The Influence of Social Information on Norms of Cooperation

Dissertation

zur Erlangung des akademischen Grades
doctor rerum politicarum

an der Fakultät für Wirtschafts- und Sozialwissenschaften
der Ruprecht-Karls-Universität Heidelberg

vorgelegt von
Tillmann Christian Eymess
geboren am 13. Oktober 1990 in Göttingen

im April 2021

The Influence of Social Information on Norms of Cooperation

Tillmann Eymess

April 2021

Acknowledgments

I am thankful to my doctoral supervisor Florian Diekert. His research and his vision for a joint project laid the groundwork for a productive collaboration, not only in the role of student and supervisor, but also as colleagues and co-authors. With his unconditional availability for feedback and the encouragement to independently explore own research ideas, Florian provided far more than the necessary support for working on this dissertation.

The research in my dissertation rests on two waves of data collection in the Lake Victoria region of Tanzania. Both research trips greatly benefited from extraordinary work by Joseph Luomba and Florian Diekert, who helped to assemble great teams of motivated research assistants. Those include Halima Adam, Julieth Tibanywana, Salma Emmanuel, and Elizabeth Mlahagwa. I am greatly thankful for their help and the ever welcoming atmosphere they created. I further express my gratitude to the Tanzania Fisheries Research Institute and the University of Dar es Salaam (in particular Razack Lokina) for their collaboration and gratefully acknowledge funding from the European Research Council through the "NATCOOP" project (ERC-StGr 678049).

To many colleagues, companions, and friends before and during my doctoral studies, I owe a debt of gratitude. I am thankful to Susan Steiner and Friederike Lenel for introducing me to field research, Sanam Vardag and the Heidelberg Center for the Environment for putting me in front of an interdisciplinary audience, Robbert Schaap for our everyday office antics, and Anca Balietti for jointly preparing the next step of my career.

Lastly, there is a special group of people that has absorbed a lot of unseen stress and helped me keep a healthy balance. Among all of them, my deepest gratitude belongs to Laura.

Tillmann Eymess
Heidelberg, April 30th 2021

Table of Contents

Introduction

The provision of information on the behavior and attitudes of others is an effective tool to induce behavioral change. Practitioners and researchers alike use social information to dissuade individuals from harmful practices or nudge them towards socially desirable actions (Miller and Prentice, 2016). For example, if the aim is to reduce a household's energy consumption, information on the low energy use of one's neighbors is a powerful incentive to save electricity (Allcott, 2011; Allcott and Kessler, 2019). Likewise, information on the generosity of others increases charitable donations (Croson et al., 2009; Goeschl et al., 2018) and information on previous election turnout can induce people to vote (Gerber and Rogers, 2009). So far, economists have focused on the use of social information in situations where agents do not directly interact. But the scant evidence on the effectiveness of social information in strategic interactions is surprising. After all, social norms exist to govern social behavior (Binmore, 2010). In situations where selfish incentives are at odds with what is socially beneficial, social norms may be utilized to induce behavioral change.

This dissertation studies the effect of social information on cooperative behavior in a social dilemma. Especially when formal regulations are either ineffective, dysfunctional, or missing, social norms may help to reduce behavior considered harmful to the collective. A specific context of relevance is the use of natural resources in developing countries where issues of environmental pollution, deforestation, or overfishing are often poorly addressed due to limited state capacity, lack of political will, or corruption (Ostrom, 2008). Without effective formal regulations, the mutual cooperation necessary to facilitate sustainable resource use has to be established through self-management and voluntary efforts. Here, policies may use social information to make an impact. If interventions establish and maintain a social norm of cooperation, they can help to combat the overexploitation of natural resources.

The empirical work in this dissertation focuses on the behavior of actual resource users in two economic lab-in-the-field experiments. The data was collected with participants from the Lake Victoria fisheries in Tanzania, a socio-ecological setting that is prototypical for a collective action problem of common-pool resource use under weak formal institutions. De jure, the fisheries are managed through official regulations

that are monitored and enforced by elected community representatives. De facto, violations are common and co-management structures are undermined by issues of corruption and a lack of trust in the devolution of governmental power (Nunan et al., 2018; Etiegni et al., 2020). Therefore, the Lake Victoria fisheries are a particularly good setting to study the influence of social information on norms of cooperation. Social norms can introduce a true form of bottom-up governance that gives resource users much needed agency in natural resource management.

Studies on the effect of social information should be considered as part of the broad literature on approaches that try to induce, enhance, and maintain cooperation for the management of public goods. These generally rest on the universal finding that humans are not strictly self-interested decision makers, but have other-regarding preferences: they tend to cooperate and generate shared benefits (Vollan and Ostrom, 2010; Cooper and Kagel, 2016). In particular, Velez et al. (2009) show that actual resource users are not only driven by their self-interest but that their behavior in economic experiments is well-explained by preferences for conformity. Furthermore, cooperative resource management may be facilitated through the sanctioning of free-riders (Ostrom et al., 1992; Fehr and Gächter, 2000) or the possibility for direct communication (Ostrom and Walker, 1991; Charness and Dufwenberg, 2006), and depends on several situational variables such as the quality of leadership (Gutiérrez et al., 2011) or temporal and spatial dynamics (Janssen et al., 2010).

Experiments on cooperative behavior have long recognized the need to leave the abstract frame of laboratory setups with western student populations (Henrich et al., 2010). Instead, lab-in-the-field and framed field experiments look towards studying social dilemmas in more realistic settings, gradually lifting the assumption that agents only act in their self-interest. Thereby, they have introduced and developed a behavioral theory of human action that is equipped to explain more nuanced aspects of social interactions (Vollan and Ostrom, 2010; Anderies et al., 2011). One main finding is that behavior is influenced by one's peers. In fact, peers can affect behavior by simply being present during the decision making process such that the decision itself is no longer anonymous (Alpizar et al., 2008; Lopez et al., 2012). Among the models that can explain such peer effects are norm-based approaches, *i.e.*, theories that postulate that own actions are influenced by others' behavior and beliefs. In other words, behavior follows social norms. It is exactly this relationship between own actions and others' behavior and beliefs where social information interventions (like the ones studied in this dissertation) try to induce behavioral change.

The dissertation begins with an exercise on identifying the mechanism through

which social information induces a behavioral change by *"Dissecting Social Norms of Cooperation"* in chapter one. The second chapter, *"The Creation of Social Norms under Weak Institutions"*, studies under which conditions a social information treatment is able to create and maintain a social norm. The first two chapters show that social information is only effective when combined with a mechanism of social enforcement and exerts more influence on agents that are in close social proximity to their peers. In chapter three, *"Changing Collective Action"*, the effect of social information is tested when decisions are made by teams. It addresses the fact that many social dilemmas such as community-based resource management need to be solved by groups of individuals that make joint decisions. Chapter three ends with an intriguing secondary result: the effect of social information is driven by agents with leadership experience. Chapter four, *"Captains of Change"*, tests the robustness of this relationship. The study establishes that it is indeed resource users in leadership positions that should be the target of social information interventions that try to induce a pro-social change in natural resource management.

In sum, the dissertation contributes to the literature by studying the use of social information in an explicitly social context: cooperation behavior in a social dilemma. It shows under which conditions social norms may become policy tools and establishes that social information not only works with individual actors but also when decisions are made by teams.

Social Norms

The use of social information rests on the assumption that social norms may be created, activated, or changed to induce behavioral change. Although they are the underlying subject in each of the four chapters, none of them has room for an in-depth discussion on social norms. To guide the reader's thinking, I therefore address the following questions. How are social norms defined, how can they be operationalized for empirical research, and why is social information able to leverage social norms to induce behavioral change?

There is no consensus on the definition of a social norm. Beginning with Elster (1989) and Cialdini et al. (1990), economists have adopted distinctions within the broader term of "norms" from the literature in social psychology. Elster (1989, p.99) distinguishes between moral norms and social norms: "For norms to be social, they must be shared by other people and partly sustained by their approval and disapproval". Elster's distinction describes two important characteristics of social norms.

First, while moral norms guide personal notions of behavior, social norms exist to regulate morality in groups. Second, social norms are enforced. They govern social interactions and transgressions are punished to maintain that norms are followed. To better describe the influence of norms on behavior, Cialdini et al. (1990) distinguish between injunctive and descriptive norms. Injunctive norms describe what one ought to do and thereby encompasses one's own moral stance as well as a perception of what is approved by others. Descriptive norms describe what most people do.

These conceptualizations shed light on why social norms motivate behavior. For one, descriptive norms are evidence of common behavior or even a group consensus. They imply correctness and provide an indication on how to act in a given situation (Cialdini and Trost, 1998). Further, injunctive norms give behavioral guidance by implying what behavior is met with social approval, and crucially, which actions are met with disapproval and potentially entail further social sanctions. Commonly, but not necessarily, descriptive and injunctive norms overlap. What most people do is a good sign about which action will be approved of and if actions are approved, they are likely to be common behavior (Cialdini et al., 1990). Hence, social norms describe both a shared morality and a shared behavioral practice. They are equilibrium selection devices for coordination problems, especially when rationality provides limited guidance or incentives are in conflict (Binmore, 2010). Social norms are therefore well-suited to solve a social dilemma.

Despite of their importance for decision making, social norms are difficult to operationalize for empirical research. Most definitions do not give an indication on how to measure social norms or how to identify their influence on behavior (Bicchieri and Dimant, 2019). Instead, economists use the terms "norm" and "social norm" to describe many behavioral phenomena, in particular when a comparison to peer behavior is possible. But without an explicit measurement, social norms cannot be identified as causal drivers of behavior (Fehr and Schurtenberger, 2018). Bicchieri (2006, 2017) formulates a definition of social norms that illustrates a possible operationalization. She defines a social norm as a rule of behavior that is followed conditional on two social beliefs. Agents have to believe that (i) others in their reference network follow the norm (empirical expectations) and that (ii) others in their reference network consider the prescribed behavior to be appropriate (normative expectations). Her definition implies that social norms can only be identified when both behavior is observed and when social beliefs are measured (Bicchieri et al., 2018).

Bicchieri's definition is helpful to understand why social norms can be an effective policy tool. The underlying idea that since the behavior that follows social norms

is conditional on social beliefs, it is enough to change those beliefs. In other words, norms can be manipulated (Elster, 1989) and the manipulation can be measured. By providing information about the behavior of others or about what others consider appropriate, social information aims to change either empirical or normative expectations (or both), *i.e.*, it attempts to change the perceived social norm. If the intervention is successful in doing so and actions are indeed conditional on social beliefs, behavior will change. Yet, it is not always necessary to *change* social norms. As social norms are not always at work when agents make decisions (Cialdini et al., 1990), it is sometimes enough to *activate* them by making them salient. And if the behavior in question does not yet underlie any form of shared understanding, social information may *create* social norms. The mechanism is the same for any of these forms: social information prompts the desire to conform.

So far, I have purposely avoided the term *conformity*. It deserves separate attention. In studies on social norms, the action that is in line with what is viewed as normative behavior is usually termed conformity but a precise definition is rarely given. Attempts to define conformity include seminal contributions by Bernheim (1994) and Cialdini and Goldstein (2004). Both view conformity as the adjustment of behavior to match the responses of others. The concept is useful, especially when assessing the impact of an interventions that is designed to change behavior by changing norms. If an intervention induces an adjustment of behavior towards the norm, conformity describes the success of the treatment. Yet, the adjustment concept limits conformity to a somewhat narrow aspect. It misses those that follow norms even without additional influence. For them, conformity does not involve the adjustment of behavior.

A number of theoretical contributions formalize conformity to explain why norms are followed, see Sugden (2000), Bénabou and Tirole (2006), Kimbrough and Vostroknutov (2016), or Michaeli and Spiro (2017). Generally, these models pose that an agent's utility depends on (i) the material gain from own behavior and (ii) a discomfort that is increasing in the mismatch between own actions and the social norm (*i.e.*, others' actions and normative beliefs). Thereby, they formalize the intuition that nonconformity may cause disutility. Several motivations can trigger a response that avoids this disutility. A universal one is a preference for conformity as such. That is, one

may have a preference that own behavior matches the behavior of others.[1] Further motivations depend on the behavioral context. For example, in a social dilemma, preferences for fairness or trust can guide normative considerations and dictate which behavior is necessary to avoid the disutility of non-conformity.

An additional motivation for conformity with social norms is the threat of social sanctions. In the case of norm transgression, others are feared to show resentment and respond with ostracism (Bénabou and Tirole, 2011) or various other forms of social or institutionalized punishment, see Engl et al. (2020) for a review. Sanctions represent looming consequences of violating the norm and can be modeled as a weight on the perceived discomfort when own actions deviate from the normative prescription. Oftentimes, the mere threat of publicly exposing non-conformity is enough to ensure that norms are followed (Cardenas, 2011; Lopez et al., 2012). But without such enforcement mechanisms, there are no consequences to violation other than the internal guilt of skirting morality (Elster, 1989).

One last aspect of conformity with social norms that is useful to the reader is the importance of a reference network highlighted in the earlier mentioned definition by Bicchieri (2006, 2017). The intuition is simple: social norms are more important for own actions when it regulates a social interaction among friends or family rather than an interaction among strangers. Other commonly important reference networks are built on shared ethnicity, gender, or religious orientation. Again, the discomfort of non-conformity is at work as the social impact of one's actions varies with the importance or the strength of the reference network (Latané, 1981). Hence, the operationalization of social norms through the provision of social information is tied to concepts such as social cohesion (Cialdini and Trost, 1998) and an individual's social proximity to one's peers (Bicchieri et al., 2019; Dimant, 2019).

In empirical research, behavioral experiments are a powerful tool to test the effect of social information. The standard approach to assess the success of an intervention is the evaluation of behavioral differences between control and treatment group or between two different social information messages.[2] Studies may then focus on the

[1] Related fields of research address why agents have such a preference for conformity. Influential work by Tajfel and Turner (1979), and Akerlof and Kranton (2000) leads a discussion on the effect of group affiliation and social identity on decision making. They argue that people categorize themselves into social categories, identify themselves with certain social groups, and follow the behavior that is prescribed to this social identity. This idea relates to the concept of in-group favoritism and out-group hostility (Goette et al., 2006; Meier et al., 2016).

[2] Allcott and Kessler (2019) propose an alternative approach and assess the effectiveness of a social information intervention in terms of social welfare, *i.e.*, how much participants are willing to pay for a nudge.

evolution of behavioral differences over time or study conditional treatment effects with respect to additional treatment conditions or participant characteristics. Also, insights into the mechanism through which social information affects behavior may be gained by studying the underlying structure of moral and social beliefs. It is therefore useful to combine the provision of social information with an incentivized belief elicitation.

Synopsis

The dissertation includes four papers, presented in the order in which they were written. Read in sequence, the progress of research along the central theme of providing social information to induce cooperative behavior becomes apparent. While the general topic is the same, each paper is a self-contained study such that the dissertation will be repetitive at certain points. In the following, I will describe how each chapter ties into the next. The first three chapters are manuscripts in the review process for publication while the fourth chapter is in preparation to enter a submission process. Replication files for the analyses in all chapters and supplementary material is available under `https://doi.org/10.11588/data/AB9OAL`.

Chapter 1, *"Dissecting Social Norms of Cooperation: A Conditional Process Analysis"*, studies the mechanism of how the combination of social information message and a social sanctioning institution induces a behavioral change. The experiment is a three-person prisoner's dilemma with fishermen from Lake Victoria, Tanzania that was primarily designed to test the research question addressed in chapter two. The experiment varies (i) whether participants receive low or high social information and (ii) whether they have the possibility to weakly sanction others' choices.

The study has two main objectives. First, it tests whether the transmission of the effect of social information through a change in the perceived social norm depends on a sanctioning institution. Results suggest that a sanctioning institution is indeed necessary for social information to initiate meaningful behavioral change. Second, the paper motivates and utilizes a conditional process analysis to test for causal and conditional effects of social information on cooperation. It thereby showcases the importance of exploring causal paths via the inclusion of contextual variables in the framework of social information, norm conformity, and behavioral change.

Chapter 2, *"The Creation of Social Norms under Weak Institutions"*, co-authored with Florian Diekert, Joseph Luomba, and Israel Waichman, then studies the repeated

form of the prisoner's dilemma with fishermen at Lake Victoria. The study finds that the provision of social information is able to *create* norms of cooperation in the experiment, defined as a stable pattern of behavior that is supported by a consistent set of social beliefs. Results support that the effect of social information is conditional on a social sanctioning institution. That is, the interventions fails without the threat of sanctions.

The paper formalizes a simple model of conformity to illustrate that the results are well explained by norm-based theories. In particular, the social norm differs between low (non-cooperative) and high (cooperative) social information suggesting that agents prefer to conform with what they expect others to do. Furthermore, the study uses random treatment assignment across individual characteristics to examine an agent's social proximity to others as an additional social influence for conformity. The social proximity measure includes information on defining features of the social structure in fishing communities and is a significant predictor of the results. Those with close social proximity to the others in their session drive the social information treatment effect.

Chapter 3, *"Changing Collective Action: Social Information Increases Cooperation of Teams in a Prisoner's Dilemma"*, co-authored with Florian Diekert, rethinks the use of social information as a solution to a social dilemma. Motivated by the observation that teams are the key actors in many social dilemmas, the paper poses the simple question whether social information is also effective in inducing a behavioral change when decisions are made by teams. A long standing literature in social psychology and economics suggests that the answer is not obvious. Teams make more rational and selfish decisions than individuals. Hence, the documented success of social information on individual behavior cannot be used to infer their effect on team decisions. The paper therefore links the literatures on behavioral change and team decision making and makes the prediction that social information should be successful when decisions are made by teams.

The prediction is tested in a two-team prisoner's dilemma experiment with naturally occurring teams from the Lake Victoria fisheries in Tanzania. Results suggest that social information can indeed increase cooperative behavior when decisions are made by teams. Moreover, the field context informs an additional variation in the experimental design. The experiment randomly imposes two decision making mechanisms commonly used in fishing crews: teams either decide through an egalitarian or hierarchical approach. With naturally occurring teams, the paper identifies expe-

rience with the respective decision making mechanism as an important determinant for desirable behavioral change. When teams decide through dictatorial choice, the treatment effect is driven by fishermen with authority and leadership experience in real life. This secondary result suggests that social information is more effective with resource users in leadership positions.

Chapter 4, *"Captains of Change: The Effect of Social Information on Resource Users in Leadership Positions"*, co-authored with Philipp Händel, raises the policy relevant question of who should be the target of a social information intervention that aims to induce a pro-social change in natural resource management. Results in chapter three intriguingly suggests that resource users with leadership experience (*i.e.*, captains and vessel owners) may be leveraged as facilitators of collective change.

To test the robustness of the leadership result in chapter three, the paper re-analyzes the three-person prisoner's dilemma studied in chapters one and two. It finds that resource users in leadership positions are indeed exceptionally receptive to a social information message. In contrast, the intervention fails with regular crew members. By studying the treatment effect of an additional unrelated social information intervention, the paper shows that the main result is not explained by the fact that those fishermen who work as captains and owners are more strongly affected by social information per se. Findings may inform policy makers on who they should target as their "captains of change" in natural resource management.

Policy Implications and Research Outlook

The dissertation establishes that social information is an effective tool to induce behavioral change in a social dilemma. In doing so, the presented experimental evidence builds an important bridge between the laboratory and the field and thereby addresses the need to link academic research on social information to real-world scenarios (Kinzig et al., 2013). The field context at Lake Victoria does so along several lines. First, the effect of social information is tested with actual resource users that need to solve a social dilemma under weak formal institutions in their everyday lives. Second, social information is tested when decisions are made by naturally occurring teams, addressing the fact that the key actors in many social dilemmas such as common pool resource use are groups of individuals that make joint decisions. Third, the analyses on the effect of social information incorporate aspects of real life community structures including the role of social enforcement, social proximity, and leadership. Findings suggest that the pro-social forces of communication, social cohesion, pun-

ishment, and leadership (Janssen et al., 2010; Gutiérrez et al., 2011) may be jointly leveraged as a policy tool for governing a social dilemma.

To induce mutual cooperation and thereby help to avert overexploitation of natural resources, policy-oriented research may focus on designing and testing social information interventions in the field. Results in this dissertation suggest that social information should be supported by mechanisms for social enforcement and may be targeted towards resource users in leadership positions. Policy makers will also be interested in the cost-effectiveness and persistence of social information interventions. While the literature following Allcott's (2011) influential study on household energy use finds that effects persist and scalability is cost-effective (Ferraro et al., 2011; Ayres et al., 2013; Costa and Kahn, 2013), social norms are highly context dependent. Especially in developing countries, the implementation and evaluation of effective policies require community trust, cross-institutional collaboration, and the credibility of social information itself.

There are still unknowns on how social information induces behavioral change. Future research needs to study how preexisting social norms interact with newly introduced or activated ones. For example, does an existing social norm for reciprocity reinforce the treatment effect of a social information intervention that introduces a norm of fairness? How do outcomes change when social information attempts to activate a norm that is in conflict with established rules of behavior? Disentangling the complexity of social interactions is a daunting task. Yet, behavioral and experimental approaches are well-suited to continue testing whether models of normative influence can be operationalized to leverage social norms as policy tools in situations where individual and social incentives are at odds.

In examining the effect of social information, this dissertation does not address the role of direct communication. In fact, direct communication is prohibited in the two underlying experiments. While this retains some experimental control to better isolate a treatment effect, it removes an important form of social interaction. As such, communication is shown to facilitate cooperation, both in interactions between individuals (Andrighetto et al., 2013) and between groups (Iida and Schwieren, 2016). Yet, there is only limited evidence on how communication interacts with social information. Abrahamse and Steg (2013) argue that communication increases the effect of social information as it makes social identities and the associated norms more salient. Also, communication may help to enforce social norms, even before transgressions occur (Kinzig et al., 2013). It seems intuitive that social information may be more effective if people can use communication as an additional tool to coordinate on a

mutually cooperative response. Here, more research is necessary to pin down how exactly social information and communication interact.

