Hazard Model on the hazard of a country dyad to sign an allocation treaty. Coefficients are displayed. The robust standard errors are presented in the parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The results also reveal that it is important to control separately for the available water in the downstream and the upstream state's basin. The available water per person in the basin of the upstream state, *RenWaterSupplyUp*, remains insignificant in all regressions. It is remarkable that the regression results provide no evidence that water scarce upstream countries are less willing to agree to the signature of an allocation treaty. *RenWaterSupplyDown*, in contrast, has a significantly negative effect on the likelihood of signing a treaty in all regression models at least at a 10% level. This indicates that treaties are more common in situations where the downstream state suffers from water scarcity.

Countries that are in an active alliance agreement are not more likely to sign an allocation treaty than other countries, as indicated by the variable *Allies*. The coefficient of the control variable *Adjacent*, which controls whether the two states involved share a common border, is at least weakly significant in two specifications. The coefficients are quite high. Adjacent countries are on average around 70%[19] more likely to sign treaties than non-adjacent countries if the coefficients report the true values. Having a common political system has no impact on the formation of allocation treaties.

Turning to the regressions that test the exposure of the states to drought and flood reveals that *ExpoDroughtUp* is at least weakly significantly positive in all three models that include the variable. A high variation in river runoff in the downstream state's basin also increases the likelihood for a treaty as indicated by the coefficient of the variable *ExpoDroughtDown*. However, the variation in the downstream state's basin has a higher coefficient and a more robust effect. Thus, allocation treaties become more likely when the upstream state and in particular the downstream state face high insecurity in their water supply because of a high variation in runoff. The level of economic exposure to floods in the upstream basin, *ExpoFloodUp*, is not significantly different from zero. The coefficients of *ExpoFloodDown*, in contrast, have a positive impact on the formation of treaties.

As it seems that powerful downstream states do not coerce upstream states into cooperation, it is interesting to check whether issue-linkage and side payments play a role in the formation of the treaties. The International Freshwater Treaties Database (Giordano et al. 2014; Transboundary Freshwater Dispute Database 2018) provides the information whether the treaties use issue-linkage or compensations. Examining the share of allocation treaties in the dataset that uses these mechanisms reveals that 12 (18%) treaties include some form of compensation and 8 (12%) link the treaty to non-water issues.[20] This demonstrates that only a minority of the treaties in the dataset use these mechanisms as an incentive for the formation of allocation treaties. This result is in line with previous studies that find that side payments are not common in water allocation treaties (Dinar 2006).

[19] $e^{0.538}-1=0.71$ and $e^{0.531}-1=0.70$

[20] For example, in the 1951 'Agreement between Finland and Norway on the Transfer from the Course of the Näätämo (Neiden) River to the course of the Gandvik River of Water from the Garsjoen, Kjerringvatn and Forstevannene Lakes', Norway pays for Finish losses in power generation because of water allocation.

In order to test the robustness of the results, Table 4.3 considers dyads to be at risk to sign an allocation treaty only starting from the year 1950. This starting point was also chosen by Tir and Ackerman (2009). It allows to add the economic interdependence, measured as the value of trade in the dyad divided by the combined GDP, as a control variable. Data on international trade before 1950 is too scarce to include it as a control variable. Common trade has been shown to improve cooperation concerning international environmental problems (Sigman 2003). Moreover, trade is strongly correlated to the signature of river treaties (Tir and Ackerman 2009). Dyads that already signed an allocation treaty at that time were excluded from the analysis.

The regression models (1) and (2) confirm the finding that the downstream state's GDP relative to that of the upstream state has no impact on the formation of treaties. The relative military expenditure of the downstream state, in contrast, enters on a 5% and a 10% significance level. The more the downstream state spends on its military relative to the upstream state the more likely treaties become. The interpretation of the coefficient is not straightforward. An increase by one unit of *MilExp* increases the estimated hazard by 16%[21] if the effect is indeed linear as assumed in the model. In the case where both states spend the same amount on their military an increase by one unit of *MilExp* would mean that the downstream state increases its spending by 172%.[22] Military power seems to play a role, but its effect is only noticeable when the downstream state's capacity exceeds that of the upstream state to a large extent.

GDP per capita is significantly correlated with the signature of treaties in both models that include the variable. The negative sign of the coefficients indicates that relative high wealth in the upstream country compared to the downstream country is associated with a higher likelihood for an allocation treaty. So, while treaties are more likely at the presence of a militarily powerful downstream state, treaties occur also more frequently at the presence of a relatively poor downstream state. It is possible that richer states feel a stronger obligation to cooperate with poorer states.

The economic interdependence between the two countries, *Trade*, is not significant. This result is surprising but shows that omitting the variable might not alter the results presented in the first set of regressions. The capacity of hydropower dams in the upstream basin area is highly significant. Hydropower dams constructed in the downstream basin are insignificant as well as irrigation and water supply dams in the upstream or downstream area. The upstream state's water supply remains insignificant. Also the downstream country's water supply, which was still significantly negative in the regressions presented in Table 4.2, is insignificant.

---

[21] $(e^{0.148}-1)*100=16\%$

[22] The ratio of military spending is equal to 1 if both states spend the same amount on their military. The natural logarithm of this ratio is 0. In order to increase the logarithm to 1 the downstream state's spending needs to increase by the factor e → $(e-1)*100=172\%$

Table 4.3 Cox regressions starting in 1950

	(1)	(2)	(3)	(4)	(5)	(6)
GDP	0.0552	0.0750				
	(0.0975)	(0.0882)				
MilExp			0.134^{*}	0.148^{**}		
			(0.0769)	(0.0711)		
GDPpc					-0.568^{**}	-0.492^{**}
					(0.264)	(0.222)
Trade	-34.48	-12.62	-32.34	-11.62	-37.84	-14.19
	(59.61)	(44.23)	(58.79)	(43.69)	(62.01)	(46.05)
HydroDamUp	22.80^{***}	23.94^{***}	25.29^{***}	26.05^{***}	21.70^{***}	23.12^{***}
	(6.062)	(4.633)	(5.351)	(4.347)	(4.625)	(3.605)
HydroDamDown	-3.995	-0.887	-5.021	-1.784	-3.363	0.186
	(7.248)	(7.167)	(7.581)	(7.484)	(7.699)	(6.953)
IrrigationDamUp	-0.569	-5.201	-0.512	-4.788	-7.974	-17.55
	(17.63)	(17.28)	(18.16)	(17.60)	(18.24)	(18.34)
IrrigationDamDown	-4.817	-3.132	-7.917	-6.310	-2.175	-0.884
	(6.535)	(6.757)	(6.655)	(6.885)	(6.844)	(7.073)
RenWaterSupplyUp	-8.365	-5.347	-8.397	-5.500	-8.270	-5.030
	(11.74)	(8.608)	(11.87)	(8.921)	(11.24)	(7.712)
RenWaterSupplyDown	-7.723	-6.362	-7.414	-6.218	-6.833	-5.684
	(5.812)	(5.186)	(5.605)	(4.907)	(5.410)	(5.153)
Allies (d)	-0.388		-0.345		-0.345	
	(0.420)		(0.422)		(0.410)	
Adjacent (d)	0.659		0.599		0.792	
	(0.505)		(0.489)		(0.497)	
Common regime (d)	-0.0698		-0.108		-0.00431	
	(0.392)		(0.391)		(0.389)	
ExpoDroughtUp		0.250		0.289		0.161
		(0.209)		(0.215)		(0.201)
ExpoDroughtDown		0.382		0.350		0.458^{*}
		(0.275)		(0.276)		(0.277)
ExpoFloodUp		1.371		0.981		2.569^{**}
		(1.188)		(1.109)		(1.130)
ExpoFloodDown		3.201^{***}		3.352^{***}		2.886^{***}
		(0.797)		(0.759)		(0.821)
Observations	19290	19459	19105	19274	19290	19459
Number of dyads	507	498	507	498	507	498
Number of treaties	36	38	36	38	36	38

Cox Proportional Hazard Model on the hazard of a country dyad to sign an allocation treaty. Coefficients are displayed. The robust standard errors are presented in the parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Neither being formal allies, nor sharing a common border, nor having the same political system has an impact on the formation of treaties. Also, the exposure to droughts in the upstream and the downstream region is a factor that drives the formation of treaties. The coefficient of *ExpoFloodUp* is only significantly different from zero in the sixth regression model. The economic exposure to floods in the downstream state's basin has a more robust positive effect on the formation of treaties. Thus, the results are sensitive to the starting point of the analysis and demonstrate how it alters the results. While power was irrelevant and water scarcity and exposure to droughts in the downstream basin had a decisive effect when starting the analysis in 1907, this changes when starting the analysis in 1950 and additionally controlling for common trade.

An additional robustness check is provided in Table 4.7 in the Appendix. Dyads are only included if they share a common border. The results show that power is also irrelevant between countries that share a common border. The driving factor for the formation of treaties between adjacent states is the construction of hydropower and irrigation and water supply dams in the upstream state's basin area and water scarcity in the downstream state. Adjacent states are also found to sign treaties more often when they share a common political system.

4.5.1 Water scarcity

The UN considers an area as experiencing water stress when available renewable water per person per year falls below 1,700 m^3. Areas with less than 1,000 m^3 available renewable water are considered water scarce (United Nations 2012). In the following, I test whether the findings also hold true for water stressed and scarce regions where competition over water is higher and conflicts over the allocation of water are more common. Only basins in which the renewable water per person and year is below 1700 m^3 are included in the analysis. Table 4.4 shows the results of Cox regressions when only water scarce basins are included in the analysis. This reduces the number of dyads in the sample by almost 80% and the number of treaties to ten to twelve, depending on the explanatory variables that are used. Almost no dams are constructed in the sample, therefore they are not included in the analysis.

The GDP of the downstream state relative to that of the upstream state has no impact on the formation of allocation treaties. The downstream state's relative military spending, in contrast, has a significant impact. Treaties in water scarce regions are more common when the downstream state is relatively powerful in military terms. The coefficient is much higher compared to the models presented in Table 4.3 An increase by one unit increases the hazard by 30%.[23] So while military appears to be more important in water scarce regions, it still only has a noticeable effect if the downstream state's capacity exceeds that of the upstream state multiple times. GDP per capita remains insignificant.

The level of water scarcity has no effect on the formation of treaties. Neither very high levels of scarcity in the upstream basin nor in the downstream basin play a role. Also, the other control variables are insignificant, except for the exposure to drought in the downstream state's basin.

[23] $(e^{0.262}-1)*100=30\%$

Treaties are more likely when downstream states face higher variability in discharge and, thus, higher insecurity with regard to droughts.

Table 4.4 Cox regressions: Water scarce basins only

	(1)	(2)	(3)	(4)	(5)	(6)
GDP	0.192	0.204				
	(0.170)	(0.171)				
MilExp			0.262^{**}	0.323^{**}		
			(0.115)	(0.139)		
GDPpc					0.414	0.182
					(0.578)	(0.572)
RenWaterSupplyUp	-25.33	-17.06	-26.36	-24.63^{**}	-25.19	-16.37
	(22.51)	(11.35)	(21.82)	(11.82)	(21.63)	(13.88)
RenWaterSupplyDown	-19.21	-28.37	-24.20	-32.74	-18.31	-32.60
	(20.41)	(26.04)	(19.62)	(21.10)	(22.23)	(37.21)
Allies (d)	0.301		0.287		0.179	
	(0.647)		(0.547)		(0.579)	
Adjacent (d)	-0.266		0.266		-0.249	
	(1.271)		(0.831)		(1.272)	
Common regime (d)	-0.328		-0.494		-0.251	
	(0.897)		(0.986)		(0.858)	
ExpoDroughtUp		-0.523		-0.743		-0.535
		(0.458)		(0.470)		(0.511)
ExpoDroughtDown		0.901^{**}		0.948^{**}		0.913^{*}
		(0.424)		(0.395)		(0.550)
ExpoFloodUp		-112.2		-174.8		-87.53
		(104.3)		(111.9)		(98.73)
ExpoFloodDown		-74.27		-75.70		-53.52
		(54.85)		(47.67)		(64.63)
Observations	5027	5058	5468	5517	5027	5058
Number of dyads	121	119	126	124	121	119
Number of treaties	10	11	11	12	10	11

Cox Proportional Hazard Model on the hazard of a country dyad to sign an allocation treaty. Coefficients are displayed. Only water scarce basins are included in the analysis. The robust standard errors are presented in the parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Compensation is applied in only one of the 12 allocation treaties (8.3%) and linkage to non-water issues in four of them (33.3%). This serves as an indication that linkage might be applied more often in water scarce regions to establish cooperation. However, these results are based only on a small number of cases. The proportion is also too low to explain the emergence of treaties in these regions.

4.5.2 Power

So far, the results suggest that relative economic power has no effect on the likelihood to sign allocation treaties. When considering all countries there is also only weak evidence that military power is relevant. There are two possible explanations for the finding that the downstream state's power is not the driving factor in the emergence of allocation treaties in most cases. The first explanation is that states cannot utilize their economic or military power to coerce other states into cooperation. Hegemons might be more vulnerable than sheer numbers reveal (Petersen-Perlman and Fischhendler 2018). Economic and military power might also not be able to offset the advantages that a country holds due to its geographical location. A second explanation could be that it is less relevant who has the power but how strongly states differ with regard to their power. The theory of hegemonic stability argues that hegemons are necessary to create stable cooperation because hegemons are able to offer the public good of stability (Keohane 1984). The concept has also been applied to international rivers (Daoudy 2009; Warner and Zawahri 2012). Its effect on treaty formation has also been tested in a number of empirical studies (Espey and Towfique 2004; Tir and Ackerman 2009; Dinar, Dinar, and Kurukulasuriya 2011).

In order to further investigate the role of power with regard to cooperation over the allocation of water in international rivers, Table 4.5 provides regressions that test whether power asymmetries in the dyad have an impact on the formation of treaties. The variables *GDPAsymmetry*, *MilExpAsymmetry* and *GDPpcAsymmetry* were created by dividing the stronger state's value by the weaker state's value before taking the logarithm. This approach allows to test whether the presence of a hegemon facilitates cooperation.

The regression models (1) and (2) in Table 4.5 show that treaties are more likely to occur when dyad states are asymmetric with regard to their economy. The same is true for differences in military power and wealth. The positive sign of the coefficients shows that the presence of a hegemon enables cooperation independent of its location. Thus, despite important differences in the setup of the analysis the results support the findings of previous studies. Treaties occur more likely in the presence of high asymmetries in military capacity (Tir and Ackerman 2009) and economic power (Song and Whittington 2004).

The results for the other variables are almost identical to the results in Table 4.2. The regressions confirm that the strongest drivers for the formation of treaties are the construction of hydropower dams in the upstream basin, the upstream state's water scarcity and the downstream state's exposure to droughts and floods.

Table 4.5 Cox regressions: Power asymmetry

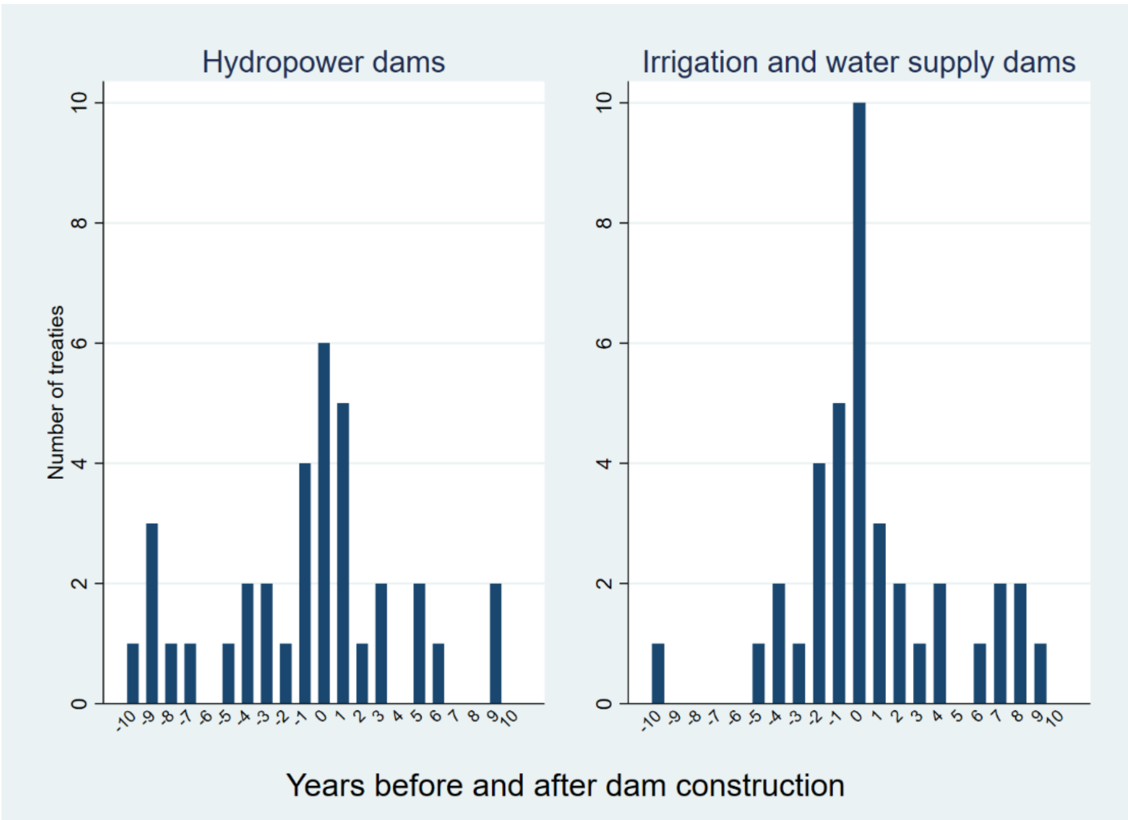
	(1)	(2)	(3)	(4)	(5)	(6)
GDPAsymmetry	0.243**	0.178**				
	(0.0999)	(0.0849)				
MilExpAsymmetry			0.180**	0.173**		
			(0.0752)	(0.0728)		
GDPpcAsymmetry					0.729***	0.469*
					(0.261)	(0.262)
HydroDamUp	21.33***	23.44***	24.26***	24.96***	24.12***	25.46***
	(4.495)	(3.686)	(4.305)	(3.651)	(4.060)	(3.646)
HydroDamDown	-5.375	-3.598	-6.354	-5.669	-2.497	-1.408
	(7.263)	(6.449)	(6.360)	(5.806)	(5.948)	(5.735)
IrrigationDamUp	24.68	4.646	27.49*	8.908	25.45*	5.493
	(15.29)	(18.21)	(15.43)	(17.72)	(14.50)	(17.93)
IrrigationDamDown	-2.775	-2.737	-4.022	-3.966	-2.570	-2.484
	(5.061)	(5.271)	(4.907)	(5.001)	(5.185)	(5.379)
RenWaterSupplyUp	-0.0259	-0.00128	0.0106	0.0224	-0.0305	-0.0116
	(0.165)	(0.0668)	(0.0610)	(0.0570)	(0.102)	(0.0593)
RenWaterSupplyDown	-7.604**	-5.086*	-8.276**	-6.362*	-6.546**	-4.420*
	(3.408)	(2.820)	(3.643)	(3.457)	(2.977)	(2.377)
Allies (d)	0.00655		-0.0531		-0.0167	
	(0.310)		(0.284)		(0.307)	
Adjacent (d)	0.590**		0.384		0.678**	
	(0.270)		(0.259)		(0.280)	
Common regime (d)	0.00447		-0.0445		0.0849	
	(0.290)		(0.284)		(0.290)	
ExpoDroughtUp		0.356**		0.311*		0.302*
		(0.178)		(0.181)		(0.183)
ExpoDroughtDown		0.638***		0.669***		0.694***
		(0.247)		(0.247)		(0.251)
ExpoFloodUp		1.818*		1.896*		1.770*
		(1.071)		(1.082)		(1.006)
ExpoFloodDown		2.225***		2.266***		2.423***
		(0.754)		(0.756)		(0.712)
Observations	24307	24486	26959	27267	24307	24486
Number of dyads	558	550	592	584	558	550
Number of treaties	58	60	62	64	58	60