The use of social information has limits. When there are uncertainties about the reference group or when agents have reason to doubt the credibility of the information provided, the formation of self-serving beliefs may undermine the intervention's effect (Bicchieri and Dimant, 2019; Charness et al., 2019). Also, any social information has a secondary meaning. That is, a message that attempts to induce desirable behavior by pointing towards common actions reveals that there are indeed some people that do the opposite. If agents disproportionately latch onto this interpretation, interventions may backfire. Further research is necessary to understand under which conditions such "boomerang effects" (Schultz et al., 2007) occur, to what extent they compromise the overall effectiveness of social information, or how avoid to them entirely. A related concern may be the overuse of social information. In a world where norm nudges are used to regulate too many aspects of life, they may start to fall on deaf ears as the preferences that induce norm conformity become oversaturated.

Social norms are not a panacea. There is a reason why in some aspects of human life, laws and their enforcement are necessary. Yet, social norms exist to regulate social interactions. They may help to find solutions to the many unregulated social dilemmas that permeate life. Social norms can address collective action problems at the individual and group level, are cost-efficient, flexible, may persist across generations, and appeal to a very basic idea of human cooperation.

Bibliography

Abrahamse, W. and Steg, L. (2013). Social influence approaches to encourage resource conservation: A meta-analysis. *Global Environmental Change*, 23(6):1773–1785.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.

Allcott, H. and Kessler, J. B. (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics*, 11(1):236–76.

Alpizar, F., Carlsson, F., and Johansson-Stenman, O. (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in costa rica. *Journal of Public Economics*, 92(5-6):1047–1060.

Anderies, J. M., Janssen, M. A., Bousquet, F., Cardenas, J.-C., Castillo, D., Lopez, M.-C., Tobias, R., Vollan, B., and Wutich, A. (2011). The challenge of understanding decisions in experimental studies of common pool resource governance. *Ecological Economics*, 70(9):1571–1579.

Andrighetto, G., Brandts, J., Conte, R., Sabater-Mir, J., Solaz, H., and Villatoro, D. (2013). Punish and voice: Punishment enhances cooperation when combined with norm-signalling. *PloS one*, 8(6):e64941.

Ayres, I., Raseman, S., and Shih, A. (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *The Journal of Law, Economics, and Organization*, 29(5):992–1022.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Bénabou, R. and Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics*, 126(2):805–855.

Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy*, 102(5):841–877.

Bicchieri, C. (2006). *The Grammar of Society*. Cambridge University Press, Cambridge, 1. publ. edition.

Bicchieri, C. (2017). *Norms in the Wild – How to Diagnose, Measure, and Change Social Norms*. Oxford University Press.

Bicchieri, C. and Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public Choice*, pages 1–22.

Bicchieri, C., Dimant, E., Gächter, S., and Nosenzo, D. (2019). Social proximity and the evolution of norm compliance. Technical report, Working Paper.

Bicchieri, C., Muldoon, R., and Sontuoso, A. (2018). Social norms. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition.

Binmore, K. (2010). Social norms or social preferences? *Mind & Society*, 9(2):139–157.

Cardenas, J. C. (2011). Social norms and behavior in the local commons as seen through the lens of field experiments. *Environmental and Resource Economics*, 48(3):451–485.

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

Charness, G., Naef, M., and Sontuoso, A. (2019). Opportunistic conformism. *Journal of Economic Theory*, 180:100–134.

Cialdini, R. B. and Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, 55:591–621.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015–1026.

Cialdini, R. B. and Trost, M. R. (1998). Social influence: Social norms, conformity and compliance.

Cooper, D. J. and Kagel, J. H. (2016). Other-regarding preferences. In Kagel, J. H. and Roth, A. E., editors, *The Handbook of Experimental Economics, Volume 2*, pages 217–289. Princeton University Press.

Costa, D. L. and Kahn, M. E. (2013). Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3):680–702.

Croson, R., Handy, F., and Shang, J. (2009). Keeping up with the Joneses: The Relationship of Perceived Descriptive Social Norms, Social Information, and Charitable Giving. *Nonprofit Management and Leadership*, 19(4):467–489.

Dimant, E. (2019). Contagion of pro-and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, 73:66–88.

Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives*, 3(4):99–117.

Engl, F., Riedl, A., and Weber, R. A. (2020). Spillover effects of institutions on cooperative behavior, preferences, and beliefs. *Preferences, and Beliefs (August 3, 2020)*.

Etiegni, C. A., Irvine, K., and Kooy, M. (2020). Participatory governance in lake victoria (kenya) fisheries: whose voices are heard? *Maritime Studies*, pages 1–19.

Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

Fehr, E. and Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, 2(7):458–468.

Ferraro, P. J., Miranda, J. J., and Price, M. K. (2011). The persistence of treatment effects with norm-based policy instruments: evidence from a randomized environmental policy experiment. *American Economic Review*, 101(3):318–22.

Gerber, A. S. and Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, 71(1):178–191.

Goeschl, T., Kettner, S. E., Lohse, J., and Schwieren, C. (2018). From social information to social norms: Evidence from two experiments on donation behaviour. *Games*, 9(4).

Goette, L., Huffman, D., and Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review*, 96(2):212–216.

Gutiérrez, N. L., Hilborn, R., and Defeo, O. (2011). Leadership, social capital and incentives promote successful fisheries. *Nature*, 470(7334):386–389.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3):61–83.

Iida, Y. and Schwieren, C. (2016). Contributing for myself, but free riding for my group? *German Economic Review*, 17(1):36–47.

Janssen, M. A., Holahan, R., Lee, A., and Ostrom, E. (2010). Lab experiments for the study of social-ecological systems. *Science*, 328(5978):613–617.

Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.

Kinzig, A. P., Ehrlich, P. R., Alston, L. J., Arrow, K., Barrett, S., Buchman, T. G., Daily, G. C., Levin, B., Levin, S., Oppenheimer, M., et al. (2013). Social norms and global environmental challenges: the complex interaction of behaviors, values, and policy. *BioScience*, 63(3):164–175.

Latané, B. (1981). The psychology of social impact. *American Psychologist*, 36(4):343.

Lopez, M. C., Murphy, J., M., S. J., and Stranlund, J. K. (2012). Comparinng the effectiveness of regulation and pro-social emotions to enhance cooperation: Experimental evidence from fishing communities in colombia. *Economic Inquiry*, 50(1):131–142.

Meier, S., Pierce, L., Vaccaro, A., and La Cara, B. (2016). Trust and in-group favoritism in a culture of crime. *Journal of Economic Behavior & Organization*, 132:78–92.

Michaeli, M. and Spiro, D. (2017). From peer pressure to biased norms. *American Economic Journal: Microeconomics*, 9(1):152–216.

Miller, D. T. and Prentice, D. A. (2016). Changing norms to change behavior. *Annual Review of Psychology*, 67:339–361.

Nunan, F., Cepić, D., Mbilingi, B., Odongkara, K., Yongo, E., Owili, M., Salehe, M., Mlahagwa, E., and Onyango, P. (2018). Community cohesion: Social and economic ties in the personal networks of fisherfolk. *Society & Natural Resources*, 31(3):306–319.

Ostrom, E. (2008). The challenge of common-pool resources. *Environment: Science and Policy for Sustainable Development*, 50(4):8–21.

Ostrom, E. and Walker, J. (1991). Communication in a commons: cooperation without external enforcement. *Laboratory Research in Political Economy*, pages 287–322.

Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86(2):404–417.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429–434.

Sugden, R. (2000). The motivating power of expectations. In *Rationality, Rules, and Structure*, pages 103–129. Springer.

Tajfel, H. and Turner, J. C. (1979). An integrative theory of intergroup conflict. *Organizational Identity: A Reader*, pages 56–65.

Velez, M. A., Stranlund, J. K., and Murphy, J. J. (2009). What motivates common pool resource users? experimental evidence from the field. *Journal of Economic Behavior & Organization*, 70(3):485–497.

Vollan, B. and Ostrom, E. (2010). Cooperation and the commons. *Science*, 330(6006):923–924.

Chapter 1

Dissecting Social Norms of Cooperation: A Conditional Process Analysis

Abstract: Social information and sanctioning mechanisms can induce behavioral change. In combination, the two measures are however sparsely tested and the underlying behavioral mechanism is poorly understood. With a conditional process analysis, I focus on the interplay between social information and sanctioning and study the drivers of behavioral change when both measures are combined. Results suggest that social information messages induce a behavioral change in two ways. First, their effect on cooperation is, in part, mediated by empirical expectations. Those who receive information about cooperative behavior of others, expect them to cooperate and subsequently cooperate themselves. Second, social information directly affects cooperation when combined with a sanctioning opportunity. Moreover, I highlight the methodological importance to explore causal paths via the inclusion of contextual variables in the framework of social information, norm conformity, and behavioral change.

Keywords: social information; sanctioning; social norms; cooperation

1.1 Introduction

When formal regulations fail to effectively govern collective action problems, policy makers and researchers seek informal means of inducing pro-social behavior. Among the measures that can help solve a social dilemma are (i) social sanctioning mechanisms, and (ii) the activation of social norms. First, social sanctions of undesirable behavior or the threat thereof are shown to facilitate cooperation (Fehr and Gächter, 2002; Engl et al., 2020). Second, the provision of social information can induce desirable behavior by changing the perceived social norm and prompting a desire for conformity (Bicchieri and Xiao, 2009; Croson et al., 2009; Goeschl et al., 2018). Social information and sanctions appear to benefit from obvious synergies. When norm transgression can be punished, compliance may increase. At the same time, sanctions may be more forceful when backed by strong social norms (Herrmann et al., 2008; Andrighetto et al., 2013).

When combined with a sanctioning possibility, so-called norm-based interventions or "norm-nudges" have had mixed success in inducing behavioral change. While Fehr and Schurtenberger (2018) find that the effectiveness of social information depends on a punishment opportunity, Bicchieri et al. (2020b) show that their combination can also have a detrimental effect on cooperation. Importantly, the underlying behavioral mechanism is still poorly understood. Policy makers that want to use a combination of sanctioning mechanism and social information as an informal regulation tool will look to leverage synergies and induce conformity but "must understand the limitations of nudging" (Bicchieri and Dimant, 2019, p.7). As a first contribution, I therefore focus on the interplay between social information and a sanctioning opportunity and study the drivers of behavioral change when both measures are combined.

To dissect the underlying behavioral mechanism into conditional and causal processes and thereby link the provision of social information under sanctioning to a change in cooperation behavior in a public goods game, I utilize a conditional process analysis. The conditional process analysis is a versatile methodological approach to analyze whether certain causal processes are conditional on other influences such as additional treatments, individual characteristics, or circumstance (Hayes, 2018). In particular, my study builds on Croson et al. (2009) who use a mediation analysis to establish a causal link between the provision of social information and pro-social behavior that works through influencing the perceived descriptive social norm. As a second contribution, I additionally model the possibility to sanction and be sanctioned by others as a moderating influence and thereby showcase the importance to

explore causal paths via the inclusion of contextual variables in the framework of social information, norm conformity, and behavioral change.

I follow the definition of a social norm by Bicchieri (2017) and define *empirical expectation* as the belief about what others do and *normative expectation* as the belief about what others think that one should do.[1] Agents are considered to follow a social norm if they condition their behavior on their expectations regarding the beliefs and behavior of others. Norm conformity is therefore directly tied to the idea that social beliefs correlate with own decisions. By leveraging this correlation, interventions can try to influence normative and empirical expectations in order to initiate behavioral change.

The related literature has established that behavioral change can be induced through the provision of information about others' beliefs or behavior. Pioneering contributions focus on promoting environmentally-friendly consumer behavior including recycling (Schultz, 1999) and energy use reduction (Allcott, 2011; Costa and Kahn, 2013). Since then, social information has been successfully used to induce charitable giving (Croson et al., 2009; Goeschl et al., 2018), tax compliance (Hallsworth et al., 2017), and cooperation in strategic experimental settings (Fehr and Schurtenberger, 2018; Bicchieri et al., 2020b; Diekert et al., 2021). Interventions generally aim to affect either normative or empirical expectations and results suggest that, with exceptions, both message types can be effective in inducing desirable behavior (Farrow et al., 2017).

I link the provision of social information with endogenous weak sanctioning, a second mechanism used to discourage undesirable behavior and maintain cooperation in the lab and in the field. Masclet et al. (2003) and Tyran and Feld (2006) present evidence that non-deterrent sanctions, *i.e.*, punishment that does not change payoff incentives from defection to cooperation, can activate pro-social norms in public goods experiments and increase cooperation. Fehr and Fischbacher (2004), Xiao and Houser (2011), and Andrighetto et al. (2013) add that punishment works as an ancillary mechanism that gives salience to the norm and thereby leads participants to better recognize the consequences of violation. Peers punish or threaten to punish norm transgression and thus cause an increase in pro-social behavior (Fehr and Gächter, 2000, 2002).

[1]The distinction between a normative and an empirical motivation for behavior matches the idea brought forth by Cialdini et al. (1990) who differentiate between injunctive and descriptive norms, between what ought to be done and what is done. The definition by Bicchieri (2017) considers expectations about others' behavior (empirical expectation) and beliefs (normative expectation), making a clear distinction to personal normative beliefs, one's own moral stance on what ought to be done.

To my knowledge, only two studies explicitly look at the interaction between a sanctioning mechanism and the provision of social information. First, Fehr and Schurtenberger (2018) find that normative priming only leads to sustained cooperation when peer punishment is allowed. Otherwise, pro-social behavior deteriorates over time suggesting that punishment is necessary for norm conformity. Second, Bicchieri et al. (2020b) disentangle the effects of punishment, normative or empirical information, and the respective combinations thereof on return behavior in a modified trust game. They establish an effect of punishment and social information on pro-social behavior when used in combination but find no such relationship when either treatment is isolated. While the combination of punishment and normative information can increase pro-social behavior, empirical information combined with punishment can have detrimental effects that occur because punishment is perceived to be unjustified when agents have reason to doubt the validity or applicability of the message.[2] In sum, Fehr and Schurtenberger (2018) and Bicchieri et al. (2020b) establish a joint effect of sanctioning opportunity and social information on pro-social behavior, but they do not focus on the behavioral mechanism linking information treatment to a change in behavior with a possible dependency on a sanctioning mechanism.

I use a one-shot prisoner's dilemma game which varies (i) whether participants receive high or low social information and (ii) whether they have the possibility to weakly sanction others' choices. Data comes from a lab-in-the-field experiment with fishermen from Lake Victoria, Tanzania, in which the interaction in the prisoner's dilemma is repeated. The repetition was only announced to participants after the first round was completed, making the first interaction a true one-shot procedure that rules out any forward looking motivations that could influence the causal and conditional processes subject to this study. In a comprehensive analysis on the strategic interactions in the repeated game, Diekert et al. (2021) find that under sanctioning, the provision of social information creates different social norms of cooperation.

In addition to the observation of contribution decisions, I elicit normative and empirical expectations in an incentivized setting. That allows me to directly apply Bicchieri's (2017) concept of a social norm to the data and study the belief structure that may transmit the effect of social information to cooperation behavior. In particu-

[2]Bicchieri et al. (2020b) allow for high and low costs of conformity. In particular, investors could transfer none, half or their full endowment while sending a costless request message, asking to return half of the transfer. The setup makes a trustee's compliance relatively cheap when half the endowment was transfered, compared to the case where the full endowment is sent. With cheap compliance, the combination of normative information and punishment increases returns by the trustees. With costly compliance, the combination of empirical information and punishment causes the message to backfire.

lar, the proposed conditional process analysis models both social beliefs as mediating variables and the sanctioning opportunity as a moderating influence on all causal paths linking treatment to outcome. The approach enables a careful identification of norm-based channels for the effect of social information on pro-social behavior under sanctioning and demonstrates how conformity interacts with a sanctioning mechanism to drive behavioral change.

Results suggest that the provision of social information in the experiment with a sanctioning opportunity induces a behavioral change in two ways. First, social information increases cooperation through a change in empirical expectations. Those who receive information about cooperative behavior of others, expect them to cooperate and subsequently cooperate themselves. The effect is small but independent of the sanctioning mechanism. Second, I find that social information induces a direct, significant increase in cooperation when combined with a sanctioning possibility. When holding participants' beliefs fixed, information on the pro-social behavior of others leads to more cooperation only when participants are able to weakly sanction others.

1.2 Experimental Design and Procedure

The experiment is a one-shot three-person prisoner's dilemma game with prior, incentivized belief elicitation. It is varied whether participants receive a high or low social information message and whether they enter a coordinated sanctioning stage after their contribution decision.

Table 1.1: Payoff matrix

		Number of other cooperators		
		0	*1*	*2*
Own decision	*cooperate*	2pt	4pt	6pt
	defect	4pt	6pt	8pt

Participants are endowed with four points that they can either put towards their private account (defect) or towards a group account (cooperate). The group account is shared with two other players. Only the full endowment can be transferred to either account, *i.e.*, participants only have the option to put zero or four points to the group or private account. For every group account contribution, all group members receive two points towards their private account, a marginal per capita return of 0.5. Defection is the payoff-dominant strategy, see Table 1.1.

Sanctions are based on the exclusion from receiving one extra point that is available

for each group member after individual contribution decisions. A voting procedure within each group determines the group's rule for exclusion. Every group member casts a vote on either (i) excluding everyone who put the points to their private account, (ii) everyone who put their points to the group account, or on (iii) excluding no one. The majority decides which exclusion rule is implemented.[3] Exclusion votes are cast without knowledge of others' behavior but the sanctioning mechanism itself is introduced before contribution decisions are made. It is therefore possible that a vote is cast on an option that is not played by anyone in the respective group. Even with the potential exclusion of defectors from the extra point, defection remains the payoff-dominant strategy.

Prior to the contribution decision, participants are asked for their belief regarding what they think is the right thing to do, their *personal normative belief*; what they think most others think is the right thing to do, their *normative expectation*; and what they think others will actually do, their *empirical expectation*. Before beliefs are elicited, a social information message is verbally given by the instructor to all participants, see Table 1.2.[4]

Table 1.2: Messages in social information treatments

Treatment	*Message*
High information	In a previous survey, it was found that many participants chose to put the points to the group account and not to put them to their private account.
Low information	In a previous survey, it was found that many participants chose to put the points to their private account and not to put them to the group account.

The message is designed to affect empirical expectations in the direction of either cooperative (high) or non-cooperative (low) behavior. After personal normative beliefs and normative expectations are elicited, the message is given a second time before participants state their empirical expectation. For personal normative belief as well

[3]This implies that any option that receives two or three votes within a group becomes the majority rule. If all options receive exactly one vote, *i.e.*, no majority is reached among the three group members, the rule implemented is the exclusion of no member from the bonus point. Participants are fully informed about the majority voting.

[4]For clarification, the experiments' procedure is as follows: (1) general game instructions and social information message, (2) belief elicitation, (3) instructions on the sanctioning opportunity, (4) contribution decision, (5) sanctioning decision. Participants learn about contribution choices of others only after the exclusion rule is determined. Therefore, whether participants were effectively excluded does not have any implications for the analysis at hand.

as normative expectation, options are to (i) put the points to the private account, (ii) group account, or to (iii) do what others so. The latter option is not available for empirical expectation. Normative as well as empirical expectations are incentivized by offering an extra point for each correct guess.

1.3 The Conditional Process Model

The aim of the study is to understand the interplay of social information and a sanctioning mechanism within the framework of norm conformity. It is helpful to reiterate the definition of a social norm. Following Bicchieri (2006, 2017), a social norm is defined as a rule of behavior that is followed conditional on two beliefs: *empirical expectation*, defined as the belief about what others do, and *normative expectation*, defined as the belief about what others think that one should do. Importantly, a social norm is distinguished from the *personal normative belief*, defined as what oneself thinks one should do. While the personal normative belief undoubtedly plays a role in determining individual behavior, it is considered to be a fixed, non-social belief that is unaffected by a social information treatment. To understand the effect of social information on cooperation and to uncover the behavioral mechanism that causally links treatment to outcome variable, I look towards empirical and normative expectations as potential channels for a transmission of the effect.

Following suggestions by VanderWeele (2015) and Hayes (2018), I proceed with a conditional process analysis on the effect of social information on cooperation through empirical and normative expectations. I deliberately depart from the simple mediation analysis proposed by Baron and Kenny (1986). While either approach is able to model the effect of social information on cooperation through beliefs, the conditional process analysis adds the possibility to model processes contingent on another treatment variable. With the experimental design at hand, the conditional process model allows me to answer the research question of whether the effect of social information on pro-social behavior is dependent on a sanctioning possibility. Therefore, I propose a parallel multiple mediator model of the effect of social information on cooperation through participants' beliefs about the social norm that is moderated by sanctioning. In what follows, the individual components of the conditional process analysis are explained alongside motivation for the expected effects in regard to the related literature.

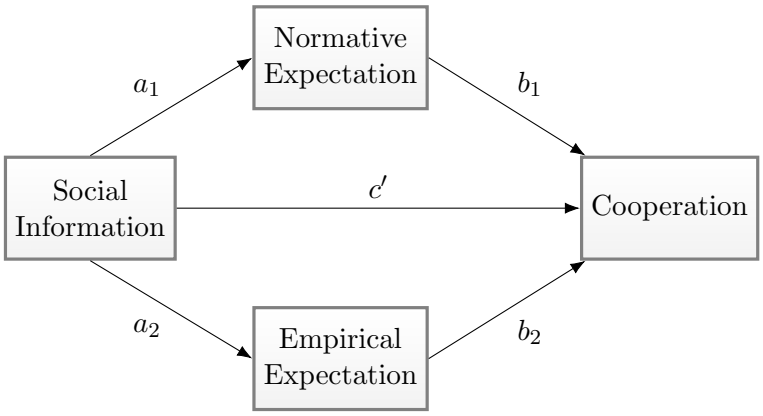


Fig. 1.1: Conceptual diagram of the parallel multiple mediator model

1.3.1 The Parallel Multiple Mediator Model

The effect of social information on cooperation transmitted through beliefs about the social norm is modeled as a parallel multiple mediator model, depicted in the conceptual diagram of Figure 1.1.

In the mediation model illustrated by Hayes (2018), the treatment has total effect c on the outcome. The total effect describes how differences between low ($X = 0$) and high ($X = 1$) social information X map onto differences in group means of cooperation $\bar{Y}$. Holding all else constant,

$$c = [\bar{Y}|(X = 1)] - [\bar{Y}|(X = 0)].$$

The total effect is in part due to a direct effect c' of the information treatment on cooperation, holding beliefs M_i with $i = 1, 2$ constant,

$$c' = [\bar{Y}|(X = 1, M_i = m)] - [\bar{Y}|(X = 0, M_i = m)],$$

and an indirect effect of the treatment on the outcome. Here, a_1 and a_2 depict the effect of the treatment on the respective mediator. They quantify the difference in group means of mediator $\bar{M}$ between the two different levels of social information X for the respective belief i,

$$a_i = [\bar{M}_i|(X = 1)] - [\bar{M}_i|(X = 0)] \text{ for } i = 1, 2.$$

The first step in establishing a causal connection is to study whether the infor-

mation message has a significant effect on the perceived social norm, *i.e.*, whether a_i is different from zero for normative and empirical expectation. In a study with an information message very similar to the one given to participants in this study, Bicchieri and Xiao (2009) find a positive effect of social information on normative and empirical expectation. In the treatment that only targets empirical expectations and is therefore most alike to the condition studied in the present analysis, a positive and significant effect on both beliefs is identified. Comparable results are obtained by Croson et al. (2009) who find that participants in high information treatments report a higher descriptive social norm. Participants take added knowledge about others' behavior into account when forming incentivized beliefs. Consequently, I expect that the social information message affects both normative and empirical expectations.

Hypothesis 1 *The high social information treatment has a positive effect on the perceived social norm, i.e.,* $a_i \geq 0$ *for* $i = 1, 2$.

Hypothesis 1 describes the first part of the indirect effect.

In the second part, each mediator affects the outcome variable. A change in participants' beliefs may lead to different cooperation behavior, depicted by b_1 and b_2. They quantify the difference in group means of cooperation $\bar{Y}$ for different levels M of the belief i, holding treatment X and the other belief constant,

$$b_i = [\bar{Y}|(M_i = 1, X = x)] - [\bar{Y}|(M_i = 0, X = x)] \text{ for } i = 1, 2.$$

The second step in establishing a causal link between the provision of social information and cooperation behavior is to study the effect of perceived social norms on the outcome. Given that Hypothesis 1 establishes normative and empirical expectation to be responsive to the treatment, the effect is transmitted to the outcome only when increased expectations lead to more cooperation. A significant effect is identified in the study by Croson et al. (2009) who thereby conclude that the effect of social information on contribution levels is mediated by perceived descriptive social norms. In the treatment that manipulates empirical expectations in isolation, Bicchieri and Xiao (2009) find a significant link between treatment and both normative and empirical expectation (see Hypothesis 1). However, they can only establish the effect to transmit to subsequent behavior for the empirical belief which suggests that changes in normative expectations do not necessarily translate into differences in contributions when the nature of the message is empirical.

The expected impact of normative expectation on behavior in the prisoner's dilemma

is nonetheless ambiguous. Bicchieri and Xiao (2009) reason that both normative and empirical expectation are initially affected by the provision of one isolated message since participants have no reason to distinguish between the two beliefs. That is, an empirical message carries normative meaning and vice versa. If others exhibit certain behavior, they most likely consider it appropriate. Likewise, if others consider an action to be appropriate, the message also reveals that they probably act accordingly. In general, both normative and empirical expectations are essential in forming social norms (Bicchieri, 2017) and (perceived) norms drive own choices (Cialdini et al., 1990; Bicchieri, 2006). Consequently, I expect both normative and empirical expectation to influence cooperation:

Hypothesis 2 *High normative and empirical expectation explain individual cooperation, i.e.,* $b_i \geq 0$ *for* $i = 1, 2$.

Both parts of the indirect effect jointly describe the total indirect effect. The total indirect effect of X on Y through both mediators is given by the sum of the product of the coefficient for the effect of the treatment on the mediator ($\theta_{X \to M_i}$), and the coefficient for the effect of the mediator on the outcome ($\theta_{M_i \to Y}$) over all mediators, M_i for $i = 1, 2$, such that

$$\sum_{i=1}^{2} \theta_{X \to M_i} \theta_{M_i \to Y} = (a_1 b_1 + a_2 b_2).$$

Hypotheses 1 and 2 imply a significant positive indirect effect of social information on cooperation through both mediators M_i with $i = 1, 2$, *i.e.*,

$$\theta_{X \to M_i} \theta_{M_i \to Y} = a_i b_i \geq 0. \tag{1.1}$$

Jointly, the total indirect effect and direct effect describe the total effect c (Vander-Weele, 2015; Hayes, 2018). The corresponding statistical model requires the following regressions,

$$M_1 = i_{M_1} + a_1 X + e_{M_1} \tag{1.2}$$

$$M_2 = i_{M_2} + a_2 X + e_{M_2} \tag{1.3}$$

$$Y = i_Y + c' X + b_1 M_1 + b_2 M_2 + e_Y \tag{1.4}$$

with i_{M_i} and i_Y as intercepts, and e_{M_i} and e_Y as the estimation errors of M_i for $i = 1, 2$ and Y, respectively. In equation (1.2) and (1.3), a_1 and a_2 estimate the

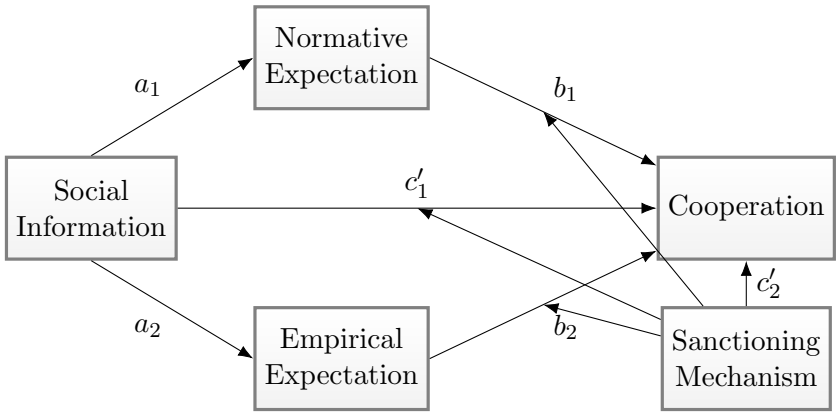


Fig. 1.2: Conceptual diagram of mediation and moderation

effect of social information on normative and empirical expectation, thereby testing Hypothesis 1. Hypothesis 2 is examined with equation (1.4), where b_1 and b_2 estimate the effect of the respective belief on cooperation behavior. Additionally, c' estimates the effect of treatment X on outcome Y holding all M_i constant.

1.3.2 Moderation

The main contribution of the paper is to study whether the underlying behavioral mechanism induced by the provision of social information described in Hypotheses 1 and 2 is contingent on the weak sanctioning treatment introduced in Section 1.2. In Hayes (2018), the analytical strategy that explores whether a certain effect or causal path is contingent on the value of another variable is described as moderation. I design a moderation model to test whether the direct influence of social information on cooperation, and its effect mediated by beliefs about the social norm is conditional on the possibility to weakly sanction others' behavior. An illustration of the moderation is given by the arrows drawn from the sanctioning treatment to the paths depicting the direct effect c_1' and the indirect effects b_1 and b_2 in Figure 1.2. Since the sanctioning mechanism is introduced to participants only after beliefs are stated, the model does not allow for moderation of the causal paths from information message to normative and empirical expectation described by a_1 and a_2.

Evidence for the supplementary effect of severe as well as weak and non-deterrent punishment with respect to social norms is manifold. Tyran and Feld (2006) show that voted upon weak sanctions, therein referred to as "mild law", activate cooperation norms and significantly increase contributions to a public good. Masclet et al. (2003)

find that informal, non-monetary sanctions such as social ostracism lead to increased cooperation in their Voluntary Contributions Mechanism. Furthermore, while comparing different punishment implementation tools, Xiao and Houser (2011) find that cooperation in groups is sustained with publicly implemented punishment and has the opposite effect when sanctions are private. Again, the ability to better express norms of cooperation is suggested as the cause of increased contributions to the public good.

Conflicting results are presented by Bicchieri et al. (2020b), who find that the combination of empirical information and weak punishment has detrimental effects on return behavior in their modified trust game. Two possible explanations for such unexpected results are given. First, participants may take advantage of the wiggle room offered by an empirical message to avoid conformity by forming self-serving beliefs that morally justify non-conformity (Konow, 2000; Dana et al., 2007). Second, the experimental design may create uncertainty about the reference group, leading individuals who do not wish to conform to form beliefs that the norm does not apply to them, but only to others (Bicchieri et al., 2020a). In her definition of social norms, Bicchieri (2017) emphasizes the importance of a reference network in motivating conformity. It is only important what others do or think one ought to do if "others" refers to a group that is relevant to the respective actor.

The overall direction and strength of the conditional effect of the sanctioning mechanism on the causal paths from social information to cooperation behavior is thus ambiguous. To clear the ambiguity, the model is designed to reveal if and how the behavioral mechanism at work differs between participants who have the sanctioning possibility and those who do not. I expect that there is a conditional indirect effect but both direction and strength of the moderation are posed as open questions. That is, the effect of perceived social norms on cooperation, *i.e.*, the paths described by b_1 and b_2, are assumed to vary with the sanctioning treatment W:

Hypothesis 3 *The processes described by Hypothesis 2 are contingent on the sanctioning possibility, i.e.,* $(\theta_{M_i \to Y|W=1}) \neq (\theta_{M_i \to Y|W=0})$ *for* $i = 1, 2$.

Hypothesis 3 translates to the statistical model as follows:

$$M_1 = i_{M_1} + a_1 X + e_{M_1} \tag{1.5}$$

$$M_2 = i_{M_2} + a_2 X + e_{M_2} \tag{1.6}$$

$$Y = i_Y + c_1' X + c_2' W + c_3' XW + b_i M_i + b_{i3} M_i W + e_Y \text{ for } i = 1, 2 \tag{1.7}$$

Note that the estimation of Y in the model with moderation now includes interaction terms of all effects assumed to be moderated by the sanctioning treatment, see equation (1.7) in comparison to equation (1.4). In particular, b_{i3} estimates by how much the difference in average cooperation Y between the two levels of belief M_i with $i = 1, 2$ changes when sanctioning W is present compared to when it is absent. The respective conditional indirect effect of belief M_i on Y is quantified by $\theta_{M_i \to Y} = b_i + b_{i3}W$ for $i = 1, 2$. If and how the causal path from normative or empirical expectation is contingent on the sanctioning mechanism will be revealed by the strength and sign of b_{13} and b_{23}, thereby testing Hypothesis 3.

Similarly, c_3' estimates by how much the difference in average cooperation Y between the two levels of social information X changes with the sanctioning mechanism. The conditional direct effect of social information treatment X on cooperation Y is thus given by $\theta_{X \to Y} = c_1' + c_3'W$. In case the combination of social information and sanctioning mechanism has an impact on individual cooperation behavior that is not explained by the causal path through either mediator, the residual will be picked up by the c_3' coefficient. In addition, the sanctioning treatment is assumed to exert its own direct effect c_2' on cooperation, see Figure 1.2. Since the sanctioning mechanism does not affect the causal path from the social information treatment to either normative or empirical expectation, equations (1.2) and (1.3) are identical to equations (1.5) and (1.6), respectively. It follows that the conditional indirect effect of X on Y through M_i is described by $a_i\theta_{M_i \to Y} = a_i(b_i + b_{i3}W)$ such that the total conditional indirect effect of X on Y through all mediators M_i with $i = 1, 2$ is given by $\sum_{i=1}^{2} \theta_{X \to M_i}\theta_{M_i \to Y} = \sum_{i=1}^{2} a_i(b_i + b_{i3}W)$.

1.3.3 The Full Conditional Process Model

A comprehensive approach to modeling causal processes between treatment X and outcome Y should explain whether the relationship is mediated and recognize possible contingencies on the experimental design, circumstance or individual differences (Hayes, 2018). Mediation and moderation are combined in a conditional process analysis.

The full conditional process model is depicted in Figure 1.3. It integrates the parallel multiple mediator and moderation model with all its properties discussed above. Furthermore, participants' personal normative beliefs are added as an explanatory variable for cooperation. Note, that the personal normative belief is not modeled as a mediator since it is assumed to be fixed and therefore independent of the social infor-

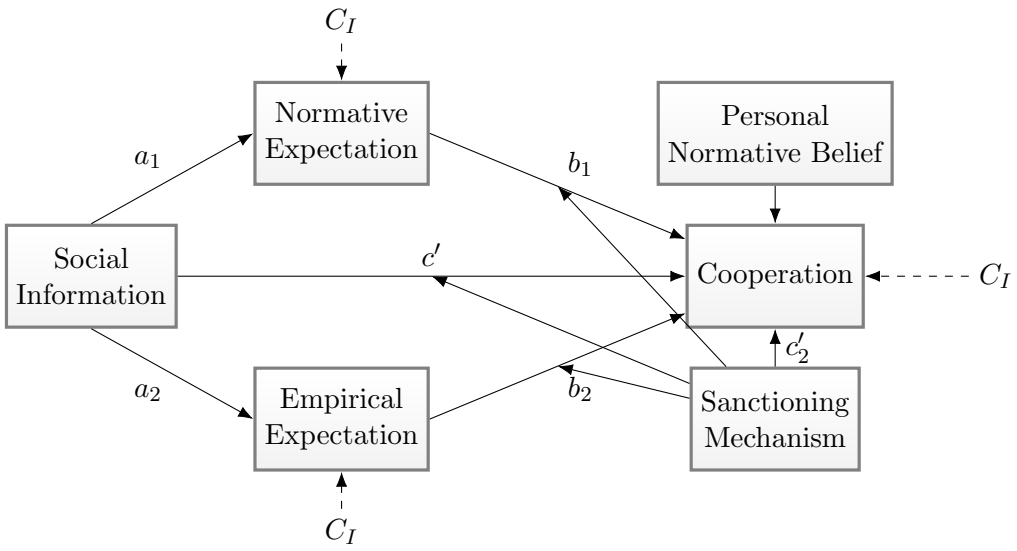


Fig. 1.3: Conceptual diagram of the conditional process model

mation treatment. Nonetheless, it is expected to be a powerful predictor of individual behavior (Bicchieri, 2017). Individual-level controls C_I are added to each equation in the regression model.[5]

With V indicating participants' personal normative beliefs, the full statistical model requires the following regressions that will be tested with a generalized structural equation model, see Section 1.5:

$$M_1 = i_{M_1} + a_1 X + f_1 C_I + e_{M_1} \tag{1.8}$$

$$M_2 = i_{M_2} + a_2 X + f_2 C_I + e_{M_2} \tag{1.9}$$

$$Y = i_Y + c_1' X + c_2' W + c_3' XW + b_i M_i + b_{i3} M_i W + dV + f_3 C_I + e_Y \text{ for } i = 1, 2 \tag{1.10}$$

I acknowledge the nature of the message provided in the experiment. The phrasing used to affect social norms explicitly addresses empirical rather than normative expectations. While both beliefs are assumed to be affected, see the reasoning for Hypothesis 1 and Bicchieri and Xiao (2009), I expect the effect to be more pronounced for empirical expectations. By extension of the argument, the effects for empirical expectations are expected to be larger throughout the entire causal path:

[5]Individual controls include participants' age, daily income, comprehension of the game's rules, and a dummy for prior participation in an economic experiment.

Hypothesis 4 *The processes described by Hypotheses 1, 2, and 3 are stronger for empirical than for normative expectation.*

Hypothesis 4 implies that without the sanctioning possibility, the indirect effect of the mediation process described in equation (1.1) is expected to be stronger for empirical than for normative expectation, *i.e.*, $a_2 \geq a_1 \geq 0$ and $b_2 \geq b_1 \geq 0$ such that,

$$\theta_{X \to M_2} \theta_{M_2 \to Y|W=0} \geq \theta_{X \to M_1} \theta_{M_1 \to Y|W=0}.$$

With a sanctioning mechanism, the same conjecture is not necessarily true. A potential detrimental effect of the sanctioning possibility on the underlying behavioral mechanism would reverse the sign of the relationship. Thus, it is expected that the moderation of the mediation effect, independent of the direction, is more pronounced for empirical expectations than for normative expectations, *i.e.*, $|b_{23}| \geq |b_{13}|$.

To summarize, both normative and empirical expectation are expected to mediate the effect of social information on cooperation. While the relationship is expected to be stronger for empirical than normative expectation, the impact of the sanctioning mechanism on each causal process is ambiguous. I shed light on the underlying behavioral mechanism that link social information to cooperation through social beliefs by studying the conditional indirect effects for normative and empirical expectations. Moreover, any effect of the combination of social information and sanctioning opportunity on the underlying behavioral mechanism that is not explained by the moderated mediation processes will be picked up by the conditional direct effect.

1.4 Data

The experiment was conducted in 20 villages with a total of 28 sessions in the Lake Victoria region of Tanzania between February and March 2018. Each session comprised 21 participants ($N = 580$) that made their living in the lake's fisheries.[6] The selection of participants prioritized participants from a field trip to the same landing sites in 2017 with a re-sampling rate just under 50%. Re-sampling will be controlled for in the analysis. The remaining participants were randomly selected from available fishermen on site.

The experimental procedure was as follows. Participants received an explanation

[6]Eight participants were dropped from the analysis as they violated the procedural rules during the belief elicitation stage. Since chapters 1 and 2 study the same experiment, participant characteristics with balance tests across treatments are shown in Appendix Table A-1 that follows chapter two.

of the general rules of the session and informed consent was obtained. Participants were guided through the prisoner's dilemma game step by step, utilizing posters for visualization and requisites for illustration of the game's mechanics. In particular, it was explained that decisions during the game would translate to real money dependent on choices of other group members. All decisions were made anonymously and no communication between participants was allowed during the session.[7] Comprehension was tested with the help of four different scenarios that asked participants to name the correct payoffs after decisions were made by all group members. Results of these questions are used as a measure of understanding in the analysis.

The experiment took place in the village's community center or directly at the landing site's beach. All decisions were made via input on tablet computers that were distributed to each participant. Together with the repeated rounds, a short questionnaire, and payout, each session lasted approximately 2.5 hours and participants earned an average of 5,000 Tanzanian Shilling (TZS).[8]

1.5 Results

I present test results on the four Hypotheses stated in Section 1.3 that build on the conditional process model explained therein. Average marginal effects from a generalized structural equation model on the basis of equations (1.8), (1.9), and (1.10) are reported in Table 1.3.

1.5.1 Indirect Effect of Social Information on Cooperation through Beliefs

Following the conditional process model described in Section 1.3, the analysis of the indirect effect includes an unconditional link between social information treatment and social beliefs as well as a conditional link between social beliefs and cooperation that tests whether the described relationship is dependent on the sanctioning mechanism.

For the first part of the indirect effect, I test whether the mediating variables

[7] Screenshots of all relevant choice situations (normative and empirical expectation, cooperation, and sanctioning) and the experimental instructions are available in the supplementary material, published online.

[8] The experiment was conducted with the help of four Tanzanian research assistants who helped with the game's explanation and assisted throughout the experiment when help was needed. The oTree software was used to implement the prisoner's dilemma game on tablets in the lab-in-the-field experiment (Chen et al., 2016). 5,000 TZS equal about US$ 2.20. The minimum payout was set to 2,500 TZS.

Table 1.3: Average marginal effects in the conditional process model

	Normative Exp.	Empirical Exp.	Cooperation
	(1)	(2)	(3)
Social Information		0.087***	-0.065
		(0.029)	(0.074)
– Defect	-0.024		
	(0.027)		
– Do what others do	0.002		
	(0.002)		
– Cooperate	0.023		
	(0.025)		
Sanctioning			0.024
			(0.027)
Social Information × Sanctioning			0.190**
			(0.080)
Empirical Expectation			0.423***
			(.071)
Normative Expectation			
– Do what others do			0.016
			(0.049)
– Cooperate			0.053
			(0.080)
Empirical Exp. × Sanctioning			-0.166
			(0.103)
Normative Exp. × Sanctioning			
– Do what others do			0.069
			(0.075)
– Cooperate			0.056
			(0.131)
Individual Controls	Yes	Yes	Yes
Village Fixed Effects	Yes	Yes	Yes
N	580	580	580

Notes: The table reports average marginal effects from a generalized structural equation model including an individual-level ordered probit regresion in column (1), and individual-level probit regressions in column (2) and (3). Robust standard errors are clustered at village level (in parentheses). All specifications have village fixed effects. Individual controls include age, age-squared, daily earnings, an indicator variable for comprehension, and whether the participant has prior experience in economic experiments. ***, ** and * indicate significance at the 1, 5 and 10% level.

are responsive to the treatment, *i.e.*, whether social information changes normative and empirical expectations. I find that social information has no effect on normative expectations, see column (1) in Table 1.3. The effect is negligible in size and indicates that participants who receive high information, on average, do not have a higher normative expectation than those participants who receive low information, *i.e.*, $a_1 = 0$.[9] In contrast, empirical expectations are significantly affected by social information ($p = .002$), see column (2) in Table 1.3. Those who receive a message about the cooperative behavior of others are on average 8.7 percentage points more likely to also expect them to cooperate, *i.e.*, $a_2 > 0$. Consequently, Hypothesis 1 is accepted for empirical but rejected for normative expectations, *i.e.*, $a_2 > a_1 = 0$.