Cox Proportional Hazard Model on the hazard of a country dyad to sign an allocation treaty. Coefficients are displayed. The robust standard errors are presented in the parentheses. $^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

4.5.3 Dams

In the following I further investigate the role of dams, since their effect has not been assessed in previous studies. The results from the regressions show that in particular hydropower dams are an important driver for the formation of treaties. In order to get a better understanding of how the timing of dams and treaties coincides, Figure 4.3 presents the number of treaties signed before and after the construction of dams in the upstream country of the dyad. The x-axis indicates the years before and after a dam construction, with ”0” representing the year of the dam construction. The negative numbers display the years before the construction of a dam and the positive numbers the years after the construction. The y-axis displays the treaties signed over the allocation of water in the same country basin. In total, the dataset includes 1475 hydropower dams and 1603 irrigation and water supply dams. The left panel in the graph demonstrates that treaties are more frequently signed in the year of the construction of hydropower dams in the upstream country than in any other year. Treaties also occur more frequently in the immediate years before and after dam constructions. Although irrigation and water supply dams are not found to increase the likelihood of treaties when tested in the regressions (except when only including adjacent states as shown in Table 4.7 in the Appendix), the pattern is also very clear for this type of dam. The right panel shows that the construction of irrigation and water supply dams and the signature of allocation treaties strongly coincide. By far most treaties are signed in the years when these dams are constructed.

Figure 4.3 Treaties and the construction of dams



This strong correlation between dams and treaties can be explained by the insecurity and externalities that dams in the upstream region create for downstream states. While dams can provide particularly inhabitants of urban areas with an improved flood control, dams are also associated with negative consequences for the well-being of the downstream population. They alter the seasonal water discharge and pose a threat to agriculture, livestock and fishing (Richter et al. 2010). Dams constructed for irrigation and urban water consumptions reduce the quantity of the river's water. But also dams designed for hydropower consume significant amounts of water due to evaporation (Mekonnen and Hoekstra 2012). Upstream states often ignore the costs resulting from the construction of dams for downstream states (Olmstead and Sigman 2015). Therefore, the construction of dams is a common reason for concerns of downstream states and conflicts between riparian states. Treaties are an attempt to limit the externalities for the downstream basin area.

4.6 Conclusion

The results from the analysis suggest that the power of the downstream state, which is the main beneficiary of allocation treaties, is only a minor factor in the formation of treaties. The economic power of downstream states relative to upstream states does not increase the likelihood of allocation treaties. The relative military power of the downstream state was found to have a small effect when choosing the year 1950 as a starting point of the analysis and no effect when starting the analysis in 1907. This result confirms the study of Dinar (2009) who reviews several agreements and argues that cooperation is the norm even in the presence of an upstream hegemon. The results also provide some evidence that relatively rich upstream states are more willing to sign allocation treaties. However, the results indicate that in water scarce basins where competition is higher, allocation treaties occur more frequently when the downstream state is militarily more powerful than the upstream state. More relevant than the downstream state's power relative to the upstream state's power is how much the states in the dyad differ in their power regardless of their location. High differences in economic power, in military power and in wealth between two countries sharing a river increase the likelihood of an allocation treaty. The more states differ in their economic and military power, the more likely allocation treaties become. This result supports the idea that hegemons might provide the stability necessary to establish cooperative regimes in international river basins. A possible solution to provide the necessary stability in basins with no clear hegemon might be the involvement of third parties, such as the UN or the EU (Bobekova 2015).

The findings also stress the importance to differentiate between the upstream and the downstream basin when examining the role of water scarcity, exposure to floods and droughts, and dam constructions, which has been neglected in previous empirical studies. The downstream state's dependency on the river is shown to be an important driver for cooperation. Downstream states that suffer from water scarcity and exposure to floods and droughts sign allocation treaties more frequently. Interestingly, I find no evidence that water scarcity or exposure to droughts and floods in the upstream basin make treaties less likely. The results also stress the importance of dam constructions in the upstream state's basin. Especially hydropower dams constructed in the upstream state's basin area are a common cause for the signature of treaties. Upstream states seem to be willing to provide downstream states with securities in the form of allocation treaties to address possible concerns arising from dam constructions.

To sum up, upstream states are willing to engage in cooperation with downstream states that face water scarcity and insecurities from droughts, floods and dam constructions. In most basins the downstream state's power is not relevant. This is an encouraging finding: Despite the lack of supranational organizations that set rules between states, cooperation emerges even when it is not induced by force. There are two possible explanations for this. The first explanation is that upstream states acknowledge the downstream state's right to the river's water to some degree. Most downstream areas are more densely populated than the upstream areas and traditionally use a larger proportion of the river's water. Although no universal international law exists that prevents upstream states from solely exploiting the river's resources, upstream states might acknowledge norms that recognize the downstream areas' need for water. The low proportion of treaties that use issue-linkage and compensations provides some support for this hypothesis. The second explanation is that downstream states are not powerless even if economically and militarily inferior. If their survival depends on a river, also weak states can be willing to risk costly conflicts, which is a credible threat also to more powerful states (Petersen-Perlman and Fischhendler 2018). Overall, the results prove that international cooperation can emerge even if the benefits of this cooperation are highly unequally distributed among countries. Only in water scarce basins, military becomes important. However, the results are based on a comparatively low number of cases. Case studies might provide a better understanding of the dynamics behind the signature of allocation treaties in these basins.

A major limitation of the model used in the analysis is that it allows only a binary coding of cooperation in the case that an allocation treaty is signed. It is difficult to control for differences in the quality of treaties. It is also important to remember that the signature of treaties is only the first step towards sustainable cooperation. An important question for future research is to what degree upstream states adapt their behavior as a result of allocation treaties and whether the water supply of downstream states actually improves once treaties are signed.

Acknowledgements: The work was financially supported by the European Union (EU) Horizon 2020 program, action ERC-2014-STG, Project HUCO, grant number 636746.

4.7 Appendix

4.7.1 Explanatory variables

In the following, I describe the creation of the explanatory variables. Data on GDP was taken from the Maddison Project database (Maddison 2010). The variable *GDP* takes the natural logarithm of the ratio of the downstream state's GDP divided by the upstream state's GDP. The second variable that is used to analyze the relative power in the dyad is the military spending, which might be a driver for conflict and cooperation in international river basins (Frey 1993). The variable was derived from the National Material Capabilities database (version 5.0) as part of the Correlates of War project (Singer, Bremer, and Stuckey 1972; Greig and Enterline 2017). MilExp is the natural logarithm of the ratio of military spending where again the downstream state is in the numerator and the upstream state is in the denominator. Thirdly, GDP per capita is used. The data on the population size, needed to calculate the value, was also taken from the Maddison Project (Maddison 2010). *GDPpc* takes the natural logarithm of the ratio of the GDP per capita. The data on the population size was taken from the National Material Capabilities database (Singer, Bremer, and Stuckey 1972; Greig and Enterline 2017). Missing values on GDP and military spending were linearly interpolated (around 5% of observations on GDP and 4.4% on military spending).

Table 4.6 Summary statistics of variables

Variable	Observations	Mean	Standard error	Minimum	Maximum
GDP	31333	.0602945	1.996854	-7.126591	7.926201
MilExp	36151	.315405	2.747588	-11.44632	9.806463
GDPpc	31333	.0562393	.7096879	-2.972589	2.818814
HydroDamUp	37189	.0005162	.0057375	0	.167013
HydroDamDown	37189	.00137	.0116211	0	.3251
IrrigationDamUp	37189	.0002336	.0027293	0	.0528295
IrrigationDamDown	37189	.0010146	.0105908	0	.162
RenWaterSupplyUp	31621	1.195152	10.41211	0	160.6046
RenWaterSupplyDown	32139	.2393076	1.725134	4.13e-09	30.31779
Allies (d)	37088	.398242	.4895424	0	1
Adjacent (d)	37189	.3254457	.4685474	0	1
Common regime (d)	35655	.6090029	.4879806	0	1
ExpoDroughtUp	32888	.4028153	.3903882	.0936	3.177283
ExpoDroughtDown	32040	.4029721	.3422175	.0936	3.177283
ExpoFloodUp	33475	.0083808	.0284797	0	.5642808
ExpoFloodDown	32634	.0253398	.0863821	0	.7354792
Trade	28508	.0025178	.0071537	0	.1824901

The capacity of new dams built in the upstream and downstream basin is controlled for by using the GRanD dataset (Lehner et al. 2011, Beames, Lehner, and Anand 2019). To this end, a list of dams in the country basin area was compiled using google earth engine. *HydroDamUp* and *HydroDamDown* measure the capacity of hydropower dams constructed in the ten proceeding

and the ten preceding years (in 10^{12} cubic meters). *IrrigationDamUp* and *IrrigationDamDown* capture the capacity of irrigation and water supply dams. This time horizon was chosen as it might be possible that treaties are signed before or after the construction of dams.
The Transboundary Water Assessment Programme (TWAP) (United Nations Environment Programme 2015) provides data on country basin level. The renewable water supply, *RenWaterSupplyDown* and *RenWaterSupplyUp*, is measured as the annual average per capita sum of discharge internally generated in the particular basin region in the respective country. For those countries with missing data on the water supply in a basin but information on the total basin runoff, population size and the share of the country's basin area of the total basin area, these variables were used to predict water supply. To this end, an OLS regression was run that used the product of the total river runoff and the share of the respective state in the basin divided by the population size as an explanatory variable for the renewable water supply per capita. The regression confirms that the variable is highly correlated with the per capita water supply ($p < 0.001$). The coefficient was subsequently multiplied with the variable in order to predict and replace missing values on water supply.
Gibler (2009) provides information on formal alliances between states. The dummy variable *Allies* takes the value one if an active defense, neutrality, nonaggression or entente treaty exists between the states in the dyad. *Adjacent* is a dummy variable that takes the value one if the two countries of the dyad are direct neighbors. The Polity IV Project (Marshall, Gurr, and Jaggers 2018) provides information on the political regimes of countries. If the two countries share a common regime (democratic or autocratic), the dummy variable *Common regime* takes the value one and zero otherwise.
The variables *ExpoDroughtUp* and *ExpoDroughtDown* measure the variation in monthly river discharge. Therefore, it is a good proxy for the vulnerability to droughts of a basin region. For missing data points of this variable imputation using the basin mean was applied. The exposure to flood, *ExpoFloodUp* and *ExpoFloodDown*, is measured as the economic exposure (in dollars) to flood hazards relative to the basin size. The TWAP provides only average values for the years 1971 to 2000. Since these values are rather constant over time, they are also used for the years before 1971 and after 2000.
The economic interdependence is measured as the common trade divided by the combined GDP in the dyad. The data was taken from Trade dataset (version 4.0) of the Correlates of War project (Barbieri, Keshk, and Pollins 2009; Barbieri and Keshk 2016).

4.7.2 Adjacent states

The regressions presented in Table 4.7 include only countries that share a common border. Also between neighbor states power is not a relevant factor. Hydropower dams in the upstream state's region have a highly significant impact. When only including adjacent states, also the effect of irrigation and water supply dams becomes significant. This shows that dams are a very important motivation for a formal regulation of the water allocation. Neither type of dam constructed in the downstream country has a significant impact on the likelihood to sign a treaty. Treaties become also more likely when the downstream state suffers from water scarcity. Two neighboring democracies and two neighboring autocracies are more likely to sign an allocation treaty than two neighbors with different political systems. Interestingly the downstream state's exposure to floods and droughts has no impact on the formation of treaties.

Table 4.7 Cox regressions: Adjacent states only

	(1)	(2)	(3)	(4)	(5)	(6)
GDP	-0.0834	-0.110				
	(0.143)	(0.156)				
MilExp			0.131	0.142		
			(0.102)	(0.107)		
GDPpc					-0.566	-0.664
					(0.468)	(0.500)
HydroDamUp	17.03***	16.82**	23.15***	25.18***	19.54***	18.70***
	(6.328)	(8.068)	(5.126)	(7.682)	(4.749)	(5.546)
HydroDamDown	6.851	10.48	10.24	13.41	6.103	13.72
	(13.69)	(14.15)	(12.49)	(13.23)	(18.97)	(17.14)
IrrigationDamUp	67.78***	42.91**	94.91***	72.90***	54.71***	30.92
	(14.73)	(19.95)	(19.10)	(17.67)	(17.25)	(19.44)
IrrigationDamDown	72.26	94.86	32.57	45.80	72.12	87.63
	(68.42)	(80.45)	(68.80)	(76.00)	(64.66)	(76.95)
RenWaterSupplyUp	-0.0274	-0.0270	-0.0581	-0.0488	-0.0421	-0.0481
	(0.0566)	(0.0732)	(0.177)	(0.111)	(0.0648)	(0.0716)
RenWaterSupplyDown	-3.627**	-3.285**	-3.871**	-3.711**	-3.634**	-3.323**
	(1.759)	(1.659)	(1.876)	(1.771)	(1.677)	(1.557)
Allies (d)	0.690		0.741		0.735	
	(0.519)		(0.524)		(0.511)	
Common regime (d)	1.197**		1.148**		1.169**	
	(0.574)		(0.582)		(0.579)	
ExpoDroughtUp		0.222		0.401		0.102
		(0.822)		(0.742)		(0.960)
ExpoDroughtDown		0.0212		-0.189		0.133
		(0.456)		(0.470)		(0.475)
ExpoFloodUp		-5.503		-1.200		-9.950
		(20.22)		(21.06)		(22.79)
ExpoFloodDown		7.770		3.793		15.95
		(18.63)		(19.29)		(20.10)
Observations	8218	7884	8677	8362	8218	7884
Number of dyads	114	109	123	117	114	109
Number of treaties	21	21	22	22	21	21

Cox Proportional Hazard Model on the hazard of a country dyad to sign an allocation treaty. Coefficients are displayed. Only adjacent states are included in the analysis. The robust standard errors are presented in the parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5 Do International Agreements Improve Water Quality in Rivers?

Sven Christens

Department of Economics, University of Kassel, 34109 Kassel, Germany

Abstract: In a world with a growing population and increasing water scarcity, water pollution is a crucial global challenge. Pollution spillovers, especially from upstream riparian states to downstream states in international rivers is a common problem that aggravates the situation. In the past decades, numerous agreements were signed that aim at reducing pollution in international rivers. However, evidence regarding their effectiveness is scarce. The present study fills this gap by trying to quantify the effect of water quality agreements on pollution in international rivers. The results show that treaties are signed in situations when the water quality in a river decreases. In order to address the endogeneity of water quality agreements, I present an instrumental variable estimation that uses other types of river treaties as an instrument. The results indicate that water quality agreements are effective in reducing pollution.

5.1 Introduction

A sufficient provision of freshwater is crucial for a healthy environment and human life. However, water is becoming scarcer in many regions of the world because of climate change and an increasing world population. In the future, a growing number of people will live in areas that face water scarcity (Gosling and Arnell 2016). International rivers are a major source of freshwater. 52% of the world's population lives in international river basins (McCracken and Wolf 2019). Not only the availability of water poses a major global challenge, but also the pollution of many rivers and lakes is alarming. Inorganic and organic pollutants pose a serious threat to the provision of food and drinking water to a growing population and to the health of ecosystems, animals and humans (Schwarzenbach et al. 2010). The issue of water quality is directly addressed in the UN 2030 Agenda for Sustainable Development: Goal 6 aims to "Ensure availability and sustainable management of water and sanitation for all" (United Nations 2020). Nevertheless, the number of people affected by polluted water from untreated waste water in rivers might increase from 1.1 billion in 2000 to 2.5 billion by 2050 (Wen, Schoups, and van de Giesen 2017), which would in turn lead to a critical increase in a number of diseases (Srinivasan and Reddy 2009). Especially in situations where rivers flow from one country to another through a border, the relation between the upstream riparian state and downstream riparian state is highly asymmetric with regard to pollution. Pollution from upstream countries creates negative externalities for downstream countries but upstream states have little incentives for cooperation (Dinar 2008, 26). It has been shown that spillovers of pollution from upstream to downstream countries is common (Sigman 2002). Water quality issues are therefore one potential source for severe international conflicts.