For the second part of the indirect effect, I test whether the mediating variables are significant predictors of the outcome, *i.e.*, whether differences in normative and empirical expectations change cooperation behavior and whether those changes are conditional on the sanctioning mechanism. The necessary regression is based on equation (1.10) and depicted in column (3) of Table 1.3.

Normative expectations do not transmit the effect of social information on cooperation. Neither when comparing those who stated the belief indicating defection and those indicating cooperation ($p = .509$), nor when comparing defection to those that stated a belief of "do what others do" ($p = .747$) the difference in cooperation behavior is significantly different from zero, *i.e.*, $b_1 = 0$. Independent of whether social information has an impact on normative expectations, a difference does not translate into significant changes in behavior.[10] Therefore, the indirect effect of social information on cooperation transmitted through normative expectation is not different from zero as Hypothesis 1 and 2 are rejected, *i.e.*, $\theta_{X \to M_1} \theta_{M_1 \to Y} = a_1 b_1 = 0$.

In contrast, empirical expectations are a highly significant predictor of cooperation ($p < .001$). Those who expect others to cooperate are on average 42.3 percentage points more likely to cooperate themselves than participants who expect others to defect. Hence, Hypothesis 2 is accepted for empirical expectations, *i.e.*, $b_2 > 0$. Together with the results from column (2), I find a positive and significant indirect effect of social information on cooperation that is transmitted through empirical expecta-

[9] Participants in the high information treatment are on average 2.4 percentage points less likely to hold the normative belief of defection and 2.3 percentage points more likely to state the cooperative belief. Neither difference is statistically significant ($p = .369$ in both cases). The effect on stating the answer "do what others do" is negligible with a difference of 0.2 percentage points ($p = .371$).

[10] The sign and strength of the coefficients for normative expectation on cooperation suggest a small positive effect. Yet, the analysis seeks to identify the effect of social information on cooperation through normative expectation. Due to the very low initial effect of social information on normative expectation, the indirect effect is smaller than 0.5 percentage points.

tions, *i.e.*, $\theta_{X\to M_2}\theta_{M_2\to Y} = a_2 b_2 > 0$. Following from equation (1.1), the effect has a strength of about 3.7 percentage points. That is, high social information leads to an increase in cooperation by 3.7 percentage points solely through inducing cooperative empirical expectations.

To examine Hypothesis 3, attention is shifted to the interaction effects of sanctioning with both normative and empirical expectation. The interaction terms indicate whether the indirect effects are conditional on the sanctioning mechanism. Results suggest that neither indirect effect is moderated by the sanctioning possibility.[11] The behavioral mechanism that links social information treatment to cooperation behavior through differences in the perceived social norm does not vary with the possibility to weakly sanction others' actions. Hypothesis 3 is rejected for both mediators, *i.e.*, $b_{i3} = 0$ for $i = 1, 2$ such that, $(\theta_{X\to M_i}\theta_{M_i\to Y|W=1}) = (\theta_{X\to M_i}\theta_{M_i\to Y|W=0})$ for $i = 1, 2$.

The rejection of Hypothesis 3 does not imply that under sanctioning, the indirect effect of social information on cooperation transmitted through empirical expectations which is established by results regarding Hypotheses 1 and 2, is insignificant. It rather indicates that the indirect effect does not get stronger or weaker when conditioned on the sanctioning possibility.[12] For normative expectation, the insignificance of the interaction with sanctioning similarly implies that the effect of different normative beliefs on cooperation does not depend on whether participants have a sanctioning opportunity. Since there is no initial mediation through normative expectation, its conditional indirect effect is trivially zero as $\theta_{X\to M_1} = 0 \implies (\theta_{X\to M_1}\theta_{M_1\to Y|W=1}) = (\theta_{X\to M_1}\theta_{M_1\to Y|W=0}) = 0$.

Lastly, Hypothesis 4 is accepted as an implication of the results described above, *i.e.*, $a_2 > a_1 = 0$, $b_2 > b_1 = 0$, and $b_{i3} = 0$ for $i = 1, 2$ such that, $\theta_{X\to M_2}\theta_{M_2\to Y} > \theta_{X\to M_1}\theta_{M_1\to Y}$. Independent of the sanctioning possibility, empirical expectations are a better predictor of cooperation in the prisoner's dilemma game and transmit a larger effect of the social information treatment.

[11] p-values are as follows: when comparing the normative belief of defection with the one indicating cooperation, $p = .700$; when comparing the normative belief of defection with the one stating "do what others do", $p = .361$; when comparing the empirical belief of defection with the one indicating cooperation, $p = .108$.

[12] For empirical expectation the coefficient is marginally insignificant ($p = .108$) and negative. Similar to the result in Bicchieri et al. (2020b), one could interpret a negative interaction term as a detrimental effect of an empirical message in combination with the sanctioning mechanism. This is not the case. The negative coefficient only suggests a decrease in the positive effect established by the 42.3 percentage point increase in the baseline effect for empirical expectation to an effect size of approx. 25.5 percentage points, diminishing the total indirect effect through empirical expectations to 2.2 percentage points.

1.5.2 Direct Effect of Social Information on Cooperation

In addition to the indirect effect, the conditional process model allows for a direct effect of social information on cooperation behavior. Therefore, I test whether the treatment directly affects the outcome, *i.e.*, whether the social information message changes cooperation behavior without affecting normative and empirical expectations and whether that change is conditional on the sanctioning mechanism, see column (3) in Table 1.3.

Results indicate that social information leads to an increase in cooperation when moderated by the sanctioning mechanism ($p = .018$), *i.e.*, $c_3' > 0$. When social information is combined with a sanctioning possibility, participants in the high information treatment are on average 19 percentage points more likely to cooperate than those in the low information treatment. The effect is not mediated by any belief about the social norm but arises directly from the information treatment. Without a sanctioning mechanism, the direct effect is insignificant ($p = .380$).

1.6 Discussion

When formal regulations fail to effectively govern collective action problems, the combination of social information and endogenous sanctions is a promising tool to induce desirable behavior. Yet, there is a limited understanding of the underlying behavioral mechanism that links social information to pro-social behavior under sanctioning. Policies that rely on social information as a tool for informal regulation will look to induce conformity but should consider possible pitfalls and limitations. With data from a one-shot prisoner's dilemma experiment that combines high and low social information, a sanctioning treatment, and incentivized belief elicitation, I use a conditional process analysis to study how norm conformity interacts with a sanctioning mechanism to drive behavioral change.

I show that social information and sanctioning opportunity induce a behavioral change in two ways. First, in line with findings by Bicchieri and Xiao (2009), Croson et al. (2009), and Goeschl et al. (2018), I show that social information increases cooperation through a change in empirical expectations, *i.e.*, beliefs about the behavior of others. Independent of the sanctioning mechanism, social information induces norm conformity. Second, I find that social information causes a direct substantial increase in cooperation that is conditional on a sanctioning possibility. Independent of social beliefs, information on the pro-social behavior of others leads to more cooperation

when there is a possibility to sanction or be sanctioned by others.

The conditional process analysis that links social information and cooperation cannot specify the nature of the conditional direct effect apart from it not being mediated by normative or empirical expectations. It should therefore be interpreted as a residual channel that picks up any effect not explained by the causal paths in the model or any of the control variables. That leaves the open question on how the combination of social information and sanctioning opportunity leads to such a substantial increase in cooperation. A likely explanation is that under the threat of sanctions, the social information message gains value for those that want to avoid the moral cost of being labeled as a norm transgressor.[13] Under sanctioning, the social information message becomes a directive for own behavior and points to a behavioral response that can shield oneself from being punished (Cialdini and Trost, 1998).

Fehr and Schurtenberger (2018) and Bicchieri et al. (2020b) present two experiments that also combine the provision of social information with an endogenous sanctioning mechanism. Fehr and Schurtenberger (2018) show that normative priming only leads to sustained cooperation when peer punishment (with a counter-punishment opportunity) is possible. While they study a finitely repeated public goods game, the observed increase in cooperation already occurs in the first period.[14] Although there are important differences with respect to participant pool, the nature of the sanctioning mechanism, and what belief is targeted by the information message, their study is conceptually close to the analysis at hand and comes to a similar conclusion: to successfully leverage social norms, social information messages should be accompanied by an enforcement mechanism.[15]

In contrast, Bicchieri et al. (2020b) find that under certain circumstances the combination of social information and sanctioning mechanism can be detrimental for pro-social behavior. In particular, when the cost of conformity in their modified trust game is high, the combination of sanctioning opportunity and empirical message backfires. The authors conclude that the information is interpreted in a way that

[13] I acknowledge that participants also factor in the material loss of being excluded. Although the sanctioning mechanism does not change the payoff-dominant strategy, the material punishment is part of a utility trade-off between material benefits and moral costs of exclusion.

[14] Interestingly, in the first round Fehr and Schurtenberger (2018) also identify a significant positive effect of social information on cooperation without punishment. The effect disappears with the first repetition and does not reappear within their 15 round game.

[15] Fehr and Schurtenberger (2018) study a western student population, use a decentralized punishment mechanism with counter-punishment opportunity, and provide a normative message while my study looks at Tanzanian fishermen, uses a weak punishment mechanism based on social exclusion, and uses an empirical message to induce behavioral change.

allows the formation of self-serving beliefs which make the message less credible and punishment unjustified. I do not replicate their high-stakes environment which could foster such self-serving biases and there is no indication that participants view the sanctioning as unjustified. After all, the mechanism is based on the exclusion of behavioral strategies instead of individual players and includes the option to not use sanctioning by voting to exclude no one.

A joint implication of Fehr and Schurtenberger (2018), Bicchieri et al. (2020b), and the study at hand is that the behavioral change induced by social information is sensitive to mechanisms of social enforcement. Here, my study contributes by highlighting that it is sufficient to supplement social information with a considerably weak form of sanctioning that (i) is costless, (ii) does not change the payoff-dominant strategy, (iii) is voluntary, and (iv) is based on the exclusion of behavioral strategies. Additionally, the conditional process analysis reveals that the majority of the behavioral change that is induced by social information directly depends on the sanctioning mechanism. Policy makers that use social information to induce desirable behavior may therefore benefit from offering mechanisms of social enforcement to boost the effectiveness of their intervention.

The study opens several avenues for future research. First, the sanctioning opportunity is only revealed after beliefs are elicited. That is, participants cannot consider the social enforcement mechanism when stating normative and empirical expectations. Whether the sanctioning mechanism would influence social beliefs and subsequently increase the indirect effects of social information on cooperation through the perceived social norm should be studied in future research. Second, the design tests the effect of an empirical message. Although an empirical message may carry normative meaning (Bicchieri and Xiao, 2009), it is not surprising that only empirical expectations transmit the effect of social information on cooperation. Would a normative message induce a similar effect when combined with a sanctioning mechanism? Bicchieri et al. (2020b) give an indication that a normative message and sanctions can induce behavioral change, at least when conformity is cheap. Yet, the underlying belief structure is unknown and could be studied to work towards a more complete picture of the interplay between social information, social norms, sanctions, and norm conformity.

Lastly, the methodological approach and the related findings in this study have important theoretical and practical implications for research. By utilizing the conditional process analysis to dissect the behavioral mechanism at work and examine its dependency on an additional treatment, I highlight the importance to explore causal paths via the addition of contextual variables in the framework of social information,

norm conformity, and related topics. Hereby, additional drivers that promote solutions to social dilemmas by initiating behavioral change, but go beyond this study, can be identified.

Bibliography

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.

Andrighetto, G., Brandts, J., Conte, R., Sabater-Mir, J., Solaz, H., and Villatoro, D. (2013). Punish and Voice: Punishment Enhances Cooperation when Combined with Norm-ignalling. *PloS One*, 8(6):e64941.

Baron, R. M. and Kenny, D. A. (1986). The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51(6):1173.

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press.

Bicchieri, C. and Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public choice*, pages 1–22.

Bicchieri, C., Dimant, E., and Sonderegger, S. (2020a). It's not a lie if you believe the norm does not apply: Conditional norm-following with strategic beliefs.

Bicchieri, C., Dimant, E., and Xiao, E. (2020b). Deviant or Wrong? The Effects of Norm Information on the Efficacy of Punishment. ECONtribute Discussion Papers Series 023, University of Bonn and University of Cologne, Germany.

Bicchieri, C. and Xiao, E. (2009). Do the Right Thing: But Only if Others Do So. *Journal of Behavioral Decision Making*, 22(2):191–208.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree - An Open-source Platform for Laboratory, Online, and Field Experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places. *Journal of Personality and Social Psychology*, 58(6):1015.

Cialdini, R. B. and Trost, M. R. (1998). Social influence: Social norms, conformity and compliance.

Costa, D. L. and Kahn, M. E. (2013). Energy conservation âĂIJnudgesâĂİ and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3):680–702.

Croson, R., Handy, F., and Shang, J. (2009). Keeping up with the Joneses: The Relationship of Perceived Descriptive Social Norms, Social Information, and Charitable Giving. *Nonprofit Management and Leadership*, 19(4):467–489.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness. *Economic Theory*, 33(1):67–80.

Diekert, F., Eymess, T., Luomba, J., and Waichman, I. (2021). The Creation of Social Norms under Weak Institutions. Heidelberg University, Department of Economics Discussion Paper No. 684.

Engl, F., Riedl, A., and Weber, R. A. (2020). Spillover effects of institutions on cooperative behavior, preferences, and beliefs. *Preferences, and Beliefs (August 3, 2020)*.

Farrow, K., Grolleau, G., and Ibanez, L. (2017). Social norms and pro-environmental behavior: a review of the evidence. *Ecological Economics*, 140:1–13.

Fehr, E. and Fischbacher, U. (2004). Social Norms and Human Cooperation. *Trends in Cognitive Sciences*, 8(4):185–190.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3):159–181.

Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868):137.

Fehr, E. and Schurtenberger, I. (2018). The dynamics of norm formation and norm decay. *Working paper series / Department of Economics*, (278).

Goeschl, T., Kettner, S., Lohse, J., and Schwieren, C. (2018). From Social Information to Social Norms: Evidence from Two Experiments on Donation Behaviour. *Games*, 9(4):91.

Hallsworth, M., List, J. A., Metcalfe, R. D., and Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148:14 – 31.

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford Press, New York, NY.

Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial Punishment Across Societies. *Science*, 319(5868):1362–1367.

Konow, J. (2000). Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, 90(4):1072–1091.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism. *American Economic Review*, 93(1):366–380.

Schultz, P. W. (1999). Changing Behavior with Normative Feedback Interventions: A Field Experiment on Curbside Recycling. *Basic and Applied Social Psychology*, 21(1):25–36.

Tyran, J.-R. and Feld, L. P. (2006). Achieving compliance when legal sanctions are non-deterrent. *Scandinavian Journal of Economics*, 108(1):135–156.

VanderWeele, T. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, New York, NY.

Xiao, E. and Houser, D. (2011). Punish in Public. *Journal of Public Economics*, 95(7-8):1006–1017.

Chapter 2

The Creation of Social Norms under Weak Institutions

Joint work with Florian Diekert, Joseph Luomba, and Israel Waichman.

Abstract: Preventing the exploitation of natural resources in developing countries is a typical situation where policies have to rely on norm-based interventions to improve outcomes. Yet, little is known about how different social norms can be created in repeated and strategic situations. Our lab-in-the-field experiment (N=588) studies how social information about high or low levels of previous cooperation affects the creation of social norms in a three-player prisoner's dilemma game with/without social sanctioning. Providing different social information succeeds in creating different norms of cooperation, but only if a sanctioning institution is available. When the initial information emphasizes cooperation, cooperation rates start high and stay high and when the initial information emphasizes defection, cooperation rates start low and stay low. Cooperation rates decline without social sanctioning, irrespective of the initial information. Exploring the role of the reference network, we find that initial information is more effective the stronger the social proximity among participants.

Keywords: collective action; common pool resources; lab-in-the-field experiment; social norms

2.1 Introduction

Managing common-pool resources is challenging, in particular under weak institutions. Lack of political will, limited state capacity, and corruption are reasons why governments fail to address problems of pollution, deforestation or overfishing (Ostrom, 2008; Barrett, 2018). A prototype of such a common-pool resource system is Lake Victoria in East Africa. Fisheries from Lake Victoria support the livelihood of four million people, contributing annually about 700 million Euro to one of the poorest regions in the world (Mkumbo and Marshall, 2015; LVFO, 2017). Preventing further depletion of the lake's resources is vital for the region, but efforts to combat overfishing are insufficient (Eggert and Lokina, 2010). In the absence of formal enforcement, resource users have to rely on voluntary efforts and self-management (Baland and Platteau, 1996; Ostrom, 2008).

Social norms are a promising tool to facilitate cooperation (Ostrom, 1990; Nyborg et al., 2016) but little is known about the mechanism and effectiveness of norm-based policies under weak institutions. Researchers face the difficulty that "without a clean empirical identification [...] almost every behavior can be rationalized as norm driven" (Fehr and Schurtenberger, 2018b, p.458). It is hence essential to identify both the factors that induce norm conformity and the relevant norms themselves. Behavioral experiments allow causal inference and are thus powerful instruments to study the effect of norm-based interventions.

This study centers on whether and how different social norms can be created in a repeated and strategic situation. Specifically, we conduct an on-site lab-in-the-field experiment with fishermen at Lake Victoria, Tanzania. In a first treatment variation, participants are given different initial social information that emphasizes either cooperation or defection, nudging participants to have high or low expectations about the cooperation behavior of others in a three-player prisoner's dilemma game. In a second treatment variation, we vary whether participants can use a coordinated weak sanctioning institution that mimics social enforcement in the field. The strategic situation is repeated to test whether different patterns of behavior and beliefs emerge, forming different social norms. Finally, we relate cooperation and expectations in the experiment to participants' social proximity to other fishermen in their reference network at Lake Victoria.

The literature on norm-based interventions suggests that providing social information can induce behavioral change (Frey and Meier, 2004; Shang and Croson, 2009). For example, social norms are found to be an effective policy tool in promoting voting

behavior (Gerber and Rogers, 2009), retirement savings (Duflo and Saez, 2003) and tax compliance (Hallsworth et al., 2017). Importantly, Allcott (2011) and Allcott and Rogers (2014) demonstrate the cost-effectiveness and simple implementation of norm-based interventions on a large scale.[1] Several lab experiments focus on the underlying mechanism of *how* social information influences behavior. Building on theories of norm conformity (Bénabou and Tirole, 2006; Kimbrough and Vostroknutov, 2016), the findings so far suggest that social information affects the perception of a social norm and prompts the desire to conform (Bicchieri and Xiao, 2009; Ferraro and Price, 2013; Goeschl et al., 2018).

Conformity depends on the behavior and opinion of others. Elster (1989, p.99) highlights the social element when distinguishing *social norms* from *moral norms*: "For norms to be social, they must be shared by other people and partly sustained by their approval and disapproval." We therefore combine social information with a mechanism of social enforcement. That is, we experimentally vary whether participants have the opportunity to show their disapproval about the behavior of others. This sanctioning institution is implemented by giving participants the option to vote on which strategy should be excluded from receiving a small financial bonus. Crucially, the financial bonus does not change the Nash equilibrium in the game, so that our study ties into the literature on weak sanctioning (Masclet et al., 2003; Tyran and Feld, 2006) and mimics informal ways of norm enforcement at Lake Victoria.[2]

We define a social norm as a stable pattern of behavior that is supported by a consistent set of beliefs about others. The definition has two implications. First, it means that "social norms cannot be identified just with observable behavior, nor can they merely be equated with normative beliefs" (Bicchieri et al., 2018). Second,

[1] The Opower Home Energy Report intervention studied by Allcott (2011) and Allcott and Rogers (2014) provide households with empirical information about energy consumption in their neighborhood. Motivation to conduct a large-scale norm-based intervention on energy conservation originates from a small set of field studies by Schultz et al. (2007), Goldstein et al. (2008), and Nolan et al. (2008).

[2] Ostrom (2000, p.149) writes: "a frequent finding is that when the users of a common-pool resource organize themselves to devise and enforce some of their own basic rules, they tend to manage local resources more sustainable than when rules are externally imposed on them", and further on p.151, "sanctions that are imposed are often so low as to have no impact on an expected benefit-cost ratio of breaking local rules (given the substantial temptations frequently involved)." While peer-to-peer punishment is found to be cooperation-enhancing in the lab (Fehr and Gächter, 2000, 2002), such a possibility of high impact punishment is not often observed in the field (Guala, 2012). Rather, some form of (not very strong) coordinated punishment is usually applied (Fehr and Schurtenberger, 2018b). A concrete example could be the exclusion from playing pool, a popular leisure activity at the shores of Lake Victoria. For an anthropological account of the social ties that form an important part of many fishermen's lives see Beuving (2010).

it means that repetition is important: A social norm is a dynamic concept (Sethi and Somanathan, 1996; Binmore, 2010; Young, 2015) that not only describes a static equilibrium in the sense that each agent prefers a given action conditional on her beliefs, but also an equilibrium in the dynamic sense that behavior and beliefs are stable and do not change over time. Therefore, we elicit normative beliefs and track participants' empirical expectations and behavior over the course of repeated one-shot interactions. The successful creation of social norms in our experiment would then show as a stable pattern of cooperation and a corresponding pattern of beliefs.

In the last decades, a plethora of approaches to induce and maintain cooperation have been tested. Following seminal contributions by Ostrom et al. (1992) and Fehr and Schmidt (1999), a number of lab-in-the-field experiments show that free-riding is reduced when cooperation can be socially enforced (Carpenter et al., 2004; Alpizar et al., 2008; Janssen et al., 2010; Rustagi et al., 2010; Hayo and Vollan, 2012). In particular, Cardenas (2011) and Lopez et al. (2012) point out that a message that highlights full contributions as the socially optimal strategy increases cooperation only when defection may be publicly exposed. Here, we combine the use of a social sanctioning institution with the provision of social information. Our focus is not so much on increasing cooperation in the experiment per se, but on whether social information (alone, or in combination with social sanctioning) creates different social norms of cooperation.