Since no supranational body sets and enforces rules regarding pollution in international rivers, states need to negotiate these rules between themselves. When states agree to cooperate, this cooperation is often formalized in the form of treaties. In the past, numerous treaties were signed that focus on environmental and water quality issues. Treaties over water quality are particularly complex because an optimal water quality management involves close coordination of all riparian states throughout the entire basin area (Giordano 2003). In this study, I will examine whether water quality in international rivers improves when a treaty is signed by the riparian states. In the analysis, treaties that directly address water quality are considered as water quality agreements (WQAs). A challenge when evaluating the effect of WQAs is that for many countries data on water quality and the important control variables is only available for a limited number of years. These years are not always the years when agreements were signed. Therefore, it is not always possible to observe pre-agreement water quality. Another major problem when evaluating the effect of environmental agreements is that the signature of treaties is endogenously determined. Countries might self-select into treaties or pollution levels might determine whether a treaty is signed. In a first step, I will test for strict exogeneity in order to assess whether simultaneity might lead to biased estimates. The results indicate that simultaneity is a problem in the sample. Analyzing the trend in pollution reveals that WQAs are commonly signed in situations when the water quality in the basin worsens. Using an instrumental variable (IV) estimation to address endogeneity indicates that WQAs are effective in reducing pollution.

The remainder of this paper proceeds as follows. Section 2 provides background information on water quality and agreements in international river basins. Section 3 describes the data used in the analysis. Subsequently, Section 4 provides an overview of the methodological challenges when estimating the effect of WQAs on water quality. Section 5 tests for possible simultaneity between WQAs and water quality and examines the trend in pollution before and after the signature of WQAs. In Section 5, an instrumental variable is introduced and used to test the effect of WQAs on water quality. Finally, Section 6 concludes the findings of this study.

5.2 Background

5.2.1 Water quality

In the analysis, water quality is measured in biological oxygen demand (BOD) and dissolved oxygen (DO) that indicate organic pollution. The two main sources of organic pollution are wastewater from cities and intensive livestock farming (Wen, Schoups, and van de Giesen 2017). Most industrialized countries were successful in reducing point source pollution from single sources such as industrial sites. Nonpoint pollution mostly from agriculture remains to be a major source of pollution. In many developing countries industrial pollution is still a severe problem (Olmstead 2010). Water quality in rivers is closely related to a country's wealth. The relation between GDP per capita and pollution can be described by the Environmental Kuznets Curve (EKC). As countries become richer and more industrialized, pollution increases. However, once a certain level in income is reached, countries put more effort into pollution control and treatment and the levels of pollution decrease as incomes increase further. Grossman and Krueger (1995) find the peak of pollution measured in BOD to be at $7853 per capita and at $2703 for DO. Other studies argue that the inverted U-shaped EKC relationship only exits for North America and Europe (Lee, Chiu, and Sun 2010). Also the trade openness of a country impacts BOD levels by affecting capital intensity and environmental regulations (Managi, Hibiki, and Tsurumi 2009). Common trade between riparian states tends to reduce pollution spillovers from upstream states to downstream states as common trade provides more room for issue-linkages that allow downstream states to gain concessions from upstream states (Sigman 2004; Bernauer and Kuhn 2010).

To improve water quality in rivers can be a costly process. For instance, the 1972 U.S. Clean Water Act has significantly improved water quality in the US but only at very high costs that might exceed the monetary benefits (Keiser and Shapiro 2019). India's national program shows only limited success in reducing water pollution, while it successfully reduces air pollution (Greenstone and Hanna 2014). A possible explanation for this relatively poor performance of water programs might be that it is relatively easy to substitute and to avoid polluted water. Another explanation of the difficulties to address water pollution is the lack of market based policies that exist for example in the case of air pollution (Olmstead 2010).

5.2.2 International agreements

Until today, there are no global framework conventions that regulate pollution in international rivers. The UN 1997 'Convention on the Law of Non-Navigational Uses of International Watercourses' was an attempt to define global principles. The convention requires states to exchange water quality data and encourages to jointly set water quality objectives. While a majority of countries voted to approve the convention only 36 have currently ratified it, which was enough for its entry into force.[24] In practice though, the concrete implementation of the principles can be difficult (Giordano 2003). Because mostly downstream states signed it, it is unlikely that it has an impact on pollution in international basins. Therefore, states have to rely on bilateral or multilateral negotiations in order to manage international basins. Cooperation between riparian states in international rivers can provide benefits but in order to do so a system-wide management is needed (Sadoff 2002). Hundreds of treaties were signed in the last century in order to establish such cooperation (Giordano et al. 2014). Hydropower and water allocation were the most common issues of treaties before 1950, while water quality and environmental issues were only minor concerns. Since then water quality has become the most frequent focus of treaties (Giordano et al. 2014). After the 1992 UN Conference on Environment and Development, addressing environmental issues in international rivers became also common in the Global South (Conca, Wu, and Mei 2006).

Previous studies identified a number of factors that impact the creation of treaties in international river basins, such as common trade, power asymmetries, common political systems of riparian states and geographical factors (Tir and Ackerman 2009; Dinar, Dinar, and Kurukulasuriya 2011; Zawahri and Mitchell 2011). Side payments can also play an important role in the formation of treaties. Rich downstream states that have a higher willingness-to-pay for environmental conservation are likely to pay poorer upstream states for pollution control (Dinar 2008, 78ff.). For instance, in the 'Convention on the Protection of the Rhine Against Pollution by Chlorides' from 1976, the Netherlands agreed to pay a higher share of the costs for pollution control than the upstream countries (Barrett 1994).

Cooley and Gleick (2011) distinguish these bilateral or multilateral agreements from general principles in international water law such as the EU Water Framework Directive or the United Nations Economic Commission for Europe Water Convention. In the analysis, I will distinguish between general frameworks and WQAs as well but focus on the effect of WQAs. Only agreements that directly address water quality are considered in the analysis. Sometimes other

[24] The current status can be found here: https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-12&chapter=27&lang=en (accessed: January 2020)

issues addressed in agreements have an influence on water quality. For instance, the allocation of water between countries is often addressed in treaties and could sometimes have an impact on water quality as a side effect. In cases when allocation of water leads to an increase in runoff from upstream states to downstream states, the water quality in downstream countries can improve. Pollution measured as per parts decreases with higher runoff. Another reason is that higher levels of runoff can help to prevent salt water intrusion in deltas (Cooley and Gleick 2011).

Previous studies that examined the effectiveness of treaties focused mostly on the question whether treaties in international river basins are able to reduce conflicts. Treaties are found to increase cooperation in the basin (Brochmann 2012) and to reduce the likelihood of the occurrence of military conflicts (Brochmann and Hensel 2011). Dinar et al. (2019) find that the number of treaties signed in a basin has a positive impact on cooperation but does not reduce conflict. Other studies highlight the importance of the institutional design of treaties when evaluating their effectiveness in reducing conflicts. Mitchell and Zawahri (2015) identify information exchange and enforcement as the most effective features of treaties in preventing the militarization of conflicts. Dinar et al. (2015) find that water allocation treaties are only effective in promoting cooperation and reducing conflict when the allocation mechanism can flexibly adapt to changes in the flow and at the same time specifies a clear rule with regard to water quantities. In addition, treaties are found to reduce potentially harmful dam constructions in upstream riparian states (Olmstead and Sigman 2015).

While a number of studies examines the formation of agreements and their effectiveness in reducing conflicts, their impact on the environment received little attention so far. To the best of my knowledge only two studies exist that test the effect of treaties on water quality in international rivers in large-N studies. Köppel and Sprinz (2019) analyze the effect of multilateral agreements on water quality in upstream-downstream rivers in Europe. The authors distinguish between legally binding and legally non-binding agreements. The results suggest that only legally binding agreements are effective in reducing pollution from upstream states to downstream states. Bernauer and Kuhn (2010) also study river pollution from upstream states to downstream states in Europe. They find that the number of environmental agreements ratified by a country is positively related to its water quality. The number of water quality agreements a country signed has no significant effect on water quality in most models. The analysis presented here differs from these two studies by testing the effect of treaties not just in Europe but worldwide. Moreover, both studies do not address endogeneity, which might lead to biased estimates.

5.3 Data

5.3.1 Water quality data

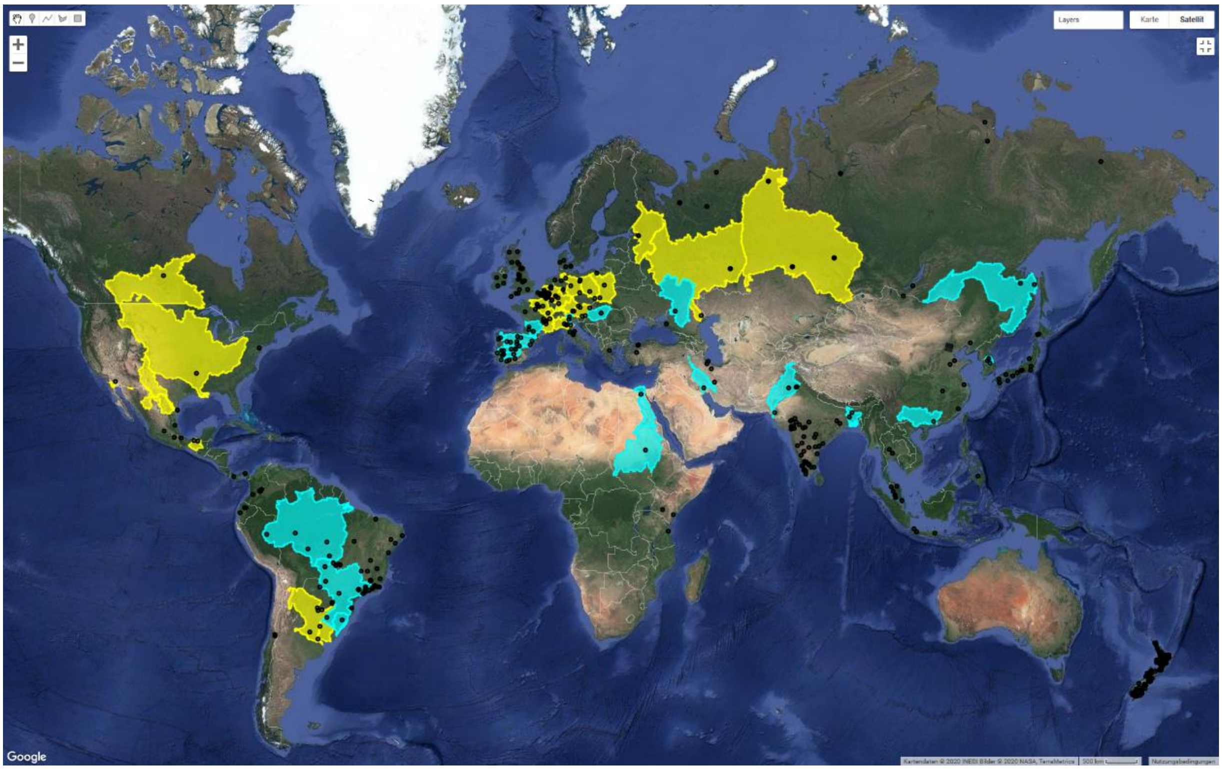
The biological oxygen demand (BOD) is chosen as the main measure for water quality in the analysis. BOD is a comprehensive indicator that is commonly used to measure water quality with higher values indicating worse quality. It measures the amount of dissolved oxygen that bacteria consume when decomposing organic matters. This decomposition is a natural process and leads to a decrease in organic pollution over time. Modern water treatment plants are able to remove organic solids almost entirely by using a mixture of mechanical, biological and chemical processes. But even simple mechanical measures (called primary treatment) are able to significantly reduce the number of organic solids and, thereby, BOD levels (Food and Agriculture Organization of the United Nations 2020). The data on BOD was provided by the United Nations' Global Environment Monitoring System Water Programme for freshwater (GEMS/Water, United Nations 2019) that gathered data from stations in rivers around the world. The data from the stations is collected and transmitted to the GEMS Water Programme by each country. Table 5.1 presents the average BOD levels by continent separated by domestic and international river basins. Water quality is better in international basins than in domestic ones on all continents except for Europe. Rivers with BOD concentrations above 8 mg/l are considered to be severely polluted (United Nations 2016). The high levels in North America are mostly driven by observations from Guatemala and Mexico.

Table 5.1: Average BOD levels by continent in domestic and international basins

	Domestic basins	International basins
Africa	5.78	3.28
Asia	6.26	2.30
Australia and Oceania	0.56	
Europe	3.12	3.83
North America	22.65	14.67
South America	5.60	3.01

Data on BOD from 493 stations could be matched with all of the control variables. The black points in Figure 5.1 represent the location of the GEMS-stations around the world. The countries with most stations are New Zealand (80), Brazil (65), India (43), Belgium (35), Russia (32) and Colombia (29). The stations are located in 43 different countries. On average the stations provide data for 7.4 years.

Figure 5.1 Stations and basins in the sample



The map displays the stations that collect data on water quality in black, international country-basin units that are regulated by a WQA in 2007 in yellow and international country-basin units with no WQA in 2007 in cyan.

Dissolved oxygen (DO) is an additional measure for water quality used in the analysis. It measures the dissolved oxygen per liter of water and is also collected by the GEMS-stations. The sample mean is 7.6 mg/l, with higher values indicating more oxygen that is available to organisms in the water and, thus, a better water quality. Low levels of DO can be a sign of too many bacteria and a resulting high oxygen demand. Both indicators, BOD and DO, measure organic pollution and are closely related. In the analysis, I will focus on BOD since it is the most commonly used measure for the effectiveness of water treatment.

5.3.2 Water quality agreements

The data on international WQAs was taken from the International Freshwater Treaties Database (Giordano et al. 2014; Transboundary Freshwater Dispute Database 2018). The database provides information on the issue areas of all different types of river treaties. Whenever a treaty is coded as relating to “water quality or water-related environmental concerns” it is included in the analysis. I included any document that was coded as a primary agreement, replacement of a primary agreement, amendment to a primary agreement or a protocol that adds further aspects to a primary agreement. Examples are the ’Agreement regarding the Colorado River Salinity’ signed between Mexico and the USA in 1965 and the ’Agreement between the Federal Republic of Germany, the Czech Republic and the Republic of Poland on Protection of the River Oder from Pollution’ signed in 1996.

The database provides only information on treaties until 2007. Therefore, the effect of agreements on water quality is analyzed only until this year. Google earth engine was used to match stations with international river basins and countries. The geospatial information on the basin area was taken from the Register of International River Basins (McCracken and Wolf 2019). Overall, 23 country-basins with WQAs and 26 country-basins from international rivers without WQA enter the analysis. Figure 5.1 shows these international country-basins. The yellow country-basins were regulated by an agreement in 2007, while the cyan ones were not. International basins without water quality data are not highlighted in the map.

5.3.3 Control variables

Since BOD is expressed in milligrams of oxygen consumed per liter, the flow or discharge of the river is an important control variable. Unfortunately, information on river discharge is not collected by the GEMS-stations. Information on river discharge was taken from the Global Runoff Data Centre (Global Runoff Data Centre 2019). In order to combine data on water quality and discharge, the GEMS-stations and the stations of the Global Runoff Data Centre were matched with the help of a list provided by the GEMS Water Programme and spatial matching using google earth engine.[25] Another important control variable is the annual mean water temperature because organic pollution decreases faster in colder water. Data on temperature is also captured by the GEMS stations. Since previous studies found an inverted U-shaped relation between water pollution and income levels, GDP per capita is included as a control variable in logarithm and in squared level. Data on GDP per capita (thousand current US$) was taken from the World Bank (2019). Table 5.8 in the Appendix provides the summary statistic of the variables. In the Appendix additional regressions (Table 5.9) are shown that use a cross-sectional pooled model. This allows to control for time-invariant variables. These are the number of people living in the upstream catchment area of the stations and whether the station is located in the upstream or the downstream area of another country.

5.4 Methodology

Evaluations of programs or treatments that are not implemented randomly need to address the problem of endogeneity in order to estimate causal effects. Endogeneity exists if an explanatory variable is correlated with the error term. In this case, the estimated coefficients of the endogenous variable would be biased. The common causes for endogeneity are omitted variable bias, simultaneity and measurement errors. An omitted variable bias exists if variables that are not included in the regression are correlated not only with water quality but also with at least one of the explanatory variables. WQAs might be prone to this bias because they are not randomly signed. Certain unobserved characteristics of the basin are likely to affect both, water

[25] GEMS-stations and the stations of the Global Runoff Data Centre were matched if they were located less than 10km from each other.

quality and the likelihood to sign a treaty. The problem of an unobserved heterogeneity can be addressed by using panel models that control for the unobserved heterogeneity of subjects (Wooldridge 2010, 247ff.), such as fixed effects models (Gormley and Matsa 2014). An important assumption is that these unobserved effects are time-invariant.

A major problem when estimating the effect of WQAs on water quality is that some agreements were signed before data on water quality was available. Because the fixed effects regressions allow only to examine variation within the stations, the effect of WQAs signed before water quality data was available cannot be estimated in fixed effects models. I, therefore, test the effect of the first WQAs signed after data on water quality becomes available. In some cases, these treaties are the second or third treaties signed in the basin. The variable used when testing the effect of the first treaty with data is labelled WQA^{*}_{it}. It is set to one in the year of the signature and keeps this value in all subsequent years.

The other important reason for endogeneity biases is the existence of simultaneity. It is likely that WQAs do not only affect water quality but that the water quality in a basin has an impact on the decision to sign agreements as well. This simultaneity between water quality and the signature of agreements would cause endogeneity. In the next section, I will provide a test for strict exogeneity in order to test for simultaneity.

Finally, measurement errors can cause an endogeneity bias. In general, data on BOD is very reliable compared to other types of emissions because the measurement is relatively cheap and widely understood (Hettige, Mani, and Wheeler 1998 in Managi, Hibiki, and Tsurumi 2009). However, to some degree strategic reporting by the countries that report water quality data cannot be excluded (Beck, Bernauer, and Kalbhenn 2010).

Throughout the empirical analysis, I will present all regressions in three different specifications with regard to the countries that are included. In the first specification, all countries are included. The second specification excludes the countries of the EU and the third specification exclusively analyzes the countries of the EU. The reason for this procedure is that previous studies could not find the same degree of free-riding of upstream states in the EU as in the rest of the world (Sigman 2002). The countries of the EU also signed treaties over water quality more frequently than countries in other regions.

5.5 Results

I start this section by providing a test for strict exogeneity. Subsequently, the trends in pollution before and after the signature of treaties are presented. In an attempt to address endogeneity, this section further provides an IV estimation. I end the section with an analysis of pollution levels in upstream basins.

5.5.1 Strict exogeneity

In the following, I will test whether simultaneity is a problem when estimating the effect of WQAs by applying a simple test for strict exogeneity. Under strict exogeneity the error term, *u*, at time *t*, is uncorrelated with the explanatory variables at any given time *s*.

$$E[u_t|x_s] = 0 \; \forall s \qquad (1)$$

This means that under strict exogeneity future levels of an explanatory variable cannot be correlated with the dependent variable. Wooldridge (2010, 325) shows how this can be tested in a fixed effects model with T>2:

$$\boldsymbol{y}_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \boldsymbol{w}_{i,t+1}\boldsymbol{\delta} + c_i + u_{it} \qquad \text{with t=1,2,…,T-1} \qquad (2)$$

where $\boldsymbol{w}_{i,t+1}$ is a subset of $\boldsymbol{x}_{i,t+1}$. Under strict exogeneity $\boldsymbol{\delta} = 0$ must hold.

Baier and Bergstrand (2007) use this method to examine the effect of trade agreements on trade. They argue that it is crucial to test for a possible feedback relation between the two variables. Trade agreements might not only impact the level of trade between two nations but the level of trade might also influence whether a trade agreement is signed. In their regression, not only the lagged levels of agreements are included but also lead of agreements (future levels). In the case of strict exogeneity, future levels of agreements must be uncorrelated with the dependent variable.

In a similar manner, I will test whether causality runs not only from WQAs to water quality but also from water quality to the decision to sign an agreement. In the regressions, the variable WQA_{it}^s takes the value one in the year when the agreement is signed. Lagged levels of WQA_{it}^s are included in order to account for lagged effects of WQAs on BOD levels. These lagged effects can for example occur when the necessary infrastructure to comply with an agreement needs several years to be constructed. In some cases, states might also need several years to ratify an agreement. Also the leads of WQA_{it}^s are included. If no causality runs from water quality to the occurrence of agreements, these variables should be insignificant.

The control variables include discharge, water temperature and income. Some of the explanatory variables might have a multiplicative effect on water quality, for example, temperatures and discharge (Sigman 2005). Therefore, the control variables enter in logarithmic form. In order to control for a possible non-linear relation between income levels and pollution, GDP per capita is also included in squared form. In order to avoid perfect collinearity, squared GDP enters in levels. Also the existence of a general water framework is controlled for in the form of the indicator variable GF_{it}. The fixed effects regression takes the following form:

$$\begin{aligned} lnPollution_{it} = {} & \beta_1 WQA_{i,t+3}^s + \cdots + \beta_4 WQA_{it}^s + \cdots + \beta_9 WQA_{i,t-5}^s + \beta_{10} lnDischarge_{it} \qquad (3) \\ & + \beta_{11} lnTemperature_{it} + \beta_{12} lnGDPpc_{it} + \beta_{13} GDPpc_{it}^2 + \beta_{14} GF_{it} \\ & + \sum_{t=1}^{T-1} \gamma_t d_t + \alpha_i + u_{it} \end{aligned}$$

where $lnPollution_{it}$ is the logarithm of BOD in year t measured at station i. In addition to the control variables already introduced, the model includes year dummies, d_t, in order to control for time trends. α_i is the time-invariant station fixed effect. This effect captures the unobserved heterogeneity of stations and, therefore, also country specific time-invariant effects. u_{it} is the idiosyncratic error term.

An important assumption in the tests is that also second, third and other subsequent WQAs signed have an impact on water quality. Section 5.7.2 in the Appendix uses the number of agreements as an explanatory variable in cross-sectional regressions in order to test if water quality differs between basins depending on the number of WQAs signed. The results provide evidence that water quality is better when the number of agreements increases. The results do not allow to conclude that WQAs have a causal effect on water quality, but encourage to also consider subsequent agreements in the analysis.

Table 5.2 presents the test for strict exogeneity. The regression models (1) and (2) include all stations in the sample, the regression models (3) and (4) exclude stations in EU countries and the regression models (5) and (6) exclusively include stations inside the EU. All models test if the signature of agreements is correlated with pollution in the three years preceding the signature. I assume that three years should be sufficient to detect possible simultaneity. In the regression models (1), (3) and (5), lags of WQA_{it}^{s} are used to estimate the effect of WQAs on water quality for each of the five years following the signature separately. In the regression models (2), (4) and (6), the indicator variable WQA_{it}^{*} is used in order to estimate the effect of WQAs on water quality. It measures the effect of WQAs not for a specific year but for the year of the signature and all proceeding years.

Discharge has no significant effect on pollution levels in the first four regression models. The variable is only statistically significant on a 10% level when testing only EU countries, as can be seen in the models (5) and (6). The coefficient takes a negative sign, as expected. Higher volumes of water flowing through a river decrease the concentration of BOD. The coefficient of *lnTemperature* is not significantly different from zero in any model. When examining all countries, a positive relation between GDP per capita and pollution can be found. In countries other than the EU, $GDPpc^2$ is significantly negative, which indicates the existence of a Kuznets curve. Income levels in the EU are not related to pollution. General water frameworks are found to significantly reduce pollution outside the EU and when testing all countries jointly.

Recalling that under strict exogeneity future levels of an explanatory variable cannot be correlated with the dependent variable, variables $WQA_{i,t+2}^{s}$ and $WQA_{i,t+1}^{s}$ draw the attention, as they are highly significant in the regression models (1), (2), (5) and (6). Only when examining exclusively non-EU countries the variables are not significant. In regression model (4), $WQA_{i,t+3}^{s}$ enters positively on a 10% significance level.

Table 5.2 Fixed effects regression testing for strict exogeneity

	(1) All countries	(2) All countries	(3) Without EU	(4) Without EU	(5) EU only	(6) EU only
Dependent variable	lnBOD					
lnDischarge	-0.00279	-0.00265	0.00612	0.00498	-0.0616*	-0.0597*
	(0.0147)	(0.0147)	(0.0159)	(0.0159)	(0.0364)	(0.0355)
lnTemperature	-0.106	-0.0921	-0.0726	-0.0857	-0.110	-0.101
	(0.0665)	(0.0669)	(0.0838)	(0.0864)	(0.0986)	(0.100)
lnGDPpc	0.311***	0.316***	0.385***	0.385***	0.0849	0.0476
	(0.0793)	(0.0782)	(0.0963)	(0.0943)	(0.241)	(0.200)
$GDPpc^2$	-0.0000902	-0.0000789	-0.000205**	-0.000202**	-0.000321	-0.000193
	(0.0000878)	(0.0000862)	(0.0000978)	(0.0000977)	(0.000214)	(0.000193)
GF	-0.226**	-0.209**	-0.364***	-0.357***	0.168	0.0790
	(0.0978)	(0.0968)	(0.132)	(0.135)	(0.198)	(0.185)
$\mathrm{WQA}^s_{i,t+3}$	0.0781	0.0995*	0.153	0.466***	0.0835	0.126**
	(0.0688)	(0.0529)	(0.0990)	(0.146)	(0.0739)	(0.0576)
$\mathrm{WQA}^s_{i,t+2}$	-0.182**	-0.176***	0.0319	0.232	-0.203***	-0.145***
	(0.0741)	(0.0508)	(0.138)	(0.151)	(0.0700)	(0.0548)
$\mathrm{WQA}^s_{i,t+1}$	-0.191***	-0.243***	0.0450	0.237	-0.165***	-0.173***
	(0.0656)	(0.0544)	(0.244)	(0.205)	(0.0623)	(0.0548)
WQA^s_{it}	-0.292***		-0.399***		-0.245***	
	(0.0801)		(0.125)		(0.0772)	
$\mathrm{WQA}^s_{i,t-1}$	-0.0824		0.00797		-0.154*	
	(0.0958)		(0.216)		(0.0874)	
$\mathrm{WQA}^s_{i,t-2}$	-0.150		-0.453**		-0.0544	
	(0.124)		(0.179)		(0.138)	
$\mathrm{WQA}^s_{i,t-3}$	-0.0638		-0.300*		0.0460	
	(0.0843)		(0.178)		(0.0973)	
$\mathrm{WQA}^s_{i,t-4}$	0.0275		-0.0660		0.0449	
	(0.0791)		(0.150)		(0.0870)	
$\mathrm{WQA}^s_{i,t-5}$	0.00935		-0.174		0.00968	
	(0.0768)		(0.141)		(0.0679)	
WQA^*_{it}		-0.259***		0.121		-0.256***
		(0.0568)		(0.214)		(0.0540)
N	3429	3432	2580	2580	849	852
R^2 within	0.175	0.176	0.205	0.202	0.233	0.238

Fixed effects regressions. Robust standard errors in parentheses. Year dummies are included but not shown.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Thus, the results provide strong evidence that strict exogeneity is violated and that simultaneity is present, in particular in the EU countries and when testing all countries jointly. WQA^s_{it} is significantly positive in all specifications, but the direction of causality is not clear. Pollution levels in the year of the signature might have an impact on the likelihood to sign a treaty, but WQAs might also reduce pollution already in the year they are signed.

The regressions (1), (3) and (5) show that pollution levels in the years after the signature are mostly unaffected by treaties. WQA_{it}^{*} is significant when examining all countries. The coefficient indicates that BOD levels are reduced by 22.8% after the signature of an agreement.[26] The estimated effect is similar for EU countries, while the variable remains insignificant for countries other than the EU. However, it is important to remember that the estimates in this model might be biased due to endogeneity.

5.5.2 Trend

In order to analyze the relation between the signature of agreements and prior water quality, in the following I estimate the trend in pollution levels before and after WQAs are signed. To this end, I introduce the variable *Trend* that takes the value 1 ten years before the signature of a treaty and counts up until ten years after the signature (*Trend*=21). The variable takes the value 11 in the year a treaty is signed.[27] The overall trend is interacted with the indicator variable $WQA*$ in order to test if the trend is significantly different after the signature of treaties than it is before the signature.[28]

The results are presented in Table 5.3. The first three regressions use BOD as the measure of pollution while the regressions (4) to (6) use DO as a measure. The variable *Trend* is significantly positive in all three models that test the effect on BOD. In the 10 years before the signature of a treaty BOD increases by 4.7% per year in countries outside the EU and by 1.8% in the EU. The insignificant interaction terms demonstrate that this trend is not significantly different after the signature of WQAs. However, in countries other than the EU, BOD is significantly shifted downwards as indicated by the variable $WQA*$. The coefficient of -1.038 is equivalent to a reduction by 64.6%. This strong drop in BOD is not offset by the positive trend.

The trend in DO is only significant for non-EU countries. The negative coefficient indicates that pollution measured in DO increases before the signature of WQAs. A possible explanation for the difference in the EU and other countries might be that the EU is comparably rich. Pollution measured in DO tends to decrease at much lower income levels than BOD (Grossman and Krueger 1995). Therefore, DO might not present a problem in the EU. The positive trend in DO in non-EU countries is brought to an end once a treaty is signed as indicated by the interaction term. Testing the trend after the signature reveals that it is not significantly different

---

[26] $(e^{-0.259} - 1) * 100 = -22.8\%$

[27] If two treaties were signed less than 20 years from each other, so that positive and negative values overlap, the variable takes the value of the year that is closer to the signature.

[28] In the remainder of the paper the indexes of WQA_{it}^{*} are no longer shown for reasons of simplicity.

from zero. While the pollution does not increase anymore, the water quality does also not improve.
In order to provide a visualization of the trends before and after WQAs are signed, the ten years before and after the treaties are coded as indicator variables and estimated in a fixed effects regression. Figure 5.2 plots the estimated dummies for each year. The year 0 presents the year a treaty is signed. The two left panels present the trend of all countries in the sample, the two panels in the middle show non-EU countries and the two right panels depict only EU countries. The upper panels show the trend in logarithmic BOD, while the lower panels show the trend in logarithmic DO.

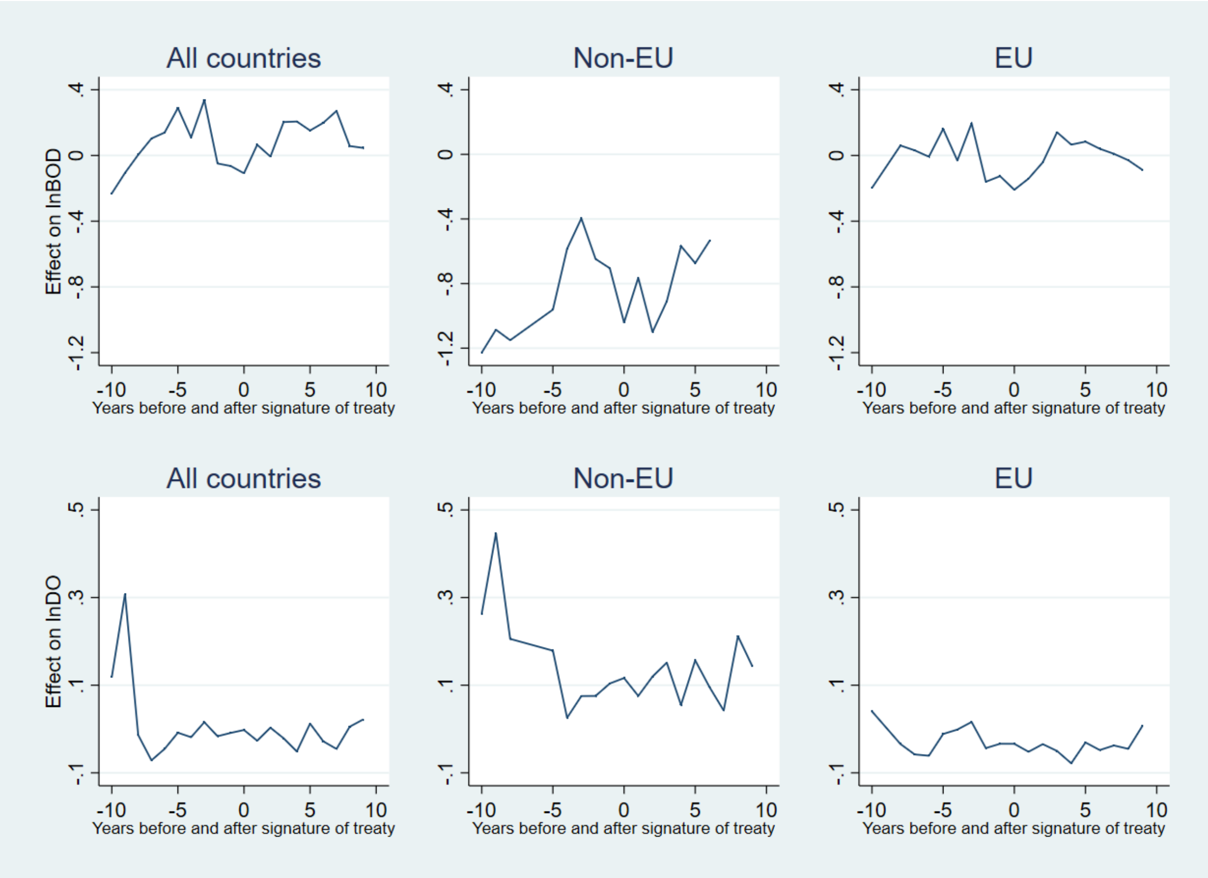
Table 5.3 Fixed effects regression: Trend

	(1) All countries	(2) Without EU	(3) EU only	(4) All countries	(5) Without EU	(6) EU only
Dependent variable		lnBOD			lnDO	
Trend	0.0186***	0.0461*	0.0175***	-0.000872	-0.00443**	-0.000763
	(0.00375)	(0.0256)	(0.00356)	(0.000714)	(0.00219)	(0.00132)
WQA*	-0.106	-1.038**	-0.0122	0.0325*	-0.116	-0.00892
	(0.140)	(0.489)	(0.140)	(0.0184)	(0.0790)	(0.0229)
Trend*WQA*	-0.0100	0.0516	-0.0183	0.0000956	0.0112**	0.00190
	(0.0100)	(0.0338)	(0.0112)	(0.00108)	(0.00516)	(0.00139)
lnDischarge	-0.00459	0.00297	-0.0592*	-0.00899	-0.0106	-0.00609
	(0.0145)	(0.0157)	(0.0354)	(0.0114)	(0.0130)	(0.0145)
lnTemperature	-0.0976	-0.0746	-0.118	-0.0922***	-0.0748***	-0.133***
	(0.0667)	(0.0849)	(0.0995)	(0.0188)	(0.0241)	(0.0417)
lnGDPpc	0.265***	0.282***	0.0484	-0.0179	-0.0139	0.0562
	(0.0672)	(0.0850)	(0.190)	(0.0192)	(0.0253)	(0.0605)
$GDPpc^2$	-0.0000554	-0.000154	-0.000228	0.0000556***	0.0000488**	0.00000143
	(0.0000807)	(0.0000952)	(0.000200)	(0.0000169)	(0.0000247)	(0.0000389)
GF	-0.236**	-0.419***	-0.173	0.0392**	0.0512**	0.0210
	(0.0976)	(0.151)	(0.288)	(0.0159)	(0.0235)	(0.0908)
N	3622	2746	876	4775	3536	923
R2 within	0.173	0.199	0.255	0.0388	0.0477	0.192

Fixed effects regressions. Robust standard errors in parentheses. Year dummies are included but not shown.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The graphs show that the trend in pollution is less clear than the regressions would suggest. The clearest decrease in water quality can be seen when looking at DO in all countries and in non-EU countries. However, the regressions do not confirm that DO significantly decreases when testing for the trend in all countries. A possible explanation is that less observations tend to be available for time periods further away from the year of signature than for the years around the signature. Thus, the estimated levels at the edges of the graph rely on fewer observations and might be driven by relatively few values. This is accounted for in the regressions but not when displaying the graphs.