The only study that investigates the relationship between a social information message and a form of social enforcement in a repeated setting is Fehr and Schurtenberger (2018a), who (parallel and unbeknownst to us) conducted a novel lab experiment on social norms of cooperation.[3] They use a public goods game and employ a 2x2 design where they (i) either do or do not provide a normative social priming, and (ii) either do or do not enable peer punishment (Fehr and Gächter, 2000, 2002) with a counter punishment possibility (Nikiforakis, 2008). They find that without punishment and irrespective of the social priming, cooperation deteriorates over time. However, when punishment and counter punishment are allowed, cooperation is stabilized in both treatments (but with a higher cooperation level when social priming was provided; see also Fehr and Schurtenberger, 2018b, Fig. 3).

[3]Bicchieri et al. (2020) present results from a trust game in a one-shot setting. They distinguish whether participants receive information about what others think the trustee ought to do, what selected trustees actually did, or no information under the presence or absence of an exogenous weak punishment institution. They find that normative information may raise the return to the trustor, but only under weak punishment, highlighting the moderating role of enforcement on norm-based interventions.

Our results are strikingly similar to those of Fehr and Schurtenberger (2018a), albeit with a participant pool of actual resource users instead of students and a different enforcement mechanism. First, we observe that the social information message has no effect on cooperation without social sanctioning. Cooperation rates decline over time, independent of whether social information emphasizes cooperation. Second, the social information message leads to significant differences in behavior with social sanctioning: When cooperation is emphasized, cooperation rates start high and stay high. In contrast, when defection is emphasized, cooperation rates start low and stay low.

Moreover, we also elicit beliefs, revealing that the differences in behavior between the treatments are accompanied by parallel differences in empirical expectations. In the social sanctioning treatment that emphasizes cooperation, cooperation rates are high and the majority expects others to cooperate. In contrast, in the social sanctioning treatment that emphasizes defection, cooperation rates are low and the majority expects others to defect. Without social sanctioning, emphasizing cooperation or defection does not produce significant differences in beliefs. Thus, the social sanctioning institution appears as a necessary condition for a stable pattern of behavior. Different initial information then leads to a self-enforcing alignment of beliefs and behavior – different social norms are created.

Research in economics, sociology, and psychology has highlighted the role of the reference network and social proximity to explain conformity with social norms (Elster, 2007; Bicchieri et al., 2019; Dimant, 2019).[4] The beliefs and behavior of others matter more for individual actions when "others" refers to a group that is relevant to the respective actor. With reference to the literature on natural resource use that connects the lab and the field (*e.g.*, Carpenter and Seki, 2011; Fehr and Leibbrandt, 2011; Kosfeld and Rustagi, 2015) we use the random treatment assignment across individual characteristics to study an individual's social proximity to others as an additional social dimension that may facilitate the creation of social norms. In particular, we measure a participant's social proximity to the others in a session by eliciting whether a fisherman belongs to the session's majority with respect to gear type (target species) and region of origin (ethnicity), two defining features of the social structure in fishing

[4]The concept of social proximity strongly relates to social identity and group affiliation. The discussion of social identity and group affiliation in economics goes back to influential work by Tajfel and Turner (1979), and Akerlof and Kranton (2000). The idea of collective identity also relates to the concept of ingroup favoritism and outgroup hostility (Goette et al., 2006; Meier et al., 2016). Further findings imply that conformity to shared norms of behavior is greater with high levels of social affiliation (Platteau, 2006). Not only do close-knit communities share an understanding through their common identity, but strong social affiliation may be a consequence of existing group norms that govern beliefs and attitudes of community members (Mason, 2006).

communities at Lake Victoria (Nunan et al., 2018). We find that those participants with close social proximity to the others in their session drive the social information treatment effect: They cooperate more when social information emphasizes cooperation and cooperate less when social information emphasizes defection.

2.2 Lake Victoria Fisheries

Lake Victoria, see Figure 2.1, is the largest lake in Africa and its resources are shared among three nations (Kenya, Tanzania, and Uganda). Despite the existence of intergovernmental structures such as the Lake Victoria Fisheries Organization, national strategic incentives limit coordinated and effective measures to curb overfishing. Within each country, weakened state capacity hampers monitoring and enforcement of the fishing regulations that aim to promote sustainable resource use. Under these weak institutions, non-compliance is common (Eggert and Lokina, 2010). Therefore, regulators should focus on compliance as the most needed form of cooperation at Lake Victoria. Promoting resource users' own cooperative efforts and community management is key to overcome the social dilemma that characterizes the use of the lake's resources.

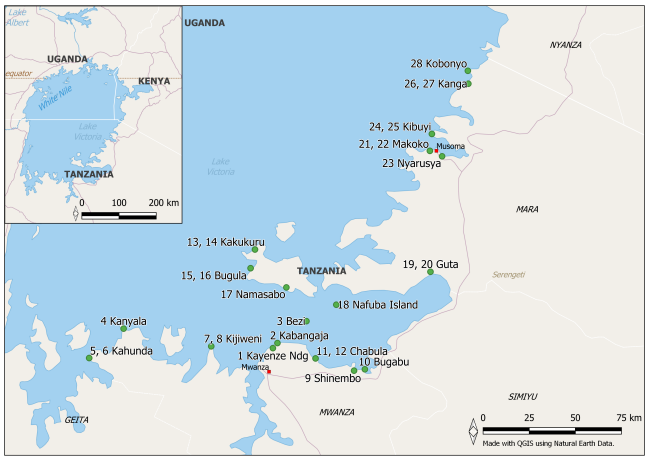


Fig. 2.1: Map of field research sites, Lake Victoria, Tanzania

In order to prevent further depletion of the fish stock and to encourage community participation, the Tanzanian government established local co-management structures known as beach management units (BMU). BMUs were introduced in 1998 and have been helpful in reducing the use of poison and dynamite but their overall effectiveness is unclear (Eggert and Lokina, 2010; Nunan and Onyango, 2017). Corruption and

kinship ties between BMU officials and fishermen make effective monitoring and enforcement difficult. Yet, BMUs are a forum for exchange and feedback at the landing site level (Luomba, 2013). This offers a potential springboard for policies that aim at changing social norms of cooperation in an environment where state institutions are weak.

Moreover, the social structure in a community plays an important role for successful management of common pool resources (Mosimane et al., 2012). Crona and Bodin (2006), Barnes et al. (2016), and Nunan et al. (2018) identify specific determinants of social cohesion in fishing communities. Two features that are particularly relevant at Lake Victoria are (1) differences in the region of origin which partly reflect ethnic differences, and (2) a fisherman's main target species which essentially splits fishermen into different life and work routines. Dagaa fishermen work at night and use solar or kerosene powered lights and small meshed seine nets, while Nile perch and Tilapia fishermen use hooks or large meshed nets to fish during the day. Moreover, dagaa is mainly sold to local and regional markets while Nile perch are collected by processing plants and exported to the world market.[5]

2.3 Experimental Design and Procedures

In the following, we describe an experiment designed to test whether a social information message can create different social norms of cooperation with or without a coordinated weak sanctioning institution. In particular, we use a factorial design where one factor is an information message on either high or low cooperation rates of previous participants in a similar setting. The other factor is a sanctioning institution, implemented as the opportunity to vote on what type of behavior should be excluded from receiving a small financial bonus.

We refer to the treatment where the sanctioning institution is available and the social information message emphasizes high cooperation as hi-S treatment. The corresponding social information treatment without a sanctioning institution is referred to as hi-noS treatment. In parallel, low-S and low-noS refer to the treatments with and without social sanctioning when the social information message emphasizes low cooperation.

[5]Interestingly, Jang and Lynham (2015) show that differences in the typical contractual sharing agreements between boat owners and crews in the dagaa and Nile perch fisheries in Kenya translate into differences in sharing behavior in the ultimatum game.

The Social Dilemma Game

Our vehicle to model the social dilemma is a three-player prisoner's dilemma. It is a simultaneous-move game such that participants cannot condition their action on the observed behavior of others. The game is played for a total of seven rounds with a perfect-stranger matching protocol. The protocol prevents directly reciprocating past behavior and precludes forward looking motivations for specific actions.[6]

Participants are randomly divided into groups of three. Each participant receives an endowment of four points that she can either allocate to a private account (hereafter "defect") or to a group account (hereafter "cooperate"). Only the full endowment can be transferred to either account. For every group account contribution, all three group members earn two points. Table 2.1 shows the individual payoff matrix.

Table 2.1: Individual payoff matrix

		Number of other cooperators		
		0	*1*	*2*
Own decision	*cooperate*	2pt	4pt	6pt
	defect	4pt	6pt	8pt

Social Information Message and Belief Elicitation

Prior to the allocation decision in the three-player prisoner's dilemma game (but after participants are fully informed about the rules), we elicit the participants' normative beliefs, normative expectations, and empirical expectations (Bicchieri, 2017). Before the belief elicitation, participants receive a verbal social information message, see Table 2.2.

Our message is designed to affect empirical expectations in the direction of either cooperative (high) or defective (low) behavior. The phrasing utilizes the word "many", making our social information manipulation very subtle. Moreover, we deliberately do not want to persuade participants that one or the other action is better. That is, we address *descriptive norms* (what do most others do) instead of *injunctive norms*

[6]Seven rounds is the maximum number of repetitions that still allows a perfect-stranger matching in groups of three when there are a total of fifteen participants in the pool. We chose this threshold because we did not know ex-ante that we would be successful in recruiting 21 participants in every session.

Table 2.2: Social information treatments

Treatment	*Message*
High information	In a previous survey, it was found that many participants chose to put the points to the group account and not to put them to their private account.
Low information	In a previous survey, it was found that many participants chose to put the points to their private account and not to put them to the group account.

(a prescription of what one ought to do, see Cialdini et al., 1990).[7]

In practice, descriptive and injunctive norms may not be fully separable. For example, participants could take the social information message as a signal of what is the right action in the eyes of others and re-evaluate their own moral belief. Therefore, we ask participants what they think one ought to do in this situation, *i.e.*, their *personal normative beliefs*. Moreover, we ask participants what they think most others think one ought to do, *i.e.*, their *normative expectations*. Participants can choose from a menu of three options. They can either state that one ought to (i) "put the points to the private account", (ii) "put the points to the group account", or (iii) "do what others do". Next, we repeat the social information message and ask participants what they think others will actually do, *i.e.*, their *empirical expectation*. Here participants can reply that (i) most other participants will allocate the endowment to the private account, or (ii) most other participants will allocate it to the group account. Both *normative expectations* and *empirical expectations* are incentivized by offering an extra point for correct prediction.

[7]Note that we did not lie to the participants. We have visited the communities in this study about a year before and conducted a survey that included the incentivized decision to put points in a private or group account. Across the whole study, many participants have put the points in the group account, and many participants have put the points in their private account. More specifically, in some communities, the majority of participants cooperated and in other communities, the majority did not cooperate. Since participants were re-sampled with a similar experimental setup (using the same tablets, logos, etc.) and the same research team, it was evident that the term "in a (the) previous survey" (the Swahili "Katika utafiti uliopita" could be both translated as specific or unspecific) referred to our last visit. Acknowledging that there are different perspectives on where to draw the line with respect to deception and manipulation in experimental economics (see Barrera and Simpson (2012) and Rousu et al. (2015) for insights into the debate), we argue that our attempt to shift empirical expectations by providing selective information is acceptable. Indeed, the crux of manipulating beliefs to study their causal effect on behavior is most commonly overcome by providing a factually accurate, but not necessarily representative statement about a previous sample, see for example Frey and Meier (2004); Bicchieri and Xiao (2009) or Fehr and Schurtenberger (2018a).

Coordinated Weak Sanctioning

To model social norm enforcement, we design a mechanism that reflects the informal institutions available to local communities. Specifically, each participant in a group may receive a bonus of one point. Prior to that, each participant casts a vote that either (i) those who allocated their endowments to their private account shall be excluded from receiving the bonus, (ii) those who allocated their points to the group account shall be excluded, or (iii) that no one shall be excluded from receiving the bonus. The exclusion rule is determined by majority, *i.e.*, the rule receiving two or three votes within a group is applied. If each alternative receives exactly one vote (no majority is reached), no member is excluded from receiving the extra point.

First, being based on the exclusion of group members from a financial bonus of one point, our sanctioning institution simulates the kind of mild ostracism observed in the field (Beuving, 2010; Guala, 2012). Second, the sanctioning institution has the character of giving a general comment on the behavior of others instead of a directed personal punishment. Participants vote on which strategy should be excluded from receiving the bonus point, and participants vote before they know the contribution decisions of their group members.

In the treatments with social sanctioning, participants are informed about the voting mechanism after beliefs are elicited, but before they make their first contribution decision. The voting procedure itself takes place after contribution decisions are made but before the actual choices of the group members are revealed. After voting, participants are informed about the allocation decisions by the three group members, the exclusion rule that was selected, and who was excluded from receiving the bonus. In the treatments without social sanctioning, the voting mechanism is not mentioned at any point during the game. After their allocation decision, participants see a report on the allocation decisions of all group members and the total points earned by themselves.[8]

Repetition

After participants finish the (one-shot) procedure explained above they are informed that the experiment will continue for six additional rounds. We employ a perfect-stranger matching, highlighting that "you will be matched with two other participants from this session that have never been in your group before and will never be in your

[8] An English translation of the instructions and screenshots of the choice situations are available in supplementary material published online.

group again". The perfect-stranger protocol is easy to communicate. Importantly, it neutralizes direct reciprocity and forward-looking motivations that participants may have for choosing a specific action. Moreover, before the allocation decision in each of the additional six rounds, we elicit participants' empirical expectations. We do not repeat the social information message because once participants gain experience in the game itself, further messages may contradict observed behavior and jeopardize the message's credibility.

Implementation

The experiment was conducted in 20 villages with a total of 28 sessions in the Lake Victoria region of Tanzania between February and March 2018, (see Figure 2.1). The sample is balanced with seven sessions per treatment. Each session comprised 21 participants such that the total number of participants is $N = 588$.[9]

The experiment took place in a community center in the village or directly at the landing site. To begin, the general rules of the session were explained and we obtained informed consent. Afterwards, tablets were distributed and participants familiarized themselves with the device by completing brief handling exercises. Participants were guided through the social dilemma game step by step, using posters for visualization and requisites for illustration of the game's mechanics. In particular, it was explained that decisions during the game would translate to real money dependent on own choices and choices of other group members. All decisions were made anonymously and no communication between participants was allowed during the session. Carton shields ensured privacy. Participants were not able to identify their group members, neither during the experiment nor afterwards. Comprehension of the game's rules was tested with the help of four different scenarios that asked participants to name the correct payoffs after specific decisions were made by all group members. We control for participant's comprehension in the analysis.

Upon completion of the experiment, a volunteer was asked to roll a die in order to determine which of the six repeated rounds would be paid out in addition to the one-shot round. Subsequently, we used the incentivized lottery-choice task by Gneezy and

[9]The current experiment is the second field trip to these communities as part of a longer project. During the first field trip to the same landing sites in 2017, participants were invited based on a random draw from the lists of registered fishermen at the respective landing site. Crew members, boat owners or fishing agents were all eligible to participate in the experiment. If a list of registered fishermen was not available we over-invited a convenience sample and randomly retained 21 participants. In the current experiment, we had the aim of re-sampling participants from the first field trip (we achieved a re-sampling rate of of just under 50%), and then used above procedure to complete sessions.

Potters (1997) to measure participants' risk aversion.[10] Participants were endowed with six points and selected how much of it to invest in a risky option with a 50% chance to lose their money and a 50% chance to triple the amount. Participants then filled out a short questionnaire on demographics, compliance to fishing regulations, official management of the landing site, and socioeconomic background. After an average of 2 hours, sessions ended with a private payout.[11]

Participant Characteristics

Table 2.3 provides an overview of the demographic characteristics of the participants.

Table 2.3: Participants characteristics

Variable	mean	sd	min	max
Age	37.95	12.03	18	93
Female	0.02	-	0	1
0 - 5,000 TZS daily earnings	0.45	-	0	1
Never moved	0.61	-	0	1
Crew size	3.79	0.84	1	6
Main gear: dagaa net	0.39	-	0	1
Risk aversion	2.9	2.51	0	6

Among all 588 participants, there are only 12 women, illustrating that the fishing sector at Lake Victoria is heavily dominated by men. The fishermen in our experiment are, on average, about 38 years old and 72% see themselves still being a fisherman in two years time. Daily earnings are low with about 45% of all participants reporting an income below 5,000 TZS (ca. US$ 2.20) per day. Over 60% of participants state that they have always lived at the respective landing site. The average crew size in our sample is 3.79 with a median value of 4. Approximately 60% of the participants work as crew and 23% report to be a boat owner. 39% of the participants target dagaa, while 22% use gillnets and 35% use hooks to target Nile perch or Tilapia. The last row of Table 2.3 shows that participants invested, on average, 2.9 out of six tokens in the risky option.

[10] As risk aversion is likely correlated with the decision to cooperate (Bohnet and Zeckhauser, 2004; Schechter, 2007), we control for it in the regression analysis.

[11] The average payout was just above 5,000 TZS (ca. US$ 2.20) which is the median daily income in our sample. The minimum payout was set to 2,500 TZS.

Social Proximity Measure

Scholars that study the role of social norms emphasize the importance of the reference network for conformity (Elster, 2007; Bicchieri, 2017). We draw from the sociological and anthropological literature on social structures in fishing communities (Crona and Bodin, 2006; Barnes et al., 2016; Nunan et al., 2018) to construct a measure of a participant's social proximity to the other participants in a given experimental session. While the prisoner's dilemma game is played with full anonymity, the set of 21 participants in one session live in the same community and know each other. Hence, the individual participant is able to infer how socially close she is to all the other participants that may be matched with her during the game. Specifically, we consider a participant to be "close" to the others in the session when she (*i*) targets the same species and (*ii*) comes from the same region as the majority of participants in a given session. These two dimensions do not fully encompass the rich structure that defines communities at Lake Victoria but it reflects important dimensions of a fisherman's social network: tribal and kinship culture as well as knowledge sharing with respect to resource use.

Based on these two questions, we construct an index that measures how close a given individual is to the typical participant in the session (social proximity, sp_i). The index can take three values: If the participant is active in the same fishery as the majority of other participants in the session *and* comes from the same region as the majority of others in the session, we set sp_i=1. If the participant is active in a different fishery *and* comes from a different region than the majority of others in the session, we set sp_i=0. Finally, if either the participant is active in the same fishery *or* comes from the same region as the majority of others (but not both), we set sp_i=0.5. We plot the distribution of the underlying data and provide an extended description of the social proximity measure in Appendix A-3.

2.4 Hypotheses

To derive Hypotheses, we first discuss standard preferences and then proceed to norm-based preferences. The individual payoff matrix (Table 2.1) illustrates that defection is the dominant strategy. Under standard preferences, the dominant strategy is neither changed by the social information message nor by the social sanctioning institution or one's social proximity to others. First, defection maximizes own payoff irrespective of the (induced) beliefs about others' behavior or considerations on who the others are.

Second, our social sanctioning institution is "weak" (see *e.g.*, Tyran and Feld, 2006; Bicchieri et al., 2020): The gain from defection is two points, while the bonus is only one point so that the dominant strategy is still to defect, even when an agent expects to be sanctioned with certainty. Hence, standard neo-classical theory predicts that agents will put their points to the private account in all treatments.

An alternative is to draw predictions from norm-based theories. Theories that describe how social norms affect utility often postulate that agents experience disutility or discomfort when choosing an action that does not conform to what they expect others to do. Equation (2.1) illustrates the mechanism (see, *e.g.*, Kimbrough and Vostroknutov, 2016; Michaeli and Spiro, 2017; Fehr and Schurtenberger, 2018b, for similar setups).

$$u(x_i, x_{-i}, ee_i) = \pi_i(x_i, x_{-i}) - \gamma_i \cdot d(|x_i - ee_i|) \tag{2.1}$$

Utility depends on an idiosyncratic component π_i that contains the material payoff from the game, as well as any moral calculations that compare a given action x_i with the personal normative belief about what is the right thing to do. The discomfort function d is increasing in the difference between the agent's own action x_i and her empirical expectation ee_i. The weight γ_i on the discomfort d differs from individual to individual. In particular, we expect that it is larger for individuals in the sanctioning treatments than in the no sanctioning treatments. Moreover, we expect that γ_i is larger for individuals with close social proximity to the other participants in their session.

The social information message is designed to affect empirical expectations and we expect agents to form their expectations accordingly. As equation (2.1) illustrates, agents that expect others to cooperate are more likely to cooperate themselves to avoid the discomfort of non-conformity.

Hypothesis 1a Initial empirical expectations about cooperation rates are higher in the hi- treatments than in the low- treatments.

Hypothesis 1b Average cooperation rates are higher in the hi- treatments than in the low- treatments.

Actual or anticipated sanctions then increase the discomfort from not conforming to ee_i. This is captured by an increased γ_i in equation (2.1). Thus, when the social information message emphasizes cooperation, and if it affects empirical expectations accordingly, we expect that agents cooperate more in the hi-S treatment than in the hi-

noS treatment. Conversely, when the social information message emphasizes defection, and if it affects the empirical expectations accordingly, we expect less cooperation in the low-S treatment than in the low-noS treatment. In other words, the effect of the sanctioning institution is not monotone, but depends on the respective message.

Hypothesis 2 The average cooperation rate is higher in the hi-S treatment than in the hi-noS treatment and lower in the low-S treatment than in the low-noS treatment.

Note that Hypothesis 2 stands in contrast to predictions that are based on preferences for the social optimum. Agents who value an outcome that maximizes group payoffs higher than conformity with empirical expectations would use the sanctioning institution to increase cooperation also in the low-S treatment, yielding lower cooperation rates in the low-noS treatment than in the low-S treatment.