Figure 5.2 Pollution trend before and after WQA signature



5.5.3 IV estimation

It is a common procedure for riparian states in international river basins to first address border and navigation issues and only if these are defined to deal with pollution. For instance, in 1960, the Netherlands and Federal Republic of Germany signed a treaty that defined their borders and navigation rights in the 'Treaty between the Kingdom of the Netherlands and the Federal Republic of Germany concerning the course of the common frontier, the boundary waters, real property situated near the frontier, traffic crossing the frontier on land and via inland waters, and other frontier questions'. The navigation rights were most important for traffic on the Rhine. 16 years later in 1976, the two states, together with France, Switzerland and Luxembourg, addressed pollution in the Rhine river with the 'Convention on the protection of the Rhine against chemical pollution' and the 'Convention on the Protection of the Rhine Against Pollution by Chlorides'. An additional protocol to the 'Convention on the Protection of the Rhine Against Pollution by Chlorides' was added in 1991 (Transboundary Freshwater Dispute

Database 2018).[29] I take advantage of this circumstance and use treaties that exclusively deal with navigation and/or border issues as an instrument for WQAs. It seems reasonable to assume that regulating the navigation and defining borders has no direct impact on water quality. Adequate instrumental variables allow to account for endogeneity.[30] However, the instrument used here has its limits. The timing of the two treaties differs substantially. The highest correlation exists when the lead by 20 to 30 years of border and navigation treaties is taken. Therefore, I will use the lead by 20 as well as 30 years of border and navigation treaties as instrumental variables in separate regressions. The instrument is also too weak to be used when analyzing only non-EU countries.

The IV regressions on the effects of BOD are presented in Table 5.4 together with fixed effects regressions. In the regression models (1), (4) and (7), *WQA** is treated as an exogenous variable in fixed effects regression models. In all other models, the variable is instrumented. In regression model (1), *WQA** shows a significantly negative sign. The estimated reduction in BOD increases substantially when *WQA** is instrumented. The coefficient is -0.765 when the lead by 20 years of border and navigation treaties is taken. This is equivalent to a reduction by 53%. The estimated reduction is 46.6% when the lead by 30 years is taken as can be seen in model (3). In the fixed effects regression for non-EU countries, *WQA** is insignificant. Because the test for strict exogeneity provided only little evidence for the existence of simultaneity in these countries, the estimate might be unbiased and indicate that WQAs are not effective in non-EU countries. At the same time, the regression models (2) and (3) find a strong effect of WQAs on water quality in all countries (which even exceeds the estimated effect in EU countries). Therefore, the results do not allow to draw a certain conclusion on the effect of WQAs on BOD in non-EU countries. When only the EU countries are analyzed the effect is similar as for all countries: the already significant effect of *WQA** is strongly increased in the IV regressions (6) and (7). The null hypothesis that *WQA** is exogenous is clearly rejected when all countries are jointly tested. For EU countries the test provides only moderate evidence of endogeneity. The F-statistic of the Kleinbergen-Paap test is clearly above the rule of thumb of 10 and, therefore, indicates that the instrument is relevant and not weak (Staiger and Stock 1997). However, there is a chance that the serial correlation of *WQA** and the instrument might inflate the values.

Table 5.5 tests if WQAs are also effective in increasing DO levels. *WQA** is estimated to only significantly reduce DO when including all countries in the sample. In the IV regressions, the

[29] Another example can be found in the basin of the Rhone. In 1959, France and Switzerland slightly redefined their border using the small river Boiron de Morges, which feeds into the Rhone, in the ‘Convention between Switzerland and France concerning the Correction of Ruisseau (Le Boiron)’. Only three years later, in 1962, the two states signed the ‘Convention concerning Protection of the Water of Lake Geneva against Pollution’, which was aiming at improving water quality in the Rhone. 12 years later, in 1971, the two states signed the ‘Exchange of Letters constituting an Agreement between France and Switzerland on the Implementation of the Convention of 16 November 1962, concerning Protection of the Waters of Lake Geneva against Pollution’.

[30] In the first stage of the IV estimation, the variance of *WQA** is estimated using OLS although the variable is binary in order to avoid the ‘forbidden regression’. Angrist and Pischke (2009, 190) offer a discussion of the issue.

effect has a higher significance but is less in scale than in the fixed effects regression. The estimated increase in DO is equivalent to 14% in model (2) and 13% in model (3). When testing the effect of *WQA** in EU countries only a weakly significant effect on DO can be found in model (7) but no significant effect in other models. The test for endogeneity suggests that it is only a problem when including all countries in the estimation. In the EU, the null hypothesis of exogeneity is not rejected. This result is in line with the trends observed before. DO tends to increase worldwide before the signature of WQAs but not in the EU. Simultaneity and the resulting endogeneity might be less of a problem when examining DO in the EU.

Table 5.4 IV regressions: BOD

	(1) All countries Fixed effects	(2) All countries IV (20 years)	(3) All countries IV (30 years)	(4) Non-EU Fixed effects	(5) EU Fixed effects	(6) EU IV (20 years)	(7) EU IV (30 years)
lnDischarge	-0.00456	-0.00324	-0.00355	0.00299	-0.0638*	-0.0708**	-0.0677**
	(0.0146)	(0.0125)	(0.0124)	(0.0158)	(0.0372)	(0.0301)	(0.0299)
lnTemperature	-0.0910	-0.0666	-0.0723	-0.0594	-0.0999	-0.101	-0.101
	(0.0667)	(0.0604)	(0.0597)	(0.0843)	(0.0992)	(0.0640)	(0.0632)
lnGDPpc	0.260***	0.289***	0.282***	0.287***	0.0523	-0.0803	-0.0220
	(0.0673)	(0.0598)	(0.0592)	(0.0809)	(0.200)	(0.183)	(0.171)
GDPpc2	-0.0000578	-0.0000470	-0.0000495	-0.000150	-0.000218	-0.0000442	-0.000121
	(0.0000808)	(0.0000499)	(0.0000493)	(0.0000943)	(0.000206)	(0.000188)	(0.000171)
GF	-0.222**	-0.196***	-0.202***	-0.361***	-0.191	-0.365	-0.289
	(0.0965)	(0.0759)	(0.0754)	(0.135)	(0.293)	(0.271)	(0.253)
WQA*	-0.172***	-0.765***	-0.627***	-0.0447	-0.186***	-0.514***	-0.369**
	(0.0632)	(0.171)	(0.134)	(0.168)	(0.0501)	(0.190)	(0.153)
N	3622	3565	3565	2746	876	845	845
Test endogeneity		0.000	0.0002			0.0587	0.2104
Kleinbergen-Paap (F-statistics)		34.242	43.766			23.916	30.259

Fixed effects regressions and IV regressions using station fixed effects. Robust standard errors in parentheses. Year dummies are included but not shown. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In additional regressions, presented in Section 5.7.3 in the Appendix, I replace missing values on discharge with zero and add a dummy that controls for missing values. This more than doubles the number of observations and allows to include additional basins and treaties. Figure 5.4 in the Appendix shows a map of the stations and basins in the sample. The procedure has been used in previous studies but with less observations (Sigman 2002; Bernauer and Kuhn 2010). The results for the EU are similar in the regressions. However, the results change when all countries are considered. *WQA** is found to improve water quality when it is included as an exogenous variable in fixed effects regressions. Once the variable is instrumented it no longer has a significant effect neither on BOD nor on DO. It is not clear whether the change in the results is the consequence of the additional agreements and basins that are analyzed or because of the artificial replacement of missing variables.

Table 5.5 IV regressions: DO

	(1) All countries Fixed effects	(2) All countries IV (20 years)	(3) All countries IV (30 years)	(4) Non-EU Fixed effects	(5) EU Fixed effects	(6) EU IV (20 years)	(7) EU IV (30 years)
lnDischarge	-0.00899	-0.00902	-0.00902	-0.0107	-0.00608	-0.00530	-0.00502
	(0.0114)	(0.00767)	(0.00767)	(0.0129)	(0.0145)	(0.0115)	(0.0115)
lnTemperature	-0.0925^{***}	-0.0957^{***}	-0.0953^{***}	-0.0748^{***}	-0.133^{***}	-0.133^{***}	-0.133^{***}
	(0.0187)	(0.0172)	(0.0172)	(0.0237)	(0.0417)	(0.0316)	(0.0317)
lnGDPpc	-0.0181	-0.0219	-0.0214	-0.0114	0.0576	0.0713	0.0763
	(0.0192)	(0.0161)	(0.0161)	(0.0250)	(0.0599)	(0.0558)	(0.0554)
GDPpc2	0.0000567^{***}	0.0000546^{***}	0.0000549^{***}	0.0000487^{**}	0.000000332	-0.0000224	-0.0000309
	(0.0000168)	(0.0000129)	(0.0000130)	(0.0000247)	(0.0000385)	(0.0000370)	(0.0000356)
GF	0.0384^{**}	0.0333^{**}	0.0340^{**}	0.0523^{**}	0.0217	0.0468	0.0560
	(0.0157)	(0.0145)	(0.0145)	(0.0220)	(0.0909)	(0.0857)	(0.0857)
WQA*	0.0312^{**}	0.133^{***}	0.120^{***}	0.00540	0.0121	0.0479	0.0611^{*}
	(0.0149)	(0.0337)	(0.0278)	(0.0293)	(0.0202)	(0.0393)	(0.0357)
N	4775	4716	4716	3536	923	892	892
Test endogeneity		0.0006	0.0004			0.3361	0.1419
Kleinbergen-Paap (F-statistics)		47.043	61.882			32.641	41.037

Fixed effects regressions and IV regressions using station fixed effects. Robust standard errors in parentheses. Year dummies are included but not shown. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

River Basin Organizations (RBOs) are a particular form of agreement that has been proposed as an effective way to manage international river basins. They offer a deeper and more institutionalized form of cooperation. In Section 5.7.4 in the Appendix, I examine pollution levels before and after the signature of the documents that lead to the creation of RBOs. It is shown that RBOs are created in situations when water quality is constant. In a first analysis no effect on water quality can be found. However, the instruments used to analyze WQAs do not perform well enough to apply them as well to RBOs.

5.5.4 Upstream basins

In order to examine whether treaties are effective in reducing free-riding in international rivers, Table 5.6 and Table 5.7 show the estimated effect in upstream stations only. These are stations whose water flows to another country. When using BOD as the dependent variable, WQAs appear to reduce pollution in upstream stations to a much larger degree than in other stations in international river basins. The effect when testing all countries and when testing the EU only is highly significant throughout all regressions. The estimated reduction for all countries is 65%

in model (2) and 50.8% in model (3), which seems very high. In the EU, the estimated reduction is 51.5% in model (6) and 39.2% in model (7). The Kleinbergen-Paap F-statistic indicates that the instrument used in model (6) is weak, while the instruments perform well in all other models.

Table 5.6 IV regressions in upstream basins: BOD

	(1) All countries Fixed effects	(2) All countries IV (20 years)	(3) All countries IV (30 years)	(4) Non-EU Fixed effects	(5) EU Fixed effects	(6) EU IV (20 years)	(7) EU IV (30 years)
lnDischarge	-0.123	-0.113	-0.117	-0.119	-0.193***	-0.174***	-0.185***
	(0.101)	(0.0838)	(0.0815)	(0.113)	(0.0561)	(0.0506)	(0.0471)
lnTemperature	-0.226**	-0.118	-0.166	0.165	0.307	-0.150	-0.0554
	(0.0999)	(0.120)	(0.108)	(0.266)	(0.243)	(0.179)	(0.154)
lnGDPpc	0.685*	1.178***	0.959***	1.082***	-0.466	-0.372	-0.320
	(0.370)	(0.357)	(0.306)	(0.342)	(0.505)	(0.494)	(0.451)
GDPpc2	-0.000438	-0.000174	-0.000292	0.00425*	0.000181	0.000145	0.000145
	(0.000267)	(0.000234)	(0.000191)	(0.00211)	(0.000152)	(0.000273)	(0.000237)
GF	-0.122	-0.305	-0.223	-1.392***	-1.321*	0.00205	-0.230
	(0.210)	(0.235)	(0.214)	(0.291)	(0.716)	(0.520)	(0.439)
WQA*	-0.284***	-1.051***	-0.709***	-0.250	-0.284***	-0.724**	-0.497**
	(0.0739)	(0.307)	(0.202)	(0.292)	(0.0684)	(0.307)	(0.193)
N	421	411	411	202	219	210	210
Test endogeneity		0.0021	0.0262			0.0632	0.1967
Kleinbergen-Paap (F-statistics)		19.399	29.979			7.184	14.767

Fixed effects regressions and IV regressions using station fixed effects. Robust standard errors in parentheses. Year dummies are included but not shown. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The results for DO are presented in Table 5.7. In the IV regressions, the estimated effect of WQAs when considering all countries is increased compared to regression model (1). DO increases significantly in upstream stations after the signature of WQAs. In the IV regressions the estimated effect is between 10.4% and 10.8%. Even in the EU a significant effect can be found, although no effect appeared when all stations in the EU were analyzed. Overall, the results show that also upstream states reduce their pollution as a consequence of treaties. The estimated reduction is even stronger than in other stations. Free-riding in the form of pollution spillovers is, thus, significantly reduced.

Table 5.7 IV regressions in upstream basins: DO

	(1) All countries Fixed effects	(2) All countries IV (20 years)	(3) All countries IV (30 years)	(4) Non-EU Fixed effects	(5) EU Fixed effects	(6) EU IV (20 years)	(7) EU IV (30 years)
lnDischarge	0.00815	0.00765	0.00762	0.0192	-0.0415	-0.0503*	-0.0508*
	(0.0167)	(0.0145)	(0.0145)	(0.0171)	(0.0306)	(0.0277)	(0.0276)
lnTemperature	-0.0583*	-0.0713**	-0.0720**	-0.104**	-0.0107	0.0149	0.0163
	(0.0309)	(0.0358)	(0.0360)	(0.0499)	(0.0574)	(0.0494)	(0.0490)
lnGDPpc	-0.0502	-0.0616*	-0.0630**	0.0448	0.226	0.242	0.243
	(0.0344)	(0.0324)	(0.0304)	(0.0873)	(0.220)	(0.163)	(0.164)
GDPpc2	-0.000026	-0.000018	-0.000017	-0.000043	-0.000071*	-0.000065	-0.000064
	(0.0000206)	(0.0000219)	(0.0000217)	(0.000214)	(0.0000359)	(0.0000559)	(0.0000572)
GF	-0.0373	-0.0290	-0.0286	-0.108	-0.130	-0.139	-0.140
	(0.0380)	(0.0329)	(0.0327)	(0.0794)	(0.117)	(0.107)	(0.109)
WQA*	0.0370*	0.0988**	0.103***	0.00576	0.0177	0.159*	0.167**
	(0.0187)	(0.0430)	(0.0325)	(0.0366)	(0.0215)	(0.0844)	(0.0688)
N	795	783	783	335	234	226	226
Test endogeneity		0.1101	0.0217			0.0501	0.0091
Kleinbergen-Paap (F-statistics)		29.471	48.662			12.063	21.166

Fixed effects regressions and IV regressions using station fixed effects. Robust standard errors in parentheses. Year dummies are included but not shown. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.6 Discussion and conclusion

Despite substantial technological improvements in water treatment, pollution remains a major global problem and a reason for conflicts between states. The signature of treaties is often proposed as an important tool to reduce pollution in international rivers. Still, their success in reducing pollution in international river basins is rarely studied. The scarcity of data on water quality and water temperatures complicates the evaluation. Nevertheless, important insights can be drawn from the existing data. Analyzing the trend before and after the signature of WQAs reveals that these agreements are commonly signed when water quality is worsening compared to the water quality in other river basins. This shows that states consider treaties to be an adequate tool to address the problem. This result also highlights the importance of addressing endogeneity problems when analyzing the effectiveness of treaties in international river basins. Using navigation and border issue treaties as instruments in IV regressions shows that WQAs tend to significantly improve water quality. Pollution is also estimated to be significantly reduced in upstream stations, which indicates that WQAs reduce free-riding in the form of pollution spillovers from upstream states to downstream states.

This study carefully examines the possible endogeneity when estimating the effectiveness of WQAs in reducing pollution in international rivers and presents an instrument that can be used in order to address the problem. However, the study has its limitations. Only a limited number of treaties outside the EU could be analyzed due to missing control variables. The instrument proposed for the IV estimations could not be used to study these countries. The analysis also relies on the strong assumption that first WQAs have the same impact on water quality as subsequent agreements. Furthermore, the high number of treaties analyzed comes at a price of a strong generalization of treaties, regions and times. Water diplomacy is complex and future quantitative and case studies will need to dig deeper into pollution dynamics in international river basins.

Acknowledgements: The work was financially supported by the European Union (EU) Horizon 2020 program, action ERC-2014-STG, Project HUCO, grant number 636746.

5.7 Appendix

5.7.1 Summary statistics

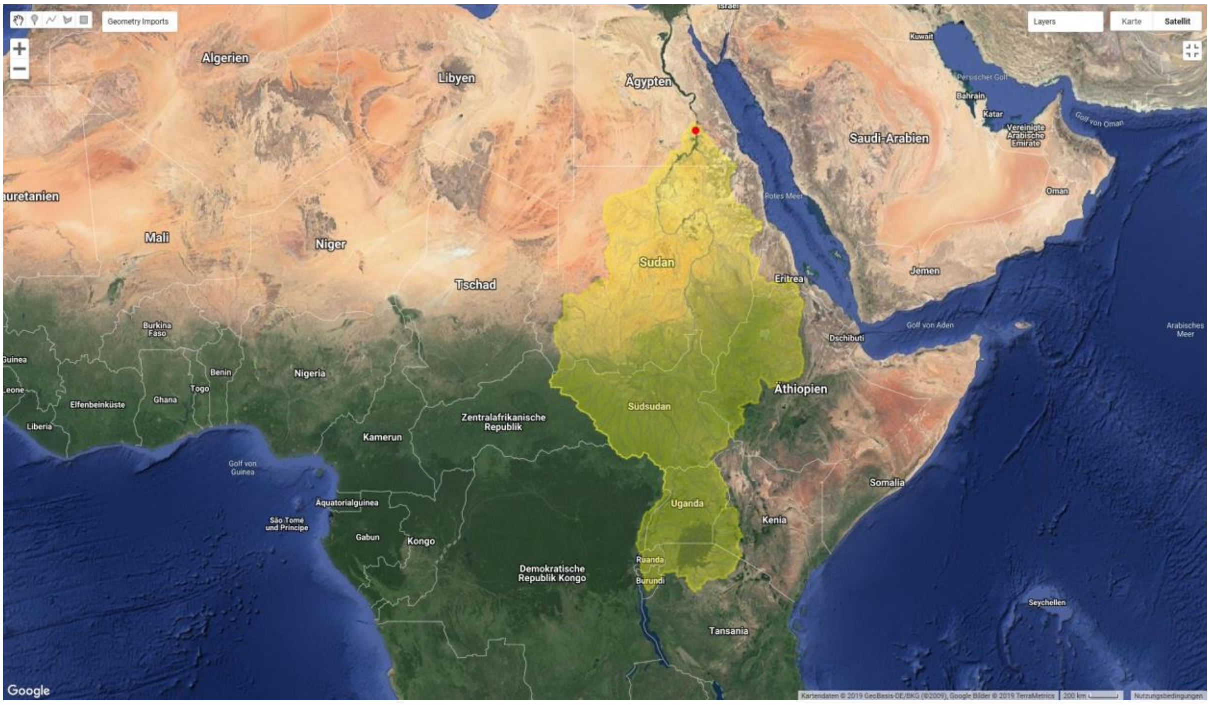
Table 5.8: Summary statistics

	Number of observations	Mean	Standard error	Minimum	Maximum
BOD	3804	2.613955	8.29208	0	368.6667
DO	3774	9.297905	2.403667	0	84.66667
WQA	3804	.1240799	.329716	0	1
Discharge	3804	1301.459	4428.95	.0285	52233.33
Water Temperature	3804	15.19471	6.228371	0	33.5
GDP p.c. (in 1000 US dollar)	3630	12.37282	9.92971	.1948047	52.83125
GF	3804	.2421136	.428419	0	1
Year	3804	1992.911	7.200431	1974	2007
Time invariant variables in between comparison					
Population in upstream area	3790	9214992	2.61e+07	2311.713	4.03e+08
Upstream	3804	.1190852	.3239313	0	1
Downstream	3804	.1650894	.37131	0	1

5.7.2 Between comparison

This section presents pooled cross-sectional regressions. This type of regression allows to compare between stations and, therefore, allows for a comparison between stations that are located in a basin with an existing WQA with stations without WQA. It is important to remember that this method does not allow to estimate the causal effect of WQAs but can serve as a first comparison of river basins with and without WQA. The analysis allows also to assess whether pollution is less of a problem in basins that are regulated by a WQA.
The between comparison also allows to include important time-invariant control variables. The specification of these variables is based on the work of Sigman (2002, 2005). A first variable is the total population living in the upstream catchment area of the station. Figure 5.3 shows the upstream catchment area of Aswan, which is located at the Nile. Google earth engine was used to compute the area. Geospatial data on population was taken from the Center for International Earth Science Information Network (2018). The indicator variable *Upstream* takes the value 1 if the station's water flows to another country. In a similar manner the variable *Downstream* captures whether a river from another country contributes to the station's water. Thus, stations can be considered to be either upstream or downstream, both at the same time or neither. Country indicator variables are included in the regression. This might already capture some of the unobserved factors that impact water quality and might also have an effect on the likelihood to sign an agreement in the specific basin. Since states are the actors deciding on the signature of treaties, it seems likely that factors that impact the formation of agreements are located at country level. Year dummies are used to control for time effects. The standard errors are clustered on station level in order to control for correlated errors within one station.

Figure 5.3 Upstream catchment area of Aswan (Nile)



The results of the cross-sectional regressions are presented in Table 5.9. The logarithm of BOD is chosen as the dependent variable. Two specifications are chosen in order to control for the effect of WQAs. In the regressions (1), (3) and (5), *WQA* simply captures whether any WQA was signed over the river's basin. The other three regressions test whether the number of agreements is related to water quality. All control variables perform as expected. In the regression models (1) to (4), higher discharge reduces pollution, higher temperatures tend to increase BOD and income levels show an inverted U-shape relation with pollution. In the EU, these variables are insignificant or only weakly significant. General water frameworks are only weakly correlated with water quality if at all. The number of people living in the upstream area is highly correlated with BOD. Interestingly, water quality in the upstream area of international rivers is not worse than in other areas of international rivers or in national rivers. In the EU, pollution in upstream areas is even lower. This confirms the results that free-riding is less of a problem in the EU (Sigman 2002). Outside the EU, water quality is better in the downstream areas of international rivers.

Water quality in international rivers is better when a WQA was signed in the basin, as indicated by the coefficient of *WQA*. BOD is around 23.6% lower in basins with a WQA. When considering non-EU and EU countries separately, the effect is only weakly significant for non-EU countries and disappears in EU countries. Also the number of WQAs signed is only significantly related to water quality when considering all countries. Each additional agreement is associated with 8.4% lower levels in BOD. This is an important indication that also additional agreements might impact water quality. The R^2 indicates that approximately more than 60% of the variation can be explained, in the EU even more than 87%. This shows that the control variables are able to capture much of the dynamics that drive water quality.

Table 5.9 Cross-sectional pooled regressions

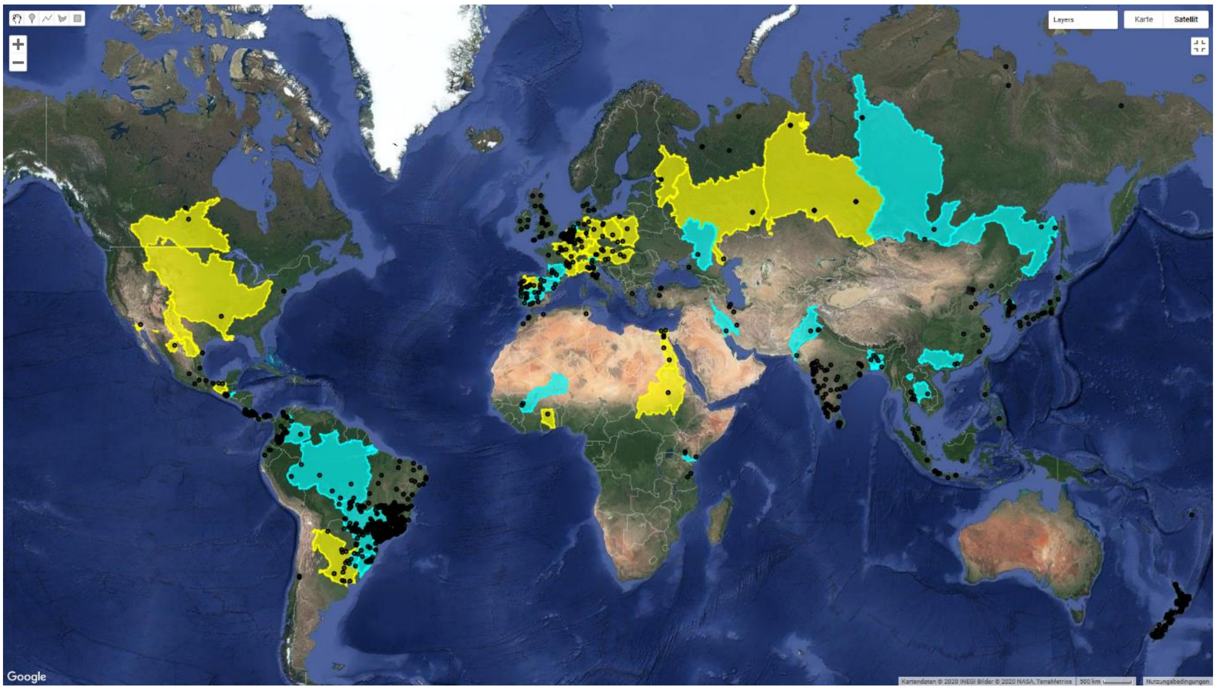
	(1) All countries	(2) All countries	(4) Non-EU	(5) Non-EU	(7) EU	(8) EU
lnDischarge	-0.0473***	-0.0448***	-0.0512***	-0.0508***	-0.0375	-0.0372
	(0.0168)	(0.0171)	(0.0189)	(0.0189)	(0.0274)	(0.0295)
lnTemperature	0.407***	0.382***	0.417***	0.398***	0.343*	0.341*
	(0.106)	(0.105)	(0.119)	(0.117)	(0.194)	(0.194)
lnGDPpc	0.378***	0.412***	0.423***	0.442***	0.455	0.399
	(0.105)	(0.108)	(0.119)	(0.122)	(0.347)	(0.349)
$GDPpc^2$	-0.000263**	-0.000282***	-0.000324**	-0.000330**	-0.000402	-0.000355
	(0.000106)	(0.000107)	(0.000142)	(0.000142)	(0.000281)	(0.000282)
GF	-0.122	-0.158	-0.246*	-0.274*	0.371	0.347
	(0.109)	(0.110)	(0.142)	(0.140)	(0.240)	(0.256)
lnPopulation	0.0803***	0.0792***	0.0721***	0.0717***	0.155***	0.146***
	(0.0199)	(0.0199)	(0.0216)	(0.0216)	(0.0363)	(0.0348)
Upstream	-0.0835	-0.0368	0.147	0.178	-0.526***	-0.534***
	(0.0994)	(0.110)	(0.134)	(0.139)	(0.114)	(0.134)
Downstream	-0.226**	-0.250***	-0.327**	-0.357**	-0.130	-0.141
	(0.0886)	(0.0874)	(0.145)	(0.144)	(0.100)	(0.0983)
WQA	-0.269***		-0.239*		-0.186	
	(0.0849)		(0.143)		(0.117)	
Number of WQAs		-0.0872***		-0.110		-0.0307
		(0.0297)		(0.134)		(0.0374)
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3608	3608	2732	2732	876	876
R^2	0.681	0.681	0.619	0.618	0.873	0.871

OLS regressions. Standard errors are clustered on station level and shown in parentheses. Year and country dummies are included but not shown. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.7.3 Replacing missing values

River discharge is an important control variable. Because pollution is measured as a concentration of pollutants in the water, a change in the discharge volume by definition should change the pollution concentration. Unfortunately, the variable is frequently missing, which reduces the number of observations considerably. In the regressions presented in Table 5.10 and Table 5.11, missing values of discharge are replaced with zero. In addition, an indicator variable is added that takes the value one when discharge was replaced and zero otherwise. This procedure has already been applied by Sigman (2002) and Bernauer and Kuhn (2010), but to a much smaller share of observations. The number of observations more than doubles when missing discharge is replaced. This also increases the number of agreements included in the analysis in particular in non-EU countries. Figure 5.4 presents a map of the stations that enter the regressions and the country-basins with WQA (yellow) and those without WQA (cyan). Future levels of border and navigation treaties are again used as instruments for *WQA**.

Figure 5.4: Stations and basins in the sample when missing discharge is replaced



The results when analyzing all countries jointly change severely. Regression model (1) demonstrates that *WQA** still has a significantly negative sign when treated as an exogenous variable in a fixed effects regression. However, in the IV regression the coefficient remains insignificant. In the regressions that did not replace missing values, the estimated effect of WQA still increased in the IV regressions. It is difficult to distinguish whether the change in the results is driven by the increase in WQAs that are tested or by the replacement of missing values by a random number. The estimates for *WQA** in EU countries resembles those from the regressions in the main paper more closely. The estimated reduction in BOD is also much higher when *WQA** is instrumented. In all of the IV regressions the border and navigation treaties are a relevant instrument.

Table 5.10 IV regressions with missing discharge replaced: BOD

	(1) All countries Fixed effects	(2) All countries IV (20 years)	(3) All countries IV (30 years)	(4) Non-EU Fixed effects	(5) EU Fixed effects	(6) EU IV (20 years)	(7) EU IV (30 years)
lnDischarge	-0.00402	-0.00349	-0.00350	-0.00165	-0.00143	-0.00545	-0.00444
	(0.00966)	(0.00834)	(0.00834)	(0.0107)	(0.0187)	(0.0183)	(0.0182)
Discharge missing (d)	0.0688	0.0735*	0.0735*	0.155**	0.0109	-0.0128	-0.00686
	(0.0532)	(0.0411)	(0.0410)	(0.0679)	(0.0858)	(0.0856)	(0.0850)
lnTemperature	-0.0693	-0.0924*	-0.0922*	-0.0794	-0.0367	-0.0307	-0.0322
	(0.0759)	(0.0549)	(0.0546)	(0.0898)	(0.0870)	(0.0641)	(0.0639)
lnGDPpc	0.0287	0.00461	0.00487	-0.000436	0.668***	0.614***	0.627***
	(0.0559)	(0.0396)	(0.0392)	(0.0679)	(0.194)	(0.161)	(0.158)
GDPpc2	-0.000156***	-0.000177***	-0.000177***	-0.000143*	-0.000606**	-0.000508**	-0.000532***
	(0.0000585)	(0.0000383)	(0.0000379)	(0.0000805)	(0.000302)	(0.000198)	(0.000194)
GF	-0.192***	-0.231***	-0.231***	-0.303***	0.0881	0.0637	0.0698
	(0.0684)	(0.0608)	(0.0602)	(0.0898)	(0.124)	(0.107)	(0.103)
WQA*	-0.209***	0.471	0.463	-0.294*	-0.0822*	-0.322**	-0.262**
	(0.0687)	(0.340)	(0.295)	(0.178)	(0.0470)	(0.149)	(0.125)
N	9463	9356	9356	8053	1410	1378	1378
Test endogeneity		0.0299	0.0120			0.0907	0.1390
Kleinbergen-Paap (F-statistics)		31.326	38.582			32.329	40.405

Fixed effects regressions and IV regressions using station fixed effects. Robust standard errors in parentheses. Year dummies are included but not shown. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The results also change when DO is used as the dependent variable. In the fixed effects regression with all countries *WQA** is only weakly significant. Instead of enhancing the effect instrumenting the variable leads to the disappearance of the significant effect. In the EU, the variable becomes highly significant when instrumented. Thus, the results for DO are almost reversed when missing discharge is replaced. While the results in the main paper suggest that WQAs are only effective when testing all countries, the results of this section suggest that they are only effective in the EU. I consider the results presented in the main paper to be more robust. Replacing missing values with zero might introduce a bias, which is one possible explanation for the change in results.

Table 5.11 IV regressions with missing discharge replaced: DO

	(1) All countries Fixed effects	(2) All countries IV (20 years)	(3) All countries IV (30 years)	(4) Non-EU Fixed effects	(5) EU Fixed effects	(6) EU IV (20 years)	(7) EU IV (30 years)
lnDischarge	-0.00160	-0.00162	-0.00161	-0.00287	0.00116	0.00309	0.00304
	(0.00400)	(0.00334)	(0.00334)	(0.00457)	(0.00985)	(0.00769)	(0.00769)
Discharge missing (d)	-0.0200	-0.0203	-0.0202	-0.0296	-0.00201	0.0116	0.0112
	(0.0209)	(0.0201)	(0.0201)	(0.0248)	(0.0449)	(0.0370)	(0.0369)
lnTemperature	-0.0674***	-0.0661***	-0.0665***	-0.0633***	-0.172***	-0.176***	-0.176***
	(0.0254)	(0.0187)	(0.0187)	(0.0229)	(0.0444)	(0.0378)	(0.0378)
lnGDPpc	-0.0132	-0.0117	-0.0122	0.0290	-0.150	-0.118	-0.119
	(0.0144)	(0.0110)	(0.0110)	(0.0207)	(0.124)	(0.0896)	(0.0884)
GDPpc2	0.0000371**	0.0000389***	0.0000383***	-0.0000114	0.0000282	-0.0000505	-0.0000484
	(0.0000186)	(0.0000108)	(0.0000107)	(0.0000219)	(0.0000656)	(0.0000557)	(0.0000541)
GF	0.0281	0.0316**	0.0306**	0.0369**	-0.356	-0.337***	-0.337***
	(0.0194)	(0.0135)	(0.0131)	(0.0177)	(0.231)	(0.101)	(0.101)
WQA*	0.0296*	-0.0316	-0.0133	0.0135	0.00471	0.156***	0.152***
	(0.0160)	(0.107)	(0.0914)	(0.0328)	(0.0254)	(0.0463)	(0.0421)
N	11762	11619	11619	9661	1474	1443	1443
Test endogeneity		0.337	0.6353			0.0002	0.0001
Kleinbergen-Paap (F-statistics)		42.699	54.992			46.932	59.164

Fixed effects regressions and IV regressions using station fixed effects. Robust standard errors in parentheses. Year dummies are included but not shown. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.7.4 River Basin Organizations

A particular form of agreements is a River Basin Organization (RBO). These organizations provide a more institutionalized framework to manage international water bodies than treaties. RBOs differ from treaties over international rivers in two ways. Firstly, while treaties are always legally binding, RBOs need not to be legally binding but must be considered politically binding by its members (Bernauer 1995 in Schmeier, Gerlak, and Blumstein 2016). Secondly, RBOs are characterized by a stronger degree of institutionalization. Schmeier, Gerlak, and Blumstein (2016) argue that three dimensions of institutionalization need to be fulfilled in order to consider an agreement as an RBO: (1) The cooperation needs to be permanent, thus of long-term nature, (2) an infrastructure, such as regular meetings, must be created and (3) RBOs need to be able to act independently from other actors such as states. RBOs often collect data and share them among their member states. This makes the monitoring of members' action and environmental water quality efforts easier (Schmeier 2015).