We operationalize our definition of a social norm by requiring (*i*) a stable pattern of behavior, and (*ii*) a corresponding pattern of beliefs. By studying the evolution of both cooperation and empirical expectations in the repeated game, we can reject social norms as an explanation for behavior when conditions (*i*) and (*ii*) are not met. Without social sanctioning, the forces that push an agent's action towards empirical expectations are weak. There are only internal consequences (*e.g.*, guilt) for non-conformity. However, when there is a sanctioning institution, agents perceive a social consequence to non-conformity. Since the message introduces a social behavior that one can conform to, we expect that the initial cooperation rate aligns with agents' empirical expectations. Consequently, agents see their empirical expectations confirmed and act accordingly in the next round of the game: a self-fulfilling prophecy of either high or low cooperation.

In sum, we predict that conditions (*i*) and (*ii*) are met in the two treatments with social sanctioning. The variation in the social information treatment induces social norms at two different levels.

Hypothesis 3 Behavior and beliefs are stable and align in the -S treatments: different social norms are created.

The sanctioning institution increases the consequence of non-conformity, but the punished behavior is not publicly exposed. Therefore, anticipated sanctions do not necessarily induce a feeling of shame in those that violate the normative prescription (Elster, 1989; Lopez et al., 2012; Schram and Charness, 2015). Additional forms

of social impact may be required to induce and maintain conformity. While group membership in the prisoner's dilemma is anonymous, participants do know that they are matched with and potentially sanctioned by some of their peers. The discomfort of non-conformity and anticipated sanctions then additionally depends on the social proximity to others in the session.

Hypothesis 4 Close social proximity is associated with more cooperation and higher expectations in the hi-S treatment and less cooperation and lower expectations in the low-S treatment.

Beyond anticipated sanctions and social proximity, the discomfort from not conforming with the expected actions of others may depend on several other motivations. Those include a preference for conformity as such, an intrinsic motivation for fairness, the guilt of disappointing or harming others, or a preferences for reciprocity. While each of these motivations may be at work to some extent, we do not aim to isolate their effect.

2.5 Experimental Results

We present our results in three steps. First, we turn to behavior and beliefs over the course of the experiment (Section 2.5.1). Second, we study the use of the social sanctioning institution (Section 2.5.2). Third, we explore the role of the participant's social proximity to their reference network (Section 2.6).[12]

2.5.1 The Evolution of Cooperation

Figure 2.2 shows cooperation rates over time in the four treatments. Without sanctioning, the social information message appears to have no effect. Irrespective of which behavior is emphasized (hi-noS or low-noS), cooperation rates decline over the course of the repeated game. With social sanctioning stable cooperation rates are maintained. Cooperation rates start high and stay high when the initial message emphasizes cooperation (hi-S), and cooperation rates start low and stay low when the initial message emphasizes defection (low-S).

To support the descriptive findings, we estimate a non-linear panel data model of binary cooperation decisions on the treatment variation and a time trend. The

[12] A companion paper (Eymess, 2021) focuses on the one-shot segment of the experiment. Our estimation strategy is described in detail in Appendix A-1.

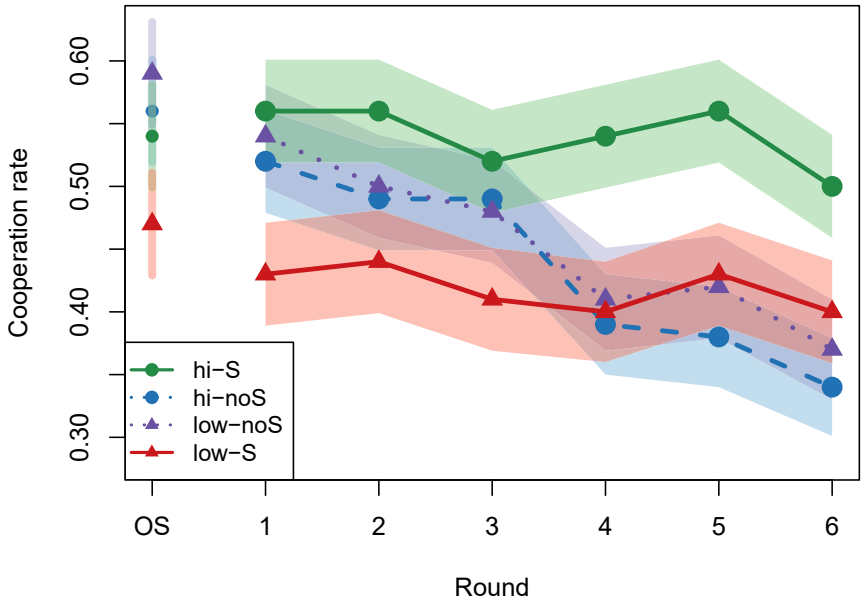


Fig. 2.2: Cooperation rates (averaged on the treatment level) over the game. The one-shot game is indicated by OS and repeated rounds by their number. The shaded area indicates ±1 SE.

regression results are shown in column (1) of Table 2.4 with hi-S as the baseline treatment. We find significantly less cooperation in the low-S treatment than in the hi-S treatment ($p<0.01$). Moreover, both treatments without social sanctioning exhibit a significant round effect; cooperation erodes with each repetition of the social dilemma game ($p<.01$). The unraveling of cooperation is not observed in the sanctioning treatments; cooperation rates are stable over rounds. Also, the cooperation rate is significantly lower when the social information message emphasizes defection rather than cooperation. Results are in line with non-parametric tests and robust to the exclusion of the last round, exclusion of the one-shot procedure (first round), to dropping participants that failed the comprehension tests, controlling for prior experience with lab-in-the-field experiments, and to choosing a linear probability model instead of the non-linear probit model (see Appendix A-2).

In sum, we find partial support for Hypothesis 1b: While average cooperation is higher in the hi-S treatment than in the low-S treatment, there is no difference between the cooperation rates in the hi-noS and the low-noS treatment. Similarly, we find partial support for Hypothesis 2: While starting from the same level, the average

Table 2.4: Behavior and beliefs: Individual level

	Cooperation	Empirical Expectation
	(1)	(2)
low-noS	-0.027	-0.686**
	(0.315)	(0.287)
low-S	-0.862***	-1.024***
	(0.237)	(0.236)
hi-noS	-0.180	-0.306
	(0.276)	(0.228)
Round	-0.025	-0.025
	(0.032)	(0.032)
low-noS × Round	-0.173***	-0.022
	(0.053)	(0.044)
low-S × Round	-0.018	0.022
	(0.041)	(0.036)
hi-noS × Round	-0.182***	-0.072**
	(0.049)	(0.036)
Session Fixed Effects	Yes	Yes
Individual Controls	Yes	Yes
N	4116	4116

Notes: The table reports random effect estimators for an individual level probit model. The baseline is set to the hi-S treatment. The model includes the one-shot game and all repeated rounds. Robust standard errors are clustered at the session level (in parentheses). Individual controls include age, age squared, an indicator variable for comprehension, a risk preference measure, type of main gear, and crew size. ***, **, and * indicate significance at the 1,5, and 10% level.

cooperation rate in the last rounds is higher in the hi-S treatment than in the hi-noS treatment and the average cooperation rate in the low-S treatment is lower than in the low-noS treatment in the early rounds, but ends at about the same level.

Next, we study participants' belief structure and analyze personal normative beliefs, normative expectations, and initial empirical expectations. Recall that the belief elicitation is conducted after the provision of social information, but before the one-shot game. Moreover, participants are not yet informed about the voting mechanism in the low-S and hi-S treatments when stating their initial beliefs. Hence, we only distinguish between the low and high social information treatment.

The left panel in Figure 2.3 shows the distribution of personal normative beliefs. Each of the options, to "put the points to the private account", to "do what others do", and to "put the points to the group account" is chosen by about one third of the participants for both social information messages. This pattern suggests that participants do not perceive a clear moral difference between contributing to the private or the group account.

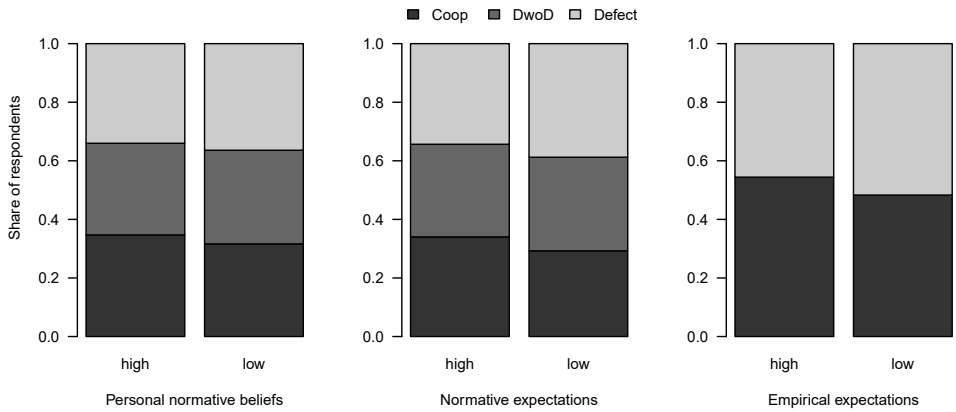


Fig. 2.3: Elicited Beliefs by Information Treatment. Left panel shows personal normative beliefs after the high and low social information message. The dark gray area shows the share of fishermen that think the right thing to do is to "put the points in the group account" (Coop). The medium gray area shows the share of fishermen that think the right thing is to "do what others do" (DwoD). The light gray area shows the share of fishermen that think the right thing is to "put the points in the private account" (Defect). Middle panel shows normative expectations, and right panel shows empirical expectations (where "do what others do" was not an option).

The middle panel in Figure 2.3 shows the distribution of normative expectations. Similar to personal normative beliefs, each option is chosen by about one third of the participants. This suggests that the provision of social information did not affect injunctive norms, at least not strongly.

The right panel in Figure 2.3 shows the distribution of initial empirical expectations. We observe a marginally significant difference in empirical expectations between the high- and low social information treatments (p=0.08, one-sided test of proportions, N_{low}=N_{high}=294). In particular, 54% of our participants expect others to cooperate in the high information treatment, while this proportion is 48% with low social information. Note that although the effect appears to be a small, it changes the expectation of the majority. Since the sanctioning institution is based on majority voting, the majority expectation may be crucial.

With respect to Hypothesis 1a, we find weak evidence that initial empirical expectations about cooperation are lower in the low- treatments than in the hi- treatments. Considering the clear differences in cooperation behavior, the small difference of initial beliefs is remarkable. It suggests that if different social norms emerge over the course of the experiment, these norms were created by the intervention and not imported

into the lab. We turn to the evidence on the emergence of different social norms next.

To study the creation of social norms, we focus on the operationalization stated in Section 2.4. A social norm requires (*i*) a stable pattern of behavior, and (*ii*) a corresponding pattern of beliefs. Cooperation declines over time in both -noS treatments, violating condition (*i*). No social norm is created. In the two -S treatments, we do observe a stable pattern of behavior (condition *i*). The question is whether we also observe a corresponding pattern of beliefs (condition *ii*).

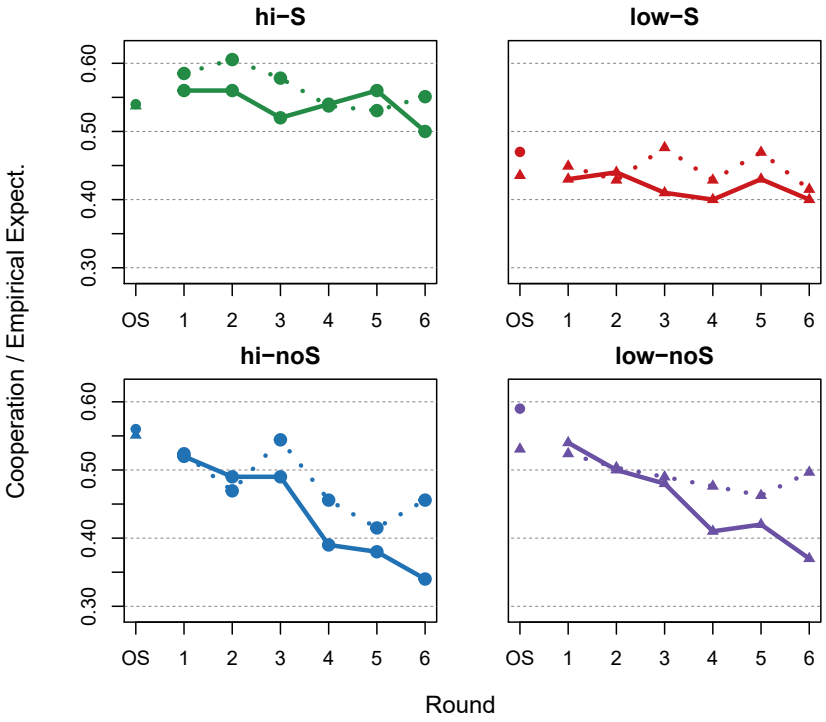


Fig. 2.4: Evolution of empirical expectations. Dotted lines show average empirical expectations, solid lines average cooperation rates. The one-shot game is indicated by OS, repeated rounds by their number.

Figure 2.4 plots average cooperation and empirical expectations over the course of the experiment for all four treatments. First, we find a marked difference in the levels of empirical expectations between the hi-S and low-S treatment. In particular, the majority of participants in the hi-S treatment expect others to cooperate, while the majority in the low-S treatment expect others to defect. Second, we find that, similar to the average cooperation rate, the average empirical expectation is stable in the two -S treatments. In the two -noS treatments, the average cooperation rate declines, and a gap to average empirical expectation opens up. Accordingly, we observe

that participants' empirical expectations in rounds 4-6 of the -noS treatments are less accurate than the empirical expectations of participants in the -S treatments ($p < 0.01$).[13]

We use a non-linear panel data model to formally establish that empirical expectations follow the same pattern as cooperation behavior. We regress empirical expectations (column (2) of Table 2.4) on the same set of co-variates as cooperation (column (1) of Table 2.4). The difference in cooperation behavior between the low-S and hi-S treatment is accompanied by parallel differences in empirical expectations. Participants have significantly lower empirical expectations in the low-S treatment ($p<0.01$). Moreover, there is no round trend for empirical expectations in the sanctioning treatments.

In sum, we confirm Hypothesis 3: Behavior and beliefs are stable in the treatments with a social sanctioning institution. This is not the case in the treatments without social sanctioning. In other words, a norm-based intervention that relies on the provision of social information alone is not sufficient: a social sanctioning institution is necessary for the creation of different social norms.

2.5.2 The Role of the Sanctioning Institution

To understand the role that coordinated weak sanctioning plays in creating social norms, we first investigate its use. Then we explain votes to exclude defectors, and finally how exclusion affects subsequent cooperation.

The sanctioning institution is sparsely used. In the low-S treatment, more than 50% of the participants vote to exclude no one in each of the repeated rounds. In the hi-S treatment, about 40% of the participants vote to exclude no one. Accordingly, the likelihood to be excluded from the bonus with a majority vote is rather low.[14] It is reasonable that punishment is rare in equilibrium (Gächter, 2012): after all, when the norm is followed, there is no need for exclusion.

Next, we focus on the votes to exclude defectors. We expect that defection is more frequently sanctioned in the hi-S treatment if the established norm of cooperation is higher than the norm in the low-S treatment. To this end, we estimate the panel

[13] While 55% and 56% of the participants' empirical expectations are accurate in round 1-3 and round 4-6 of the -S treatments, accuracy drops from 54% in round 1-3 to 44% in round 4-6 in the -noS treatments.

[14] The probability ranges from 10% to 25% for being excluded after defecting in a given round in both the hi-S treatment and low-S treatment and between 10% and 30% for being excluded after cooperating in a given round in the hi-S treatment and respectively between 3% and 25% in the low-S treatment.

Table 2.5: Non-linear estimation to explain the use and effect of the sanctioning institution

	Vote to Exclude Def.	Cooperation	Emp. Expectation
	(1)	(2)	(3)
low-S	-1.158***	-1.669***	-1.581***
	(0.358)	(0.372)	(0.321)
excluded (def in $t-1$)	0.602***	0.099	0.008
	(0.081)	(0.161)	(0.181)
excluded (def in $t-1$) x low-S	-0.212	-0.283	0.037
	(0.279)	(0.361)	(0.291)
excluded (coop in $t-1$)	0.353*	-0.315**	-0.240
	(0.207)	(0.127)	(0.263)
excluded (coop in $t-1$) x low-S	0.144	0.282	0.197
	(0.261)	(0.218)	(0.318)
coop in t	0.394***		
	(0.138)		
Round Trend	Yes	Yes	Yes
Session Fixed Effects	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes
N	1764	1764	1764

Notes: The table reports random effect estimators for an individual-level probit model. The baseline is set to the hi-S treatment. The model includes the one-shot game and all repeated rounds. Individual controls include age, age squared, an indicator variable for comprehension, a risk preference measure, the main type of gear, and crew size. Robust standard errors are clustered at the session level (in parentheses). ***, **, and * indicate significance at the 1,5, and 10% level.

data model shown in column (1) of Table 2.5. We find that there is less voting to exclude defectors in the low-S than hi-S treatment, despite the fact that there is more defection in the low-S treatment. This corroborates that indeed different social norms have formed in the two treatments.

The regression model also includes whether a participant was excluded for defection or cooperation in the previous round, a round trend, and whether a participant has cooperated in the current round. Being excluded from the bonus after defecting leads to a higher probability to vote for excluding defectors in the next round. We also see a weak effect for being excluded after cooperating. In both low-S and hi-S treatments there is no round trend in voting behavior. Finally, the positive coefficient for the variable indicating whether a participant has cooperated in the current round, shows an additional intuitive, but subtle function of the sanctioning institution: It is a self-signaling device, reinforcing that the action one has just chosen was indeed the right thing to do.

Further, we are interested in the influence of exclusion on cooperation and empirical expectation in the next round. Recall that participants cannot vote to exclude

specific group members, but they can give feedback on strategies. The estimates presented in column (2) and (3) of Table 2.5 show that being excluded for defection has no significant effect on cooperation or empirical expectation in the following round. Thus, exclusion itself does not drive conformity. In addition, the estimation shows a small negative effect on cooperation when participants are excluded for cooperation in the hi-S treatment.

Our findings on (a) the limited use of the sanctioning institution, (b) the votes to exclude defectors, and (c) the cooperation after being excluded indicate that it is the option to comment on which behavior should be excluded, rather than the exclusion itself that stabilizes cooperation.

2.6 "Norms in the Wild"

We find that providing social information only leads to stable cooperation rates when the message is combined with a sanctioning institution. The policy implication for using social norms as a tool to improve governance under weak institutions is clear: Local fora for feedback and discussion need to be established and supported to accompany norm-based interventions. The existing BMUs could be such a forum at Lake Victoria. Policy makers should strengthen these institutions when using a norm-based intervention to create "norms in the wild" (the book title of Bicchieri, 2017).

An additional policy-relevant question is whether the effect of a norm-based intervention can be amplified by the existing social structure in the targeted communities.[15] Norm-based theories make a clear prediction: the more relevant a fisherman perceives the reference group, the more weight will be attached to the social information message and the more forceful will be the threat of social sanctioning.

To study how a participant's social proximity to others affects norm conformity, we work with the natural heterogeneity of fishing communities at Lake Victoria and construct a social proximity measure that is based on two observable characteristics: main target species and region of origin. Put simply, we ask whether, *e.g.*, a dagaa fisherman from Ukerewe Island is more strongly affected by the social information message when he is in a session full of other dagaa fishermen from Ukerewe or when he is in a session full of Nile perch fishermen from Rorya. Figure 2.5 shows predictive

[15] For example, Loock et al. (2012) find that in their social comparison study on energy conservation, reference groups with close social proximity are more effective in inducing behavioral change. Similarly, Costa and Kahn (2013) test whether political orientation can predict the effectiveness of their intervention. Close to the current application, Barnes et al. (2016) highlight the importance of ethnic networks among fishermen in sharing information on shark bycatch.

margins for the interaction between our proximity measure and the treatment variation on individual cooperation (left panel) and empirical expectations (right panel) in the repeated game. The full regression results are shown in Appendix Table A-7.

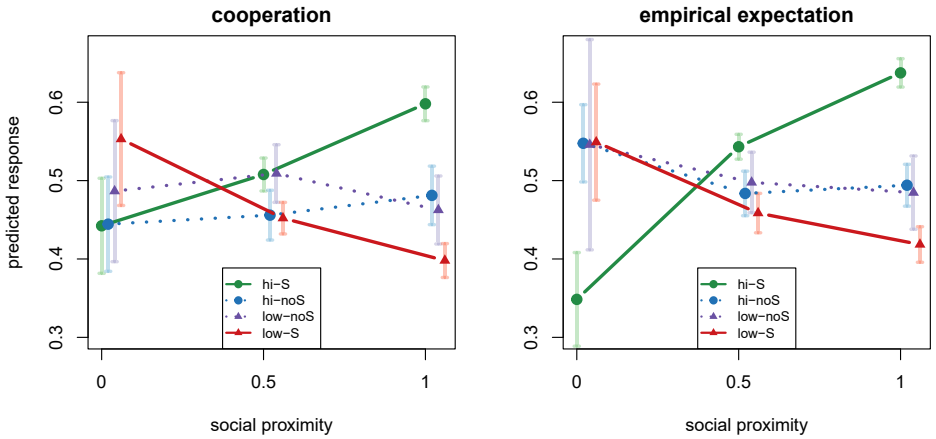


Fig. 2.5: Predictive margins for the effect of social proximity on cooperation (left panel) and empirical expectation (right panel) for the four different treatments. Whiskers indicate ±1 SE. The green solid line with round markers shows that participants in the hi-S treatment are less likely to cooperate and have low empirical expectations when they different from the session norm with respect to target fishery and region of origin (sp_i=0) and more likely to cooperate and have high empirical expectations when they have high social proximity (sp_i=1). Conversely, the red solid line with triangle markers shows that participants in the low-S treatment are more likely to cooperate and have high empirical expectations when they have low social proximity and less likely to cooperate and have low empirical expectations when they have high social proximity.