In the following, an analysis of water quality and RBOs is provided. The data on RBOs was taken from the International River Basin Organization Database (Schmeier 2013). The issues addressed by the RBOs were also coded in the database. RBOs are included if they deal with water quality as "a functional issue". Most RBOs are established through a treaty. Thus, many RBOs are already recorded in the International Freshwater Treaties Database. The instrument used for *WQA** does not perform well enough as an instrument for RBOs. These organizations tend to be signed later than water quality agreements. Their occurrence coincides, therefore, less with navigation and border treaties. I only analyze the trend before and after the document that creates an RBO is signed.

Table 5.12 presents the estimated trend before and after the signature of RBOs. Pollution worsens only when using BOD as a measure for pollution and when including all countries. In all other models no evidence for an increase in pollution can be found. Pollution measured in DO in non-EU countries even tends to decrease. It appears that RBOs are established in more stable situations than other treaties and that they do not serve as an immediate measure to deal with increasing pollution. Pollution drops significantly after the signature of RBOs, as indicated by the coefficients of *RBO**. At the same time, the interaction term indicates that in the years following the signature pollution increases significantly faster than before. In most models, the positive effect on water quality is not offset within the first years after the signature. Still, the results are ambiguous with regard to the effectiveness of RBOs to reduce pollution. Moreover, the estimated effects might be biased because of endogeneity.

Table 5.12 Fixed effects regression: Trend before and after the signature of RBOs

	(1) All countries	(2) Non-EU	(3) EU	(4) All countries	(5) Non-EU	(6) EU
Dependent variable		lnBOD			lnDO	
Trend	0.0163**	0.00718	0.00977	0.000508	0.0128***	-0.00174
	(0.00712)	(0.0465)	(0.00687)	(0.000801)	(0.00443)	(0.00158)
RBO*	-1.066***	-1.032**	-3.120**	0.0669**	0.164***	0.698**
	(0.285)	(0.514)	(1.474)	(0.0338)	(0.0527)	(0.320)
Trend*RBO*	0.0671***	0.0713	0.268**	-0.00156	-0.0140***	-0.0597**
	(0.0214)	(0.0517)	(0.125)	(0.00269)	(0.00507)	(0.0271)
lnDischarge	-0.00147	0.00538	-0.0585	-0.00917	-0.0108	-0.00646
	(0.0146)	(0.0156)	(0.0374)	(0.0114)	(0.0130)	(0.0144)
lnTemperature	-0.118*	-0.0884	-0.102	-0.0915***	-0.0771***	-0.133***
	(0.0646)	(0.0832)	(0.0989)	(0.0186)	(0.0241)	(0.0416)
lnGDPpc	0.227***	0.274***	0.182	-0.0147	-0.0104	0.0450
	(0.0659)	(0.0766)	(0.217)	(0.0194)	(0.0256)	(0.0630)
$GDPpc^2$	-0.0000587	-0.000144	-0.000332	0.0000593***	0.0000539**	0.00000952
	(0.0000811)	(0.0000916)	(0.000245)	(0.0000172)	(0.0000253)	(0.0000389)
GF	-0.191**	-0.320**	0.0478	0.0347**	0.0491**	-0.0107
	(0.0952)	(0.125)	(0.370)	(0.0157)	(0.0224)	(0.0830)
N	3622	2746	876	4775	3536	923
R2 within	0.203	0.227	0.227	0.0394	0.0502	0.192

Fixed effects regressions. Robust standard errors in parentheses. Year dummies are included but not shown.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6 References

Ahn, T. K., Elinor Ostrom, David Schmidt, Robert Shupp, and James Walker. 2001. "Cooperation in PD Games: Fear, Greed, and History of Play." *Public Choice*, 106(1/2): 137–55.

Alpizar, Francisco, Fredrik Carlsson, and Olof Johansson-Stenman. 2008. "Anonymity, reciprocity, and conformity. Evidence from voluntary contributions to a national park in Costa Rica." *Journal of Public Economics*, 92(5-6): 1047–60.

Alpízar, Francisco, and Peter Martinsson. 2013. "Does It Matter if You Are Observed by Others? Evidence from Donations in the Field*." *The Scandinavian Journal of Economics*, 115(1): 74–83.

Andreoni, James, and Ragan Petrie. 2004. "Public goods experiments without confidentiality. A glimpse into fund-raising." *Journal of Public Economics*, 88(7-8): 1605–23.

Andreoni, James, and Justin M. Rao. 2011. "The power of asking. How communication affects selfishness, empathy, and altruism." *Journal of Public Economics*, 95(7-8): 513–20.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics. An empiricist's companion*. Princeton, NJ: Princeton Univ. Press.

Auerswald, Heike, Carsten Schmidt, Marcel Thum, and Gaute Torsvik. 2018. "Teams in a public goods experiment with punishment." *Journal of Behavioral and Experimental Economics*, 72: 28–39.

Baier, Scott L., and Jeffrey H. Bergstrand. 2007. "Do free trade agreements actually increase members' international trade?" *Journal of International Economics*, 71(1): 72–95.

Barbieri, Katherine, and Omar M. G. Keshk. 2016. Correlates of War Project Set Codebook, Version 4.0. http://correlatesofwar.org.

Barbieri, Katherine, Omar M.G. Keshk, and Brian M. Pollins. 2009. "Trading Data." *Conflict Management and Peace Science*, 26(5): 471–91.

Barrett, Scott. 1994. "Conflict and cooperation in managing international water resources." *World Bank Policy Research Working Paper*, 1303.

Barrett, Scott. 2003. *Environment and statecraft. The strategy of environmental treaty-making*. Oxford: Oxford Univ. Press.

Bateson, Melissa, Daniel Nettle, and Gilbert Roberts. 2006. "Cues of being watched enhance cooperation in a real-world setting." *Biology letters*, 2(3): 412–14.

Beames, Penny, Bernhard Lehner, and Mira Anand. 2019. *Global Reservoir and Dam (GRanD) Database. Version 1.3*.

Beck, Lucas, Thomas Bernauer, and Anna Kalbhenn. 2010. "Environmental, political, and economic determinants of water quality monitoring in Europe." *Water Resources Research*, 46(11): 77.

Bernauer, Thomas. 1995. “The effect of international environmental institutions: how we might learn more.” *International Organization*, 49(2): 351–77.

Bernauer, Thomas, and Tobias Böhmelt. 2020. “International conflict and cooperation over freshwater resources.” *Nature Sustainability*, 569: 215.

Bernauer, Thomas, and Patrick M. Kuhn. 2010. “Is there an environmental version of the Kantian peace? Insights from water pollution in Europe.” *European Journal of International Relations*, 16(1): 77–102.

Biswas, Asit K. 2004. “From Mar del Plata to Kyoto. An analysis of global water policy dialogue.” *Global Environmental Change*, 14: 81–88.

Biswas, Asit K. 2011. “Cooperation or conflict in transboundary water management: case study of South Asia.” *Hydrological Sciences Journal*, 56(4): 662–70.

Bobekova, Elvira. 2015. “Third party conflict management of transboundary river disputes.” *International Journal of Conflict Management*, 26(4): 402–26.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. “hroot. Hamburg Registration and Organization Online Tool.” *European Economic Review*, 71: 117–20.

Bornstein, Gary, and Ilan Yaniv. 1998. “Individual and Group Behavior in the Ultimatum Game: Are Groups More “Rational” Players?” *Experimental Economics*, 1(1): 101–08.

Brochmann, Marit. 2012. “Signing River Treaties—Does It Improve River Cooperation?” *International Interactions*, 38(2): 141–63.

Brochmann, Marit, and Nils P. Gleditsch. 2012a. “Shared River Basin Dataset. Codebook for the Shared River Basin Dataset Version 3.0a and 3.0b.” *PRIO*.

Brochmann, Marit, and Nils P. Gleditsch. 2012b. “Shared rivers and conflict – A reconsideration.” *Political Geography*, 31(8): 519–27.

Brochmann, Marit, and Paul R. Hensel. 2011. “The Effectiveness of Negotiations over International River Claims1.” *International Studies Quarterly*, 55(3): 859–82.

Brosig, Jeannette, Joachim Weimann, and Axel Ockenfels. 2003. “The Effect of Communication Media on Cooperation.” *German Economic Review*, 4(2): 217–41.

Cason, Timothy N., and Vai-Lam Mui. 1997. “A Laboratory Study of Group Polarisation in the Team Dictator Game.” *The Economic Journal*, 107(444): 1465–83.

Center for International Earth Science Information Network. 2018. *Gridded Population of the World, Version 4 (GPWv4). Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11*.

Charness, Gary, and Matthias Sutter. 2012. “Groups Make Better Self-Interested Decisions.” *Journal of Economic Perspectives*, 26(3): 157–76.

Christens, Sven, Astrid Dannenberg, and Florian Sachs. 2019. “Identification of individuals and groups in a public goods experiment.” *Journal of Behavioral and Experimental Economics*.

Coase, Ronald H. 1960. “The Problem of Social Cost.” *The Journal of Law and Economics*(3): 1–44.

Conca, Ken, Fengshi Wu, and Ciqi Mei. 2006. "Global Regime Formation or Complex Institution Building? The Principled Content of International River Agreements." *International Studies Quarterly*, 50(2): 263–85.

Cooley, Heather, and Peter H. Gleick. 2011. "Climate-proofing transboundary water agreements." *Hydrological Sciences Journal*, 56(4): 711–18.

Cox, David. 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society*, 34(2): 187–220.

Daoudy, Marwa. 2009. "Asymmetric Power: Negotiating Water in the Euphrates and Tigris." *International Negotiation*, 14(2): 361–91.

Dawes, R. M. 1980. "Social Dilemmas." *Annual Review of Psychology*, 31(1): 169–93.

Devlin, Colleen, and Cullen S. Hendrix. 2014. "Trends and triggers redux: Climate change, rainfall, and interstate conflict." *Political Geography*, 43: 27–39.

Dinar, Ariel, Brian Blankespoor, Shlomi Dinar, and Pradeep Kurukulasuriya. 2010. "Does precipitation and runoff variability affect treaty cooperation between states sharing international bilateral rivers?" *Ecological Economics*, 69(12): 2568–81.

Dinar, Ariel, Shlomi Dinar, Stephen McCaffrey, and Daene McKinney. 2007. *Bridges Over Water*: WORLD SCIENTIFIC.

Dinar, Shlomi. 2006. "Assessing side-payment and cost-sharing patterns in international water agreements. The geographic and economic connection." *Political Geography*, 25(4): 412–37.

Dinar, Shlomi. 2008. *Routledge studies in the modern world economy*. Vol. 72, *International water treaties. Negotiation and cooperation along transboundary rivers*. London: Routledge.

Dinar, Shlomi. 2009. "Power Asymmetry and Negotiations in International River Basins." *International Negotiation*, 14(2): 329–60.

Dinar, Shlomi, Ariel Dinar, and Pradeep Kurukulasuriya. 2011. "Scarcity and Cooperation along International Rivers. An Empirical Assessment of Bilateral Treaties." *International Studies Quarterly*, 55(3): 809–33.

Dinar, Shlomi, David Katz, Lucia de Stefano, and Brian Blankespoor. 2015. "Climate change, conflict, and cooperation. Global analysis of the effectiveness of international river treaties in addressing water variability." *Political Geography*, 45: 55–66.

Dinar, Shlomi, David Katz, Lucia de Stefano, and Brian Blankespoor. 2019. "Do treaties matter? Climate change, water variability, and cooperation along transboundary river basins." *Political Geography*, 69: 162–72.

Eagly, Alice H. 1987. *Sex Differences in Social Behavior. A Social-role interpretation*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Eagly, Alice H., and Maureen Crowley. 1986. "Gender and helping behavior. A meta-analytic review of the social psychological literature." *Psychological bulletin*, 100(3): 283–308.

Eagly, Alice H., and Wendy Wood. 1991. "Explaining Sex Differences in Social Behavior. A Meta-Analytic Perspective." *Personality and Social Psychology Bulletin*, 17(3): 306–15.

Ekström, Mathias. 2012. "Do watching eyes affect charitable giving? Evidence from a field experiment." *Experimental Economics*, 15(3): 530–46.

Ellingsen, Tore, and Magnus Johannesson. 2008. "Anticipated verbal feedback induces altruistic behavior." *Evolution and Human Behavior*, 29(2): 100–05.

Engel, Christoph. 2011. "Dictator games. A meta study." *Experimental Economics*, 14(4): 583–610.

Ernest-Jones, Max, Daniel Nettle, and Melissa Bateson. 2011. "Effects of eye images on everyday cooperative behavior. A field experiment." *Evolution and Human Behavior*, 32(3): 172–78.

Espey, Molly, and Basman Towfique. 2004. "International bilateral water treaty formation." *Water Resources Research*, 40(5): 519.

Fehr, Ernst, and Ivo Schurtenberger. 2018. "Normative foundations of human cooperation." *Nature human behaviour*, 2(7): 458–68.

Festinger, Leon, A. Pepitone, and T. Newcomb. 1952. "Some Consequences of De-Individuation in a Group." *The Journal of Abnormal and Social Psychology*, 47(2, Suppl): 382–89.

Fischbacher, Urs. 2007. "z-Tree. Zurich toolbox for ready-made economic experiments." *Experimental Economics*, 10(2): 171–78.

Food and Agriculture Organization of the United Nations. 2019. Transboundary Waters. http://www.fao.org/land-water/water/water-management/transboundary-water-management/en/ (accessed July 2019).

Food and Agriculture Organization of the United Nations. 2020. *Wastewater treatment*. http://www.fao.org/3/t0551e/t0551e05.htm (accessed January 29, 2020).

Francey, Damien, and Ralph Bergmüller. 2012. "Images of eyes enhance investments in a real-life public good." *PloS one*, 7(5): e37397.

Frey, Frederick W. 1993. "The Political Context of Conflict and Cooperation Over International River Basins." *Water International*, 18(1): 54–68.

Funk, Patricia. 2010. "Social Incentives and Voter Turnout. Evidence from the Swiss Mail Ballot System." *Journal of the European Economic Association*, 8(5): 1077–103.

Gibler, Douglas M. 2009. *International Military Alliances, 1648-2008*. 2300 N Street, NW, Suite 800, Washington DC 20037 United States: CQ Press.

Gillet, Joris, Arthur Schram, and Joep Sonnemans. 2009. "The tragedy of the commons revisited. The importance of group decision-making." *Journal of Public Economics*, 93(5-6): 785–97.

Giordano, Mark, Alena Drieschova, James A. Duncan, Yoshiko Sayama, Lucia de Stefano, and Aaron T. Wolf. 2014. "A review of the evolution and state of transboundary freshwater treaties." *International Environmental Agreements: Politics, Law and Economics*, 14(3): 245–64. http://www.transboundarywaters.orst.edu.

Giordano, Meredith A. 2003. "Managing the Quality of International Rivers: Global Principles and Basin Practice." *Natural Resources Journal*, 43(1): 111–36.

Global Runoff Data Centre. 2019. *River Discharge Data*. www.bafg.de/GRDC.

Gormley, Todd A., and David A. Matsa. 2014. "Common Errors: How to (and Not to) Control for Unobserved Heterogeneity." *Review of Financial Studies*, 27(2): 617–61.

Gosling, Simon N., and Nigel W. Arnell. 2016. "A global assessment of the impact of climate change on water scarcity." *Climatic Change*, 134(3): 371–85.

Greene, William. 2010. "Testing hypotheses about interaction terms in nonlinear models." *Economics Letters*, 107(2): 291–96.

Greenstone, Michael, and Rema Hanna. 2014. "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India." *American Economic Review*, 104(10): 3038–72.

Greig, J. M., and Andrew J. Enterline. 2017. *National Material Capabilities (NMC) Data Documentation. Version 5.0*.

Grossman, G. M., and A. B. Krueger. 1995. "Economic Growth and the Environment." *The Quarterly Journal of Economics*, 110(2): 353–77.

Gupta, Joyeeta. 2016. "The Watercourses Convention, Hydro-hegemony and Transboundary Water Issues." *The International Spectator*, 51(3): 118–31.

Gupta, Joyeeta, Rhodante Ahlers, and Lawal Ahmed. 2010. "The Human Right to Water: Moving Towards Consensus in a Fragmented World." *Review of European Community & International Environmental Law*, 19(3): 294–305.

Harbaugh, William T. 1998a. "The Prestige Motive for Making Charitable Transfers." *American Economic Review*, 88(2): 277–82.

Harbaugh, William T. 1998b. "What do donations buy?" *Journal of Public Economics*, 67(2): 269–84.

Hauge, Karen, and Ole Rogeberg. 2015. "Representing Others in a Public Good Game." *Games*, 6(3): 381–93.

Hettige, Hemamala, Muthukumara Mani, and David Wheeler. 1998. "Industrial pollution in economic development: Kuznets revisited." *World Bank Policy Research Working Paper*, 1876: 1–36.

Huber, Stephan, Jochen Model, and Silvio Städter. 2019. "Ostracism in Alliances of Teams and Individuals: Voting, Exclusion, Contribution, and Earnings." *Available at SSRN: https://ssrn.com/abstract=2863890*.

Insko, Chester A., Robin L. Pinkley, Rick H. Hoyle, Bret Dalton, Guiyoung Hong, Randa M. Slim, Pat Landry, Brynda Holton, Paulette F. Ruffin, and John Thibaut. 1987. "Individual versus group discontinuity. The role of intergroup contact." *Journal of Experimental Social Psychology*, 23(3): 250–67.

Insko, Chester A., and John Schopler. 1998. "Differential distrust of groups and individuals." In *Intergroup cognition and intergroup behavior*, ed. Constantine Sedikides, John Schopler, and Chester A. Insko, 75–107. New Jersey: Lawrence Erlbaum Associates.

Insko, Chester A., John Schopler, Stephen M. Drigotas, Kenneth A. Graetz, James Kennedy, Chante Cox, and Garry Bornstein. 1993. "The Role of Communication in Interindividual-Intergroup Discontinuity." *Journal of Conflict Resolution*, 37(1): 108–38.

Insko, Chester A., John Schopler, Michael B. Pemberton, Jennifer Wieselquist, Stacy A. McIlraith, David P. Currey, and Lowell Gaertner. 1998. "Long-term outcome maximization and the reduction of interindividual–intergroup discontinuity." *Journal of Personality and Social Psychology*, 75(3): 695–710.

International Conference on Water and the Environment. 1992. *Dublin Statement on Water and Sustainable Development*.

Jenkins, Stephen P. 2005. *Survival Analysis*.

Karlan, Dean, and Margaret A. McConnell. 2014. "Hey look at me. The effect of giving circles on giving." *Journal of Economic Behavior & Organization*, 106: 402–12.

Keiser, David A., and Joseph S. Shapiro. 2019. "Consequences of the Clean Water Act and the Demand for Water Quality*." *The Quarterly Journal of Economics*, 134(1): 349–96.

Keohane, Robert O. 1984. *After hegemony. Cooperation and discord in the world political economy*. Princeton, N.J: Princeton University Press.

Kocher, Martin G., and Matthias Sutter. 2005. "The Decision Maker Matters: Individual Versus Group Behaviour in Experimental Beauty-Contest Games." *The Economic Journal*, 115(500): 200–23.

Kocher, Martin G., and Matthias Sutter. 2007. "Individual versus group behavior and the role of the decision making procedure in gift-exchange experiments." *Empirica*, 34(1): 63–88.

Kogut, Tehila, and Ilana Ritov. 2005a. "The "identified victim" effect. An identified group, or just a single individual?" *Journal of Behavioral Decision Making*, 18(3): 157–67.

Kogut, Tehila, and Ilana Ritov. 2005b. "The singularity effect of identified victims in separate and joint evaluations." *Organizational Behavior and Human Decision Processes*, 97(2): 106–16.

Köppel, Martin, and Detlef F. Sprinz. 2019. "Do Binding Beat Nonbinding Agreements? Regulating International Water Quality." *Journal of Conflict Resolution*, 63(8): 1860–88.

Kraft-Todd, Gordon, Erez Yoeli, Syon Bhanot, and David Rand. 2015. "Promoting cooperation in the field." *Current Opinion in Behavioral Sciences*, 3: 96–101.

Kugler, Tamar, Gary Bornstein, Martin G. Kocher, and Matthias Sutter. 2007. "Trust between individuals and groups. Groups are less trusting than individuals but just as trustworthy." *Journal of Economic Psychology*, 28(6): 646–57.

Kugler, Tamar, Edgar E. Kausel, and Martin G. Kocher. 2012. “Are groups more rational than individuals? A review of interactive decision making in groups.” *Wiley interdisciplinary reviews. Cognitive science*, 3(4): 471–82.

Kurzban, Robert, Kevin McCabe, Vernon L. Smith, and Bart J. Wilson. 2001. “Incremental Commitment and Reciprocity in a Real-Time Public Goods Game.” *Personality and Social Psychology Bulletin*, 27(12): 1662–73.

Lacetera, Nicola, and Mario Macis. 2010. “Social image concerns and prosocial behavior. Field evidence from a nonlinear incentive scheme.” *Journal of Economic Behavior & Organization*, 76(2): 225–37.

Lee, Chien-Chiang, Yi-Bin Chiu, and Chia-Hung Sun. 2010. “The environmental Kuznets curve hypothesis for water pollution: Do regions matter?” *Energy Policy*, 38(1): 12–23.

Lehner, Bernhard, Catherine R. Liermann, Carmen Revenga, Charles Vörösmarty, Balazs Fekete, Philippe Crouzet, Petra Döll, Marcel Endejan, Karen Frenken, and Jun Magome, et al. 2011. “High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management.” *Frontiers in Ecology and the Environment*, 9(9): 494–502.

López-Pérez, Raúl, and Marc Vorsatz. 2010. “On approval and disapproval. Theory and experiments.” *Journal of Economic Psychology*, 31(4): 527–41.

Lowi, Miriam R. 1993. *Water and Power. The Politics of a Scarce Resource in the Jordan River Basin*: Cambridge University Press.

Luhan, Wolfgang J., Martin G. Kocher, and Matthias Sutter. 2009. “Group polarization in the team dictator game reconsidered.” *Experimental Economics*, 12(1): 26–41.

Maddison, Angus. 2010. Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD. http://www.ggdc.net/maddison/.

Mahjouri, Najmeh, and Mojtaba Ardestani. 2010. “A game theoretic approach for interbasin water resources allocation considering the water quality issues.” *Environmental monitoring and assessment*, 167(1-4): 527–44.

Managi, Shunsuke, Akira Hibiki, and Tetsuya Tsurumi. 2009. “Does trade openness improve environmental quality?” *Journal of Environmental Economics and Management*, 58(3): 346–63.

Marshall, Monty G., Ted R. Gurr, and Keith Jaggers. 2018. “Political Regime Characteristics and Transitions, 1800-2017. Dataset Users’ Manual.” *Center for Systemic Peace*.

Masclet, David, Charles Noussair, Steven Tucker, and Marie-Claire Villeval. 2003. “Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism.” *American Economic Review*, 93(1): 366–80.

McCracken, Melissa, and Aaron T. Wolf. 2019. “Updating the Register of International River Basins of the world.” *International Journal of Water Resources Development*, 35(5): 732–82.

McDonald, John F., and Robert A. Moffitt. 1980. "The Uses of Tobit Analysis." *The Review of Economics and Statistics*, 62(2): 318–21.

Mekonnen, M. M., and A. Y. Hoekstra. 2012. "The blue water footprint of electricity from hydropower." *Hydrology and Earth System Sciences*, 16(1): 179–87.

Milner, Helen. 1992. "International Theories of Cooperation Among Nations: Strengths and Weaknesses." *World Politics*, 44(3): 466–96.

Mitchell, Ronald B., and Patricia M. Keilbach. 2001. "Situation Structure and Institutional Design. Reciprocity, Coercion, and Exchange." *International Organization*, 55(4): 891–917.

Mitchell, Sara M., and Neda A. Zawahri. 2015. "The effectiveness of treaty design in addressing water disputes." *Journal of Peace Research*, 52(2): 187–200.

Müller, Wieland, and Fangfang Tan. 2013. "Who acts more like a game theorist? Group and individual play in a sequential market game and the effect of the time horizon." *Games and Economic Behavior*, 82: 658–74.

Munia, H., J. H. A. Guillaume, N. Mirumachi, M. Porkka, Y. Wada, and M. Kummu. 2016. "Water stress in global transboundary river basins: significance of upstream water use on downstream stress." *Environmental Research Letters*, 11(1): 14002.

Nielsen, Kirby, Puja Bhattacharya, John H. Kagel, and Arjun Sengupta. 2019. "Teams promise but do not deliver." *Games and Economic Behavior*, 117: 420–32.

Nyborg, Karine, John M. Anderies, Astrid Dannenberg, Therese Lindahl, Caroline Schill, Maja Schlüter, W. N. Adger, Kenneth J. Arrow, Scott Barrett, and Stephen Carpenter, et al. 2016. "Social norms as solutions." *Science (New York, N.Y.)*, 354(6308): 42–43.

Olmstead, Sheila M. 2010. "The Economics of Water Quality." *Review of Environmental Economics and Policy*, 4(1): 44–62.

Olmstead, Sheila M., and Hilary Sigman. 2015. "Damming the Commons. An Empirical Analysis of International Cooperation and Conflict in Dam Location." *Journal of the Association of Environmental and Resource Economists*, 2(4): 497–526.

Ostrom, Elinor. 1990. *Governing the commons. The evolution of institutions for collective action*. Cambridge: Cambridge Univ. Press.

Ostrom, Elinor. 2009. "A general framework for analyzing sustainability of social-ecological systems." *Science (New York, N.Y.)*, 325(5939): 419–22.

Ovodenko, Alexander. 2016. "Regional Water Cooperation." *Journal of Conflict Resolution*, 60(6): 1071–98.

Peeters, Ronald, and Marc Vorsatz. 2013. "Immaterial Rewards and Sanctions in a Voluntary Contribution Experiment." *Economic Inquiry*, 51(2): 1442–56.

Petersen-Perlman, Jacob D., and Itay Fischhendler. 2018. "The weakness of the strong: re-examining power in transboundary water dynamics." *International Environmental Agreements: Politics, Law and Economics*, 18(2): 275–94.

Powell, Kate L., Gilbert Roberts, and Daniel Nettle. 2012. "Eye Images Increase Charitable Donations. Evidence From an Opportunistic Field Experiment in a Supermarket." *Ethology*, 118(11): 1096–101.

Rand, David G., Erez Yoeli, and Moshe Hoffman. 2014. "Harnessing Reciprocity to Promote Cooperation and the Provisioning of Public Goods." *Policy Insights from the Behavioral and Brain Sciences*, 1(1): 263–69.

Rege, Mari, and Kjetil Telle. 2004. "The impact of social approval and framing on cooperation in public good situations." *Journal of Public Economics*, 88(7-8): 1625–44.

Richter, Brian D., Sandra Postel, Carmen Revenga, Thayer Scudder, Bernhard Lehner, Allegra Churchill, and Morgan Chow. 2010. "Lost in Development's Shadow. The Downstream Human Consequences of Dams." *Water Alternatives*, 3(2): 14–42.

Rockenbach, Bettina, Abdolkarim Sadrieh, and Barbara Mathauschek. 2007. "Teams take the better risks." *Journal of Economic Behavior & Organization*, 63(3): 412–22.

Rogers, Peter. 1993. "The value of cooperation in resolving international river basin disputes." *Natural Resources Forum*, 17(2): 117–31.

Sadoff, C. 2002. "Beyond the river: the benefits of cooperation on international rivers." *Water Policy*, 4(5): 389–403.

Samek, Anya, and Roman M. Sheremeta. 2014. "Recognizing contributors. An experiment on public goods." *Experimental Economics*, 17(4): 673–90.

Samek, Anya, and Roman M. Sheremeta. 2016. "When Identifying Contributors is Costly. An Experiment on Public Goods." *Southern Economic Journal*, 82(3): 801–08.

Schlosser, C. A., Kenneth Strzepek, Xiang Gao, Charles Fant, Élodie Blanc, Sergey Paltsev, Henry Jacoby, John Reilly, and Arthur Gueneau. 2014. "The future of global water stress: An integrated assessment." *Earth's Future*, 2(8): 341–61.

Schmeier, Susanne. 2013. *The Institutional Design of River Basin Organizations. Introducing the RBO Institutional Design Databaseand its main Findings. Transboundary Freshwater Dispute Database*.

Schmeier, Susanne. 2015. "The institutional design of river basin organizations – empirical findings from around the world." *International Journal of River Basin Management*, 13(1): 51–72.

Schmeier, Susanne, Andrea K. Gerlak, and Sabine Blumstein. 2016. "Clearing the muddy waters of shared watercourses governance: conceptualizing international River Basin Organizations." *International Environmental Agreements: Politics, Law and Economics*, 16(4): 597–619.

Schopler, John, Chester A. Insko, Stephen M. Drigotas, Jennifer Wieselquist, Michael B. Pemberton, and Chante Cox. 1995. "The Role of Identifiability in the Reduction of Interindividual-Intergroup Discontinuity." *Journal of Experimental Social Psychology*, 31(6): 553–74.

Schwarzenbach, René P., Thomas Egli, Thomas B. Hofstetter, Urs von Gunten, and Bernhard Wehrli. 2010. "Global Water Pollution and Human Health." *Annual Review of Environment and Resources*, 35(1): 109–36.

Shepherd, Lee, Russell Spears, and Antony S. R. Manstead. 2013. "'This will bring shame on our nation': The role of anticipated group-based emotions on collective action." *Journal of Experimental Social Psychology*, 49(1): 42–57.

Sigman, Hilary. 2002. "International Spillovers and Water Quality in Rivers: Do Countries Free Ride?" *American Economic Review*, 92(4): 1152–59.

Sigman, Hilary. 2003. "Does Trade Promote Environmental Coordination?: Pollution in International Rivers." *The B.E. Journal of Economic Analysis & Policy*(3): 1–26.

Sigman, Hilary. 2004. "Does Trade Promote Environmental Coordination? Pollution in International Rivers." *Economic Analysis & Policy*, 3.

Sigman, Hilary. 2005. "Transboundary spillovers and decentralization of environmental policies." *Journal of Environmental Economics and Management*, 50(1): 82–101.

Singer, David, Stuart Bremer, and John Stuckey. 1972. "Capability distribution, uncertainty, and major power war, 1820-1965." In *International studies quarterly, Peace, war and numbers*, ed. Bruce M. Russett, and Betty C. Hanson. Beverly Hills, Calif, London: Sage Publ.

Small, Deborah A., and George Loewenstein. 2003. "Helping a Victim or Helping the Victim: Altruism and Identifiability." *Journal of Risk and Uncertainty*, 26(1): 5–16.

Small, Deborah A., and George Loewenstein. 2005. "The devil you know. The effects of identifiability on punishment." *Journal of Behavioral Decision Making*, 18(5): 311–18.

Soetevent, Adriaan R. 2005. "Anonymity in giving in a natural context—a field experiment in 30 churches." *Journal of Public Economics*, 89(11-12): 2301–23.

Song, Jennifer, and Dale Whittington. 2004. "Why have some countries on international rivers been successful negotiating treaties? A global perspective." *Water Resources Research*, 40(5): 317.

Srinivasan, Jeena T., and V. R. Reddy. 2009. "Impact of irrigation water quality on human health: A case study in India." *Ecological Economics*, 68(11): 2800–07.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65(3): 557.

Tir, Jaroslav, and John T. Ackerman. 2009. "Politics of Formalized River Cooperation." *Journal of Peace Research*, 46(5): 623–40.

Tir, Jaroslav, and Douglas M. Stinnett. 2012. "Weathering climate change: Can institutions mitigate international water conflict?" *Journal of Peace Research*, 49(1): 211–25.

Toset, Hans P. W., Nils P. Gleditsch, and Håvard Hegre. 2000. "Shared rivers and interstate conflict." *Political Geography*, 19(8): 971–96.

Transboundary Freshwater Dispute Database. 2018. *International Freshwater Treaties Database. 2018 update*.

United Nations. 1973. *Report of the United Nations Conference on the Human Environment*. Stockholm 1972.

United Nations. 1977. *Report of the United Nations Water Conference*. Mar del Plata.

United Nations. 1997. *United Nations Convention on the Law of the Non-Navigational Uses of International Watercourses*.

United Nations. 2012. Vol. 4. *Managing Water under Uncertainty and Risk. The United Nations World Water Development Report*(1).

United Nations. 2016. *The United Nations World Water Development Report 2016*.

United Nations. 2019. *Global Environment Monitoring System Water Programme for freshwater*. gemstat.org.

United Nations. 2020. *The 2030 Agenda for Sustainable Development*. https://sustainabledevelopment.un.org/ (accessed January 29, 2020).

United Nations Environment Programme. 2015. "Transboundary River Basins: Status and Future Trends." *UNEP, Nairobi*.

Vollstädt, Ulrike, and Robert Böhm. 2019. "Are groups more competitive, more selfish-rational or more prosocial bargainers?" *Journal of Behavioral and Experimental Economics*, 78: 146–59.

Warner, Jeroen, and Neda Zawahri. 2012. "Hegemony and asymmetry. Multiple-chessboard games on transboundary rivers." *International Environmental Agreements: Politics, Law and Economics*, 12(3): 215–29.

Wen, Yingrong, Gerrit Schoups, and Nick van de Giesen. 2017. "Organic pollution of rivers: Combined threats of urbanization, livestock farming and global climate change." *Scientific reports*, 7: 43289.

Wildschut, Tim, Brad Pinter, Jack L. Vevea, Chester A. Insko, and John Schopler. 2003. "Beyond the group mind: a quantitative review of the interindividual-intergroup discontinuity effect." *Psychological bulletin*, 129(5): 698–722.

Wooldridge, Jeffrey M. 2002. *Introductory econometrics. A modern approach*. 2nd ed. Mason, Ohio: Thomson South-Western.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. 2nd ed. Cambridge, Mass.: MIT Press.

World Bank. 2019. *World Bank national accounts data*.

Yoeli, Erez, Moshe Hoffman, David G. Rand, and Martin A. Nowak. 2013. "Powering up with indirect reciprocity in a large-scale field experiment." *Proceedings of the National Academy of Sciences of the United States of America*, 110 Suppl 2: 10424–29.

Yoffe, Shim, Aaron T. Wolf, and Mark Giordano. 2003. "Conflict and Cooperation Over International Freshwater Resources. Indicators of Basins at Risk." *Journal of the American Water Resources Association*, 39(5): 1109–26.

Yoffe, Shira, Greg Fiske, Mark Giordano, Meredith Giordano, Kelli Larson, Kerstin Stahl, and Aaron T. Wolf. 2004. "Geography of international water conflict and cooperation: Data sets and applications." *Water Resources Research*, 40(5): 12.

Zawahri, Neda A., Ariel Dinar, and Getachew Nigatu. 2016. "Governing international freshwater resources: an analysis of treaty design." *International Environmental Agreements: Politics, Law and Economics*, 16(2): 307–31.

Zawahri, Neda A., and Sara M. Mitchell. 2011. "Fragmented Governance of International Rivers. Negotiating Bilateral versus Multilateral Treaties." *International Studies Quarterly*, 55(3): 835–58.

Zimbardo, Philip G. 1969. "The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos." In *Nebraska Symposium on Motivation*, ed. W. D. Arnold, and D. Levine, 237–307.

I herewith give assurance that I completed this dissertation independently without prohibited assistance of third parties or aids other than those identified in this dissertation. All passages that are drawn from published or unpublished writings, either word-for-word or in paraphrase, have been clearly identified as such. Third parties were not involved in the drafting of the content of this dissertation; most specifically I did not employ the assistance of a dissertation advisor. No part of this thesis has been used in another doctoral or tenure process.

____