We find that, on average, an increase in the proximity of a fisherman to his peers in the session leads to *more* cooperation in the hi-S treatment, but *less* cooperation in the low-S treatment. In other words, fishermen with a strong reference network in the session conform to the behavior that is emphasized by the social information message. Fishermen with weak social ties to the others in their session, however, appear unaffected by the social information message and do not conform to it. In contrast, social proximity has no effect on cooperation in the -noS treatments. This is not surprising since without a sanctioning institution, different social norms of cooperation are not created.

For empirical expectations, we document even stronger effects: Fishermen with close social proximity to other participants expect their peers to cooperate more in the hi-S treatment and less in the low-S treatment. We thus confirm Hypothesis 4. In

contrast, fishermen that differ with respect to both dimensions of target species and ethnicity expect the others in the session to defect when the social information message emphasizes cooperation. As expected, social proximity has no effect on empirical expectations when no social norm is created in the two -noS treatments.

2.7 Discussion

Reducing fishing pressure at Lake Victoria is a prototypical situation where formal institutions are weak and norm-based interventions may be a promising tool to enhance cooperation and support communal self-management. In this setting, we study the moderating role of a social sanctioning institution when initial information on the behavior of others aims to create different social norms of cooperation. The design of our experiment includes repeated observation of both actions and empirical expectations, and is thereby able to detect or reject the pattern of behavior and beliefs that characterizes a social norm.

We show that the provision of social information can lead to the creation of different cooperation norms conditional on the presence of a social sanctioning institution. Without sanctioning, cooperation rates decline. With sanctioning, cooperation rates start high and stay high when the initial social information emphasizes cooperation. In contrast, cooperation rates start low and stay low when the initial social information emphasizes defection. The dynamic pattern of behavior is mirrored by a parallel pattern of empirical expectations. Our results are well explained by norm-based theories where individuals prefer to conform with what they expect others in their reference network to do.

One could object to a norm-based explanation by arguing that the social sanctioning institution has changed the game from a social dilemma game to a coordination game where the initial information provides the focal point. However, our weak sanctioning institution is deliberately designed in a way such that the Nash equilibrium of the material game is not changed.

Nevertheless, the game may have changed in participants' utility space (Ostrom, 1998; Cárdenas and Ostrom, 2004). Without knowing the non-monetary component of participants' utility function, the argument that social sanctioning changes the game form in utility space is impossible to refute. What we can do is to inspect participants' stated personal normative beliefs and normative expectations. Here, we see that about a third of the participants hold the normative belief that one ought to "do

what others do" (Figure 2.3). Associating this normative belief with a preference for conformity or conditional cooperation, we can study the behavior of these conditional cooperators. If the observed pattern is explained by the coordination of conditional cooperators on different focal points, we would need to observe that conditionally cooperative participants are more likely to cooperate when the initial information emphasizes cooperation and less likely to cooperate when the initial information emphasizes defection. However, we find that conditional cooperators are more likely to cooperate than not in all treatments (see Appendix A-2). In fact, the predicted probability that conditional cooperators contribute to the group account is indistinguishable from the probability of unconditional cooperators to do so. In contrast, participants that think that one ought to defect are less likely to contribute to the group account across all treatments. Hence, while we cannot rule out that the treatments induce different game forms in participants' utility space, we find no evidence for such an argument, either. That said, a view where social sanctioning creates a coordination game out of a social dilemma and initial information provides a focal point exactly accords to what some, for example Binmore (2010), call a social norm.

Fehr and Schurtenberger (2018a) present the only other experiment that interacts social priming with sanctioning that we are aware of. Their lab experiment also documents stable cooperation rates when a norm-nudge is combined with a sanctioning institution, but declining cooperation rates without a sanctioning institution. The design of Fehr and Schurtenberger (2018a) is sufficiently similar to ours so that the two studies can be seen as conceptual replications of each other.[16]

We believe that three lessons can be learned from the comparison of the two studies. First, while our sanctioning institution consists of three components: (i) giving/receiving feedback on others'/own behavior, (ii) a financial loss from being sanctioned, and (iii) a voting stage that could induce or reinforce consistency of (expressed) opinions, the sanctioning institution of Fehr and Schurtenberger (2018a) contains components (i) and (ii). Thus, the two studies suggest that cooperation

[16] Despite their similarity, there are several differences in the design of the two studies, most notably: (i) Fehr and Schurtenberger (2018a) attempt to affect participant's *normative* expectations while we attempt to affect participant's *empirical* expectations. (ii) Different participant pools: students from Nottingham vs. fishers from Lake Victoria. (iii) Both sanctioning institutions are 'weak' but in different manners: Fehr and Schurtenberger (2018a) weaken their peer punishment institution (Fehr and Gächter, 2000, 2002) by allowing for counter-punishment (Nikiforakis, 2008), while we chose a punishment that is weak in that it does not change the Nash equilibrium. (iv) The punishment in Fehr and Schurtenberger (2018a) is decentralized and is directed towards participants, while our sanctioning is directed towards a strategy and only inflicted when coordinated through a majority rule. (v) Fehr and Schurtenberger (2018a) use partner-matching for fifteen rounds, while we use perfect-stranger matching for seven rounds.

is affected by an attempt to activate social norms as long as there is a social enforcement institution that contains elements of feedback and sanctioning. Second, Fehr and Schurtenberger (2018a) observe stable cooperation rates with sanctioning and deteriorating rates without sanctioning over 15 rounds while we observe this pattern over seven rounds, indicating that our results may hold in a considerably longer experiment. Third, Fehr and Schurtenberger (2018a) attempt to activate the moral component of social norms, while our intervention is based on a descriptive message. The results suggest that both approaches work in a similar way.[17]

Our study opens a number of important avenues for future research: Theoretical and experimental work is needed to understand the incentives that norm-based interventions create. For example, how do preferences for conformity with social norms differ from preferences for fairness, reciprocity, or guilt aversion? How does norm-activation and norm-compliance respond to changes in the structure of the underlying game, or the sanctioning institution?

We choose a repeated prisoner's dilemma because of its simplicity and because it captures the essential cooperation problem for fishermen at Lake Victoria: compliance with official regulations under weak formal institutions. Based on our findings, future work should study the creation of social norms in a more complicated extraction game with dynamic spillovers. Furthermore, it is important to study the implementation of norm-based interventions on actual conservation efforts. Understanding the effectiveness of "social norms as solutions" is imperative, particularly in settings where formal regulation of natural resource use is challenging (Nyborg et al., 2016).

In sum, our study provides strong evidence that norm-based interventions have the potential to enhance cooperation in social dilemma situations, but they require social enforcement institutions. We take a first step towards bridging the lab and the field and provide policy recommendations from the experimental test-bed. To improve governance through norm-based interventions, they ought to be accompanied by the opportunity for social enforcement via supporting fora for feedback and discussion within relevant reference networks. In the context of Lake Victoria, existing co-management structures such as the BMUs may facilitate social enforcement by holding regular community meetings that focus on issues of non-compliance, or by

[17] A reason could be that a normative message contains information about the behavior that can be expected from others if these others often act in accordance with their normative beliefs. For the same reason, a descriptive message may also contain normative content (Bicchieri and Xiao, 2009). However, the empirical equivalence of a descriptive and normative message should be further investigated in the social dilemma and other settings. Bicchieri et al. (2020), for example, do not find such equivalence in a one-shot trust game.

adopting bylaws that mandate the exclusion of members that violate existing regulations. Moreover, our finding that social information has a stronger effect for those participants that are closer to the others in their session, suggests that targeting representative members as social multipliers and building community cohesion could be important auxiliary measures to improve governance under weak institutions.

Bibliography

Ai, C. and Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1):123–129.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.

Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10):3003–37.

Alpizar, F., Carlsson, F., and Johansson-Stenman, O. (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in costa rica. *Journal of Public Economics*, 92(5-6):1047–1060.

Baland, J.-M. and Platteau, J.-P. (1996). *Halting degradation of natural resources: is there a role for rural communities?* Food & Agriculture Org.

Barnes, M. L., Lynham, J., Kalberg, K., and Leung, P. (2016). Social networks and environmental outcomes. *Proceedings of the National Academy of Sciences*, 113(23):6466–6471.

Barrera, D. and Simpson, B. (2012). Much ado about deception: Consequences of deceiving research participants in the social sciences. *Sociological Methods & Research*, 41(3):383–413.

Barrett, S. (2018). Choices in the climate commons. *Science*, 362(6420):1217–1217.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Beuving, J. J. (2010). Playing pool along the shores of lake victoria: Fishermen, careers and capital accumulation in the ugandan nile perch business. *Africa*, 80(2):224?248.

Bicchieri, C. (2017). *Norms in the Wild – How to Diagnose, Measure, and Change Social Norms.* Oxford University Press.

Bicchieri, C., Dimant, E., Gächter, S., and Nosenzo, D. (2019). Observability, social proximity, and the erosion of norm compliance.

Bicchieri, C., Dimant, E., and Xiao, E. (2020). Deviant or Wrong? The Effects of Norm Information on the Efficacy of Punishment. ECONtribute Discussion Papers Series 023, University of Bonn and University of Cologne, Germany.

Bicchieri, C., Muldoon, R., and Sontuoso, A. (2018). Social norms. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition.

Bicchieri, C. and Xiao, E. (2009). Do the Right Thing: But Only if Others Do So. *Journal of Behavioral Decision Making*, 22(2):191–208.

Binmore, K. (2010). Social norms or social preferences? *Mind & Society*, 9(2):139–157.

Bohnet, I. and Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, 55(4):467 – 484. Trust and Trustworthiness.

Cardenas, J. C. (2011). Social norms and behavior in the local commons as seen through the lens of field experiments. *Environmental and Resource Economics*, 48(3):451–485.

Cárdenas, J.-C. and Ostrom, E. (2004). What do people bring into the game? experiments in the field about cooperation in the commons. *Agricultural systems*, 82(3):307–326.

Carpenter, J. and Seki, E. (2011). Do social preferences increase productivity? field experimental evidence from fishermen in toyama bay. *Economic Inquiry*, 49(2):612–630.

Carpenter, J. P., Daniere, A. G., and Takahashi, L. M. (2004). Cooperation, trust, and social capital in southeast asian urban slums. *Journal of Economic Behavior & Organization*, 55(4):533–551.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015–1026.

Costa, D. L. and Kahn, M. E. (2013). Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3):680–702.

Crona, B. and Bodin, Ö. (2006). What you know is who you know? communication patterns among resource users as a prerequisite for co-management. *Ecology and Society*, 11(2).

Dimant, E. (2019). Contagion of pro-and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, 73:66–88.

Duflo, E. and Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly Journal of Economics*, 118(3):815–842.

Eggert, H. and Lokina, R. B. (2010). Regulatory compliance in lake victoria fisheries. *Environment and Development Economics*, 15(2):197–217.

Elster, J. (1989). Social norms and economic theory. *Journal of economic perspectives*, 3(4):99–117.

Elster, J. (2007). *Social Norms*, pages 353–372. Cambridge University Press.

Eymess, T. (2021). Dissecting social norms of cooperation: A conditional process analysis. Unpublished manuscript, Heidelberg University.

Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868):137.

Fehr, E. and Leibbrandt, A. (2011). A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of Public Economics*, 95(9-10):1144–1155.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868.

Fehr, E. and Schurtenberger, I. (2018a). The dynamics of norm formation and norm decay. *Working paper series/Department of Economics*, (278).

Fehr, E. and Schurtenberger, I. (2018b). Normative foundations of human cooperation. *Nature Human Behaviour*, 2(7):458–468.

Ferraro, P. J. and Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1):64–73.

Frey, B. S. and Meier, S. (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *The American Economic Review*, 94(5):1717–1722.

Gächter, S. (2012). In the lab and the field: Punishment is rare in equilibrium. *Behavioral and Brain Sciences*, 35(1):26?28.

Gerber, A. S. and Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, 71(1):178–191.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645.

Goeschl, T., Kettner, S. E., Lohse, J., and Schwieren, C. (2018). From social information to social norms: Evidence from two experiments on donation behaviour. *Games*, 9(4).

Goette, L., Huffman, D., and Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review*, 96(2):212–216.

Goldstein, N., Cialdini, R., and Griskevicius, V. (2008). A room with a viewpoint: Using norm-based appeals to motivate conservation behaviors in a hotel setting. *Journal of Consumer Research*, 35(3):472–482.

Greene, W. and Zhang, Q. (2019). Nonlinear and related panel data models. In *Panel Data Econometrics*, pages 45–96. Elsevier.

Guala, F. (2012). Reciprocity: Weak or strong? what punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, 35(1):1–15.

Hallsworth, M., List, J. A., Metcalfe, R. D., and Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148:14 – 31.

Hayo, B. and Vollan, B. (2012). Group interaction, heterogeneity, rules, and co-operative behaviour: Evidence from a common-pool resource experiment in south africa and namibia. *Journal of Economic Behavior & Organization*, 81(1):9–28.

Jang, C. and Lynham, J. (2015). Where do social preferences come from? *Economics Letters*, 137:25 – 28.

Janssen, M. A., Holahan, R., Lee, A., and Ostrom, E. (2010). Lab experiments for the study of social-ecological systems. *Science*, 328(5978):613–617.

Karaca-Mandic, P., Norton, E. C., and Dowd, B. (2012). Interaction terms in nonlinear models. *Health Services Research*, 47(1pt1):255–274.

Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.

Kosfeld, M. and Rustagi, D. (2015). Leader punishment and cooperation in groups: Experimental field evidence from commons management in ethiopia. *American Economic Review*, 105(2):747–83.

Loock, C.-M., Landwehr, J. R., Staake, T., Fleisch, E., and Pentland, A. S. (2012). The influence of reference frame and population density on the effectiveness of social normative feedback on electricity consumption.

Lopez, M. C., Murphy, J., M., S. J., and Stranlund, J. K. (2012). Comparinng the effectiveness of regulation and pro-social emotions to enhance cooperation: Experimental evidence from fishing communities in colombia. *Economic Inquiry*, 50(1):131–142.

Luomba, J. (2013). Role of beach management units in implementing fisheries policy: A case study of two bmus in lake victoria, tanzania. Technical report, UNU-Fisheries Training Programme.

LVFO (2017). Regional Status Report on Lake Victoria Bi-ennial Frame Surveys between 2000 and 2016. Technical Report April 2017.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93(1):366–380.

Mason, C. M. (2006). Exploring the processes underlying within-group homogeneity. *Small Group Research*, 37(3):233–270.

Meier, S., Pierce, L., Vaccaro, A., and La Cara, B. (2016). Trust and in-group favoritism in a culture of crime. *Journal of Economic Behavior & Organization*, 132:78–92.

Michaeli, M. and Spiro, D. (2017). From peer pressure to biased norms. *American Economic Journal: Microeconomics*, 9(1):152–216.

Mkumbo, O. C. and Marshall, B. E. (2015). The Nile perch fishery of Lake Victoria: Current status and management challenges. *Fisheries Management and Ecology*, 22(1):56–63.

Mosimane, A. W., Breen, C., and Nkhata, B. A. (2012). Collective identity and resilience in the management of common pool resources. *International Journal of the Commons*, 6(2):344–362.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92(1-2):91–112.

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and Social Psychology Bulletin*, 34(7):913–923.

Nunan, F., Cepić, D., Mbilingi, B., Odongkara, K., Yongo, E., Owili, M., Salehe, M., Mlahagwa, E., and Onyango, P. (2018). Community cohesion: Social and economic ties in the personal networks of fisherfolk. *Society & Natural Resources*, 31(3):306–319.

Nunan, F. and Onyango, P. (2017). Inter-sectoral governance in inland fisheries: Lake victoria. pages 48–57. TBTI publication series, St John's, NL, Canada.

Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger, W. N., Arrow, K. J., Barrett, S., Carpenter, S., Chapin, F. S., Crépin, A.-S., Daily, G., Ehrlich, P., Folke, C., Jager, W., Kautsky, N., Levin, S. A., Madsen, O. J., Polasky, S., Scheffer, M., Walker, B., Weber, E. U., Wilen, J., Xepapadeas, A., and de Zeeuw, A. (2016). Social norms as solutions. *Science*, 354(6308):42–43.

Ostrom, E. (1990). *Governing the commons: The Evolution of Institutions for Collective Action.* Cambridge university press, Cambridge, MA.

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

Ostrom, E. (2000). Collective Action and the Evolution of Social Norms. *Journal of Economic Perspectives*, 14(3):137–158.

Ostrom, E. (2008). The challenge of common-pool resources. *Environment: Science and Policy for Sustainable Development*, 50(4):8–21.

Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86(2):404–417.

Platteau, J.-P. (2006). Solidarity norms and institutions in village societies: Static and dynamic considerations. *Handbook of the Economics of Giving, Altruism and Reciprocity*, 1:819–886.

Rousu, M. C., Colson, G., Corrigan, J. R., Grebitus, C., and Loureiro, M. L. (2015). Deception in experiments: towards guidelines on use in applied economics research. *Applied Economic Perspectives and Policy*, 37(3):524–536.

Rustagi, D., Engel, S., and Kosfeld, M. (2010). Conditional cooperation and costly monitoring explain success in forest commons management. *Science*, 330(6006):961–965.

Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural paraguay. *Journal of Economic Behavior & Organization*, 62(2):272 – 292.

Schram, A. and Charness, G. (2015). Inducing social norms in laboratory allocation choices. *Management Science*, 61(7):1531–1546.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429–434.

Sethi, R. and Somanathan, E. (1996). The Evolution of Social Norms in Common Property Resource Use. *American Economic Review*, 86(4):766–788.

Shang, J. and Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal*, 119(540):1422–1439.

Tajfel, H. and Turner, J. C. (1979). An integrative theory of intergroup conflict. *Organizational Identity: A Reader*, pages 56–65.

Tyran, J.-R. and Feld, L. P. (2006). Achieving compliance when legal sanctions are non-deterrent. *Scandinavian Journal of Economics*, 108(1):135–156.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Young, H. P. (2015). The evolution of social norms. *Annual Review of Economics*, 7(1):359–387.

Appendix

A-1 Estimation Strategy

Throughout the experiment's analysis, cooperation in the social dilemma is proxied with a binary contribution decision in a three-player prisoner's dilemma. Participants have the option to allocate their full endowment to either their private account or to a group account. Partial contributions to either account are not possible. To match the choice set in the contribution stage, empirical expectations are elicited as binary beliefs. Cooperation in the social dilemma and the antecedent expectation about others' play are the dependent variables in the analysis. They are each elicited a total of seven times over the course of the experiment.

We account for the binary nature of both outcome variables by employing nonlinear panel data models. Individual observations are correlated over time. That is, cooperation $y_{i,t}$ of individual i in round t is assumed to be correlated with the cooperation decision of the same individual in the preceding rounds.[18] Standard errors are clustered on the session level to account for idiosyncratic conditions in the experimental setup at each landing site. We use random effects probit models for all regressions on cooperation and empirical expectation to account for the unobserved heterogeneity in normative characteristics. That is, the emergence of normative behavior is assumed to vary across our sample. Participants differ in their preference for conformity and are not assumed to be comparable with respect to their sensitivity to social enforcement or the social information message as such.

Treatment interaction terms are included in all specifications. Let $y_{i,t}$ be the response of individual i at time t, $x_{i,t}$ be a set of time-variant and time-invariant covariates, and z_i a variable indicating the different treatments.[19] As treatments are not varied within sessions, z_i is time-invariant. With c_i as unobserved time-invariant heterogeneity and $\epsilon_{i,t}$ as idiosyncratic disturbance in the model, the general regression equation is specified as follows,

$$y_{i,t} = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_{12} x_{i,t} z_i + c_i + \epsilon_{i,t}. \tag{2.2}$$

The interpretation of interaction terms in non-linear panel data models is problematic as the interaction effect can be non-zero although β_{12} equals zero (Ai and Norton, 2003; Karaca-Mandic et al., 2012; Greene and Zhang, 2019). We therefore desist from quantifying the average treatment effect in Table 2.4 and Table A-7. To provide intuition about the relative size of the treatment effect with respect to the social proximity measure (Figure 2.5) and normative beliefs (Figure A-1), we provide treatment specific plots for marginal effects. Readers who are generally interested in the effect size of our norm-based intervention are referred to the descriptive results in Section 2.5. Also, Wooldridge (2010) describes linear probability models as a convenient approximation to the underlying outcome probability. For common values of the interacted covariates, linear probability models tend to give good estimates of partial effects. To this end, we present all relevant specifications not only as non-linear but also

[18] Due to the perfect-stranger matching protocol, we do not expect interdependencies across individuals.

[19] In the specifications of Table 2.4 and the related Table A-3 in Appendix A-2, $x_{i,t}$ includes a round trend variable. In the specifications of Table 2.5, $x_{i,t}$ includes an indicator variable of being excluded from the bonus point for defecting in the previous round, an indicator variable for being excluded from the bonus point for cooperating in the previous round, and a round trend variable. In the specification of Table A-7 and the related Table A-8 in Appendix A-2, $x_{i,t}$ includes a measure for social proximity to other fishermen in the session and a round trend variable. Note that the round trend is the only covariate that varies with time.

linear models, see Tables A-4, A-6, A-8, A-9, and A-10. The robustness of our results to the specification as a linear model alleviates concerns about the misspecification of interaction terms in the non-linear case (Wooldridge, 2010).

A-2 Additional Analysis and Robustness Tests

Participant Characteristics

The sample consists of $N = 588$ fishermen from landing sites in the Lake Victoria region of Tanzania. We present sample characteristics split by the different treatments in Table A-1. With respect to age, treatment averages are between 35 and 40 years of age. The sample is predominantly male with only 0% to 4% of female participants in the different treatments. Our sample is rather poor. The share of participants earning less than 5,000 TZS per day (ca. USD 2.20), indicated by *earndaily* is just under 50% in the varying treatments. In terms of the fishermen's reliance on fishing as an income source indicated by the variable *hhfracfish* ($1 = 0\% - 25\%, 2 = 25\% - 50\%, 3 = 50\% - 75\%, 4 = 75\% - 100\%$), participants answer that about 2/3 of their income comes from fishing. To the question on how often participants have moved in the past, about 60% of respond that they have never moved (see *movefreq*). Moreover, fishermen usually fish in crews of 3 to 4 (*crewsize*) while about 40% use the dagaa net as their main gear type (*maingear*). Lastly, on a seven-point Likert scale (0-6), participants invest about 3 points into a risky instead of a safe option (*risk_pref*).

Table A-1: Participants' characteristics: *min adj. p-val* shows the lowest p-value in all six pairwise comparisons

treatment	hi-S	hi-noS	low-noS	low-S	min adj. p-val
age	37.82	40.01	38.49	35.48	0.01
gender	0.02	0	0.04	0.02	0.24
earndaily	0.43	0.43	0.48	0.46	0.99
hhfracfish	2.53	2.68	2.75	2.76	0.52
movefreq	0.6	0.67	0.54	0.61	0.34
crewsize	3.85	3.88	3.63	3.8	0.08
maingear	0.4	0.3	0.49	0.38	0.01
risk_pref	2.9	2.97	3.01	2.71	0.87

To argue for a balanced sample with respect to the treatment variation, the rightmost column in Table A-1 displays the lowest p-value from a series of pairwise mean comparison tests between treatment averages per characteristic (we account for multiple testing hypotheses using the Benjamini-Hochberg procedure). We find that only age (highest mean in hi-noS = 40.01, lowest mean in low-S = 35.48) and an indicator variable for main gear type being the dagaa net indicated by *maingear* (highest mean in low-noS = 0.49, lowest mean in hi-noS = 0.3) display statistically significant differences at the 5% level. We control for these factors in all regressions.

Cooperation Rates Over Time

To argue for the significance of our main result, *i.e.*, the significantly different cooperation rates in the information treatments with a sanctioning institution presented in Section 2.5.1, we conduct Wilcoxon Mann-Whitney tests. The pairwise comparison between treatments are

shown in Table A-2. Column (1) displays differences in average cooperation between the high information treatment with sanctions (hi-S) and the low information treatment with sanctions (low-S). The difference is statistically significant in every iteration of the repeated game except for the last round. Column (2) shows the difference in average cooperation between the hi-S treatment and the corresponding high information treatment without a sanctioning institution (hi-noS). The cooperation rate is significantly lower in the hi-noS treatment in the last three rounds of the experiment. Column (3) indicates that in early rounds, average cooperation in the low-noS treatment is marginally higher than in the corresponding treatment with sanctions (low-S). Finally, column (4) indicates no statistical differences between hi-noS and low-noS treatments.

Note that we can also conduct the test on the session level, leaving only seven independent observation per treatment. We still find a marginally significant difference between the hi-S and low-S treatments ($p = .096$).

Table A-2: Treatment differences in average cooperation

Round	hi-S - low-S (1)	hi-S - hi-noS (2)	low-S - low-noS (3)	hi-noS - low-noS (4)
OS	+0.07	-0.03	-0.12*	-0.02
1	+0.14**	+0.05	-0.12*	-0.03
2	+0.12**	+0.07	-0.05	-0.01
3	+0.12*	+0.03	-0.07	+0.01
4	+0.16***	+0.15***	-0.03	-0.03
5	+0.13**	+0.18***	+0.01	-0.04
6	+0.09	+0.16***	+0.04	-0.03

Notes: Round-by-round Wilcoxon Mann-Whitney-test results. OS denotes the one-shot game, the numbers correspond to the rounds of the repeated part of the experiment. ***, **, and * indicate significant differences at the 1, 5, and 10% level.

Regression Results

We show robustness of our results with respect to the estimation strategy discussed in Appendix A-1. In Table A-3 we report robustness tests for Table 2.4 in Section 2.5.1. Depicted are probit models for the regression of treatment effect, round effect, and their interaction on cooperation and empirical expectations. Results are robust to the exclusion of the one-shot game, see columns (1) and (2), the exclusion of the last round, see columns (3) and (4), and the exclusion of all participants that failed the comprehension test, see columns (5) and (6). Prior participation in economic experiments is added as a co-variate in all models. In all specifications, we find that participants, on average, contribute significantly less in the low-S than in the hi-S treatment while treatments without a sanctioning institution are characterized by the significant breakdown of cooperation over the repeated game. The treatment difference between hi-S and low-S treatment is especially pronounced when excluding the one-shot game from the sample, see columns (1) and (2). Also, we report coefficients for all included control variables. For main choice of gear, small seine net users are the baseline. We observe that risk preferences are negatively correlated with cooperation. Also, participants that work on boats with large crews (six members or more) cooperate significantly less.

Furthermore, we repeat the robustness analysis for the choice of a linear probability model, see Table A-4. Here, results convey the same implication as for the non-linear model. Participants, on average, contribute significantly less in the low-S than in the hi-S treatment. Also, the cooperation rates in both treatments without sanctions, on average, decrease per repetition round. Likewise, the results in linear probability models are robust to the exclusion of the one-shot game, see columns (1) and (2), the exclusion of the last round, see columns (3) and (4), and the exclusion of all participants that failed the comprehensions test, see columns (5) and (6). Again, excluding the one-shot game from the analysis yields the strongest results.

Table A-3: Non-linear model explaining behavior and beliefs for subsets of the experiment

	Without OS		*Without Last Round*		*Without Test Q*	
	Coop	EE	Coop	EE	Coop	EE
	(1)	(2)	(3)	(4)	(5)	(6)
low-noS	-0.162	-1.005**	-0.183	-0.629**	0.074	-0.450
	(0.367)	(0.419)	(0.301)	(0.294)	(0.339)	(0.302)
low-S	-1.285***	-0.853***	-0.900***	-0.703***	-0.831***	-0.551**
	(0.356)	(0.307)	(0.246)	(0.264)	(0.257)	(0.260)
hi-noS	-0.338	-0.520*	-0.175	-0.060	-0.233	-0.194
	(0.325)	(0.290)	(0.300)	(0.259)	(0.300)	(0.261)
Round	-0.051	-0.076**	0.001	-0.029	-0.023	-0.029
	(0.047)	(0.036)	(0.049)	(0.040)	(0.037)	(0.035)
low-noS × Round	-0.139**	0.036	-0.185***	-0.041	-0.177***	-0.025
	(0.065)	(0.060)	(0.055)	(0.057)	(0.058)	(0.050)
low-S × Round	0.028	0.064	-0.063	0.056	-0.016	0.013
	(0.068)	(0.048)	(0.054)	(0.049)	(0.045)	(0.038)
hi-noS × Round	-0.171***	-0.019	-0.202***	-0.091*	-0.173***	-0.067*
	(0.055)	(0.042)	(0.068)	(0.051)	(0.050)	(0.039)
Age	0.054	0.018	0.056	0.035	0.050	0.015
	(0.048)	(0.042)	(0.041)	(0.031)	(0.060)	(0.047)
Age^2	-0.000	0.000	-0.000	-0.000	-0.000	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Main gear = hook	-0.396	-0.230	-0.349	-0.232	-0.227	0.004
	(0.293)	(0.301)	(0.269)	(0.280)	(0.291)	(0.304)
Main gear = dagaa net	-0.284	-0.048	-0.203	-0.078	-0.307	-0.049
	(0.255)	(0.276)	(0.257)	(0.260)	(0.303)	(0.267)
Main gear = other	-0.467	0.200	-0.376	-0.093	-0.215	0.170
	(0.411)	(0.451)	(0.389)	(0.378)	(0.378)	(0.360)
Crew size = 1	-0.738	-0.343	-0.529	-0.200	-1.332	-0.840
	(0.901)	(0.960)	(0.739)	(0.781)	(0.826)	(0.792)
Crew size = 2	-0.275	0.338	0.006	0.413	-0.581	0.141
	(0.746)	(0.880)	(0.736)	(0.833)	(0.826)	(0.852)
Crew size = 3	-0.817	-0.262	-0.654	-0.179	-1.140**	-0.445
	(0.499)	(0.574)	(0.450)	(0.482)	(0.556)	(0.521)
Crew size = 4	-0.830	-0.240	-0.701	-0.247	-0.871	-0.141
	(0.596)	(0.603)	(0.562)	(0.531)	(0.660)	(0.584)
Crew size = 5	-0.974*	-0.319	-0.921*	-0.315	-1.140*	-0.199
	(0.534)	(0.634)	(0.519)	(0.559)	(0.600)	(0.578)
Crew size = 6+	-2.308***	-1.729**	-2.087***	-1.567***	-2.361***	-1.598**
	(0.733)	(0.717)	(0.653)	(0.604)	(0.746)	(0.659)
Risk preference	0.211***	0.186***	0.206***	0.166***	0.218***	0.181***
	(0.035)	(0.039)	(0.037)	(0.033)	(0.036)	(0.034)
Comprehension	-0.324	-0.529**	-0.302	-0.399*		
	(0.276)	(0.256)	(0.270)	(0.228)		
Prior participation	-0.011	-0.035	-0.018	-0.045	0.014	-0.039
	(0.230)	(0.240)	(0.220)	(0.221)	(0.232)	(0.222)
Session Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	3456	3456	3456	3456	3535	3535

Notes: The table reports random effect estimators for an individual-level probit model. The baseline is set to the hi-S treatment. Robust standard errors are clustered at the session level (in parentheses). ***, **, and * indicate significance at the 1,5, and 10% level.

Table A-4: Linear model explaining behavior and beliefs for subsets of the experiment

	Without OS		*Without Last Round*		*Without Test Q*	
	Coop	EE	Coop	EE	Coop	EE
	(1)	(2)	(3)	(4)	(5)	(6)
low-noS	-0.023	-0.178**	-0.044	-0.138**	-0.001	-0.095
	(0.061)	(0.073)	(0.055)	(0.057)	(0.060)	(0.060)
low-S	-0.232***	-0.137**	-0.191***	-0.127**	-0.176***	-0.093*
	(0.061)	(0.053)	(0.049)	(0.052)	(0.048)	(0.049)
hi-noS	-0.068	-0.076	-0.037	-0.004	-0.070	-0.037
	(0.053)	(0.049)	(0.058)	(0.049)	(0.055)	(0.049)
Round	-0.009	-0.012**	-0.001	-0.006	-0.005	-0.006
	(0.007)	(0.006)	(0.009)	(0.007)	(0.007)	(0.006)
low-noS	-0.023**	0.005	-0.032***	-0.008	-0.032***	-0.005
	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)
low-S × Round	0.005	0.011	-0.012	0.011	-0.003	0.003
	(0.012)	(0.009)	(0.010)	(0.009)	(0.008)	(0.007)
hi-noS × Round	-0.029***	-0.005	-0.037***	-0.018*	-0.031***	-0.013*
	(0.009)	(0.007)	(0.012)	(0.009)	(0.008)	(0.007)
Age	0.008	0.007	0.009	0.009	0.007	0.006
	(0.008)	(0.007)	(0.008)	(0.007)	(0.011)	(0.009)
Age^2	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Main gear = hook	-0.053	-0.021	-0.052	-0.022	-0.018	0.025
	(0.054)	(0.053)	(0.052)	(0.053)	(0.055)	(0.056)
Main gear = dagaa net	-0.024	0.015	-0.025	0.012	-0.034	0.022
	(0.051)	(0.052)	(0.053)	(0.050)	(0.060)	(0.050)
Main gear = other	-0.049	0.049	-0.033	0.022	-0.003	0.066
	(0.079)	(0.083)	(0.075)	(0.079)	(0.070)	(0.075)
Crew size = 1	-0.176	-0.107	-0.128	-0.073	-0.277*	-0.223
	(0.164)	(0.185)	(0.151)	(0.170)	(0.165)	(0.164)
Crew size = 2	-0.038	0.039	0.012	0.060	-0.101	-0.012
	(0.131)	(0.165)	(0.134)	(0.160)	(0.142)	(0.158)
Crew size = 3	-0.127	-0.061	-0.109	-0.046	-0.191**	-0.109
	(0.085)	(0.107)	(0.081)	(0.094)	(0.092)	(0.093)
Crew size = 4	-0.127	-0.053	-0.118	-0.062	-0.148	-0.050
	(0.101)	(0.111)	(0.100)	(0.101)	(0.110)	(0.104)
Crew size = 5	-0.148	-0.067	-0.154	-0.073	-0.203**	-0.061
	(0.094)	(0.116)	(0.095)	(0.109)	(0.099)	(0.105)
Crew size = 6+	-0.340***	-0.295**	-0.351***	-0.313***	-0.389***	-0.312***
	(0.111)	(0.124)	(0.106)	(0.115)	(0.112)	(0.112)
Risk preference	0.037***	0.034***	0.038***	0.033***	0.039***	0.036***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Comprehension	-0.074	-0.110**	-0.070	-0.096**		
	(0.049)	(0.046)	(0.049)	(0.045)		
Prior participation	-0.006	-0.013	-0.010	-0.016	-0.000	-0.020
	(0.040)	(0.043)	(0.040)	(0.045)	(0.041)	(0.043)
Session Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.119	0.094	0.122	0.089	0.133	0.101
N	3456	3456	3456	3456	3535	3535

Notes: The table reports random effect estimators for an individual-level linear probability model. The baseline is set to the hi-S treatment. Robust standard errors are clustered at the session level (in parentheses). ***, **, and * indicate significance at the 1,5, and 10% level.

Next, we turn to a regression analysis of cooperation behavior on normative beliefs across the four different treatments. In Figure A-1, we present the predictive margins of a random effects probit model of cooperation behavior on both personal normative beliefs (left panel) and normative expectations (right panel) that were elicited during the belief elicitation stage in round one, see Section 2.3. The underlying regression results are presented in column (1) of Table A-5.

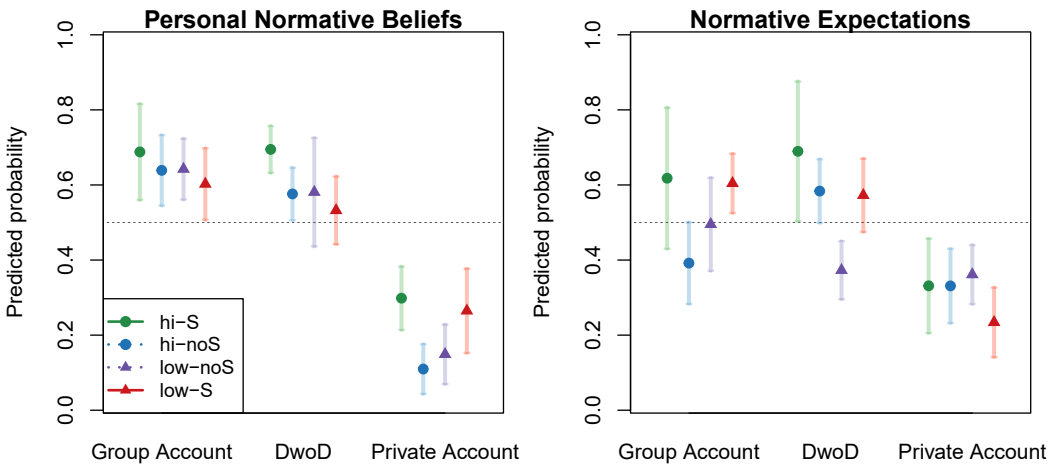


Fig. A-1: Predictive margins of cooperation for normative beliefs across treatments (based on probit model). DwoD indicates the response "Do what others do".

We expect that participants with preferences for conformity (thinking that the right thing to do is to "do what others do") are more likely to cooperate in the high treatments, where we emphasize that previous participants have cooperated. Conversely, we expect that participants with preferences for conformity are less likely to cooperate in the low treatments.

Looking at the effect of personal normative beliefs in the left panel of Figure A-1, we see that participants that think the right thing to do is to "do what others do" are indeed more likely to cooperate in the hi-S treatment than in the low-S treatment (there is no difference between the hi-noS and the low-noS treatment). However, the difference between the hi-S and the low-S treatment is not particularly large. Most importantly, for all participants with strong preferences for conformity the predicted probability is above 0.5. Thus, in all treatments, those participants cooperate more likely than not. In contrast, the difference in the predictive margin to cooperate between those that think the right thing to do is to cooperate, or to defect, is much more pronounced across all treatments.

Turning to normative expectations (right panel of Figure A-1), we first see that these expectations are a less precise predictor of cooperation than personal normative expectations: The standard errors around the point estimates are larger, the point estimates are closer to the 0.5 line, and the predictors are less consistent across treatments. For the normative expectation that most others think that one ought to "do what others do", we do see that the marginal effects in both the hi-S and hi-noS treatments are larger than the effects for low-S and low-noS treatments.

Additionally, Table A-5 shows the robustness of the result in Figure A-1 to the exclusion of the one-shot game, see column (2), and the exclusion of the last round, see column (3).

Table A-5: Non-linear model explaining the effect of normative beliefs on cooperation (PNB refers to personal normative belief, and NE refers to normative expectations)

	All Rounds Cooperation (1)	*Without OS* Cooperation (2)	*Without Last Round* Cooperation (3)
low-noS	0.337	0.241	0.186
	(0.233)	(0.302)	(0.232)
low-S	-0.207	-0.502	-0.276
	(0.260)	(0.371)	(0.226)
hi-noS	-0.304	-0.537*	-0.306
	(0.249)	(0.304)	(0.275)
Round	-0.025	-0.051	0.002
	(0.033)	(0.048)	(0.050)
low-noS × Round	-0.166***	-0.137**	-0.191***
	(0.053)	(0.064)	(0.055)
low-S × Round	-0.017	0.027	-0.057
	(0.042)	(0.069)	(0.055)
hi-noS × Round	-0.183***	-0.172***	-0.202***
	(0.050)	(0.056)	(0.069)
PNB (DwoD)	1.665***	1.731***	1.619***
	(0.187)	(0.176)	(0.243)
low-noS × PNB (DwoD)	-0.527	-0.714*	-0.624
	(0.398)	(0.407)	(0.432)
low-S × PNB (DwoD)	0.331	0.339	0.319
	(0.531)	(0.572)	(0.532)
hi-noS × PNB (DwoD)	0.612	0.561	0.599
	(0.404)	(0.453)	(0.402)
PNB (Coop)	1.634***	1.659***	1.646***
	(0.501)	(0.613)	(0.510)
low-noS × PNB (Coop)	-0.209	-0.264	-0.228
	(0.655)	(0.744)	(0.650)
low-S × PNB (Coop)	0.619	0.657	0.544
	(0.632)	(0.732)	(0.639)
hi-noS × PNB (Coop)	0.905	0.984	0.786
	(0.676)	(0.764)	(0.670)
NE (DwoD)	1.493**	1.611***	1.426**
	(0.583)	(0.586)	(0.551)
low-noS × NE (DwoD)	-0.036	-0.072	0.019
	(0.735)	(0.767)	(0.665)
low-S × NE (DwoD)	-1.444**	-1.628**	-1.341**
	(0.654)	(0.683)	(0.629)
hi-noS × NE (DwoD)	-0.454	-0.525	-0.637
	(0.681)	(0.722)	(0.659)
NE (Coop)	1.181**	1.137*	1.125**
	(0.562)	(0.679)	(0.567)
low-noS × NE (Coop)	0.407	0.383	0.413
	(0.667)	(0.785)	(0.668)
low-S × NE (Coop)	-0.632	-0.655	-0.532
	(0.698)	(0.796)	(0.704)
hi-noS × NE (Coop)	-0.921	-1.074	-0.940
	(0.719)	(0.812)	(0.710)
Session Fixed Effects	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes
N	4116	3528	3528

Notes: The table reports random effect estimators for an individual-level probit model. The baseline is set to the hi-S treatment. Robust standard errors are clustered at the session level (in parentheses). Individual controls include age, age squared, an indicator variable for comprehension, a risk preference measure, the main type of gear, and crew size. ***, **, and * indicate significance at the 1,5, and 10% level.

Table A-6 depicts robustness tests for Table 2.5 in Section 2.5.2. In particular, we show that the results for the regression of voting to exclude defectors on being excluded in the previous round and the regression of cooperation and empirical expectation on the same set of covariates are robust to the choice of a linear probability model as discussed in Appendix A-1. Other specifications are unchanged.

Table A-6: Estimates of the linear model explaining the use and effect of the sanctioning institution

	Vote to Exclude Defectors	Cooperation	Emp. Expectation
	(1)	(2)	(3)
low-S	-0.188***	-0.287***	-0.234***
	(0.061)	(0.066)	(0.056)
excluded (def in $t-1$)	0.115***	0.005	0.000
	(0.013)	(0.026)	(0.033)
excluded (def in $t-1$) x low-S	-0.038	-0.051	0.010
	(0.059)	(0.060)	(0.056)
excluded (coop in $t-1$)	0.049	-0.041**	-0.025
	(0.035)	(0.020)	(0.038)
excluded (coop in $t-1$) x low-S	0.053	0.047	0.019
	(0.053)	(0.048)	(0.054)
Round	0.003	-0.009	-0.012**
	(0.007)	(0.008)	(0.006)
Round x low-S	-0.000	0.004	0.010
	(0.010)	(0.013)	(0.009)
coop in t	-0.078***		
	(0.030)		
Session Fixed Effects	Yes	Yes	Yes
Individual Control	Yes	Yes	Yes
R^2	0.088	0.105	0.105
N	1764	1764	1764

Notes: The table reports random effect estimators for an individual-level linear probability model. The baseline is set to the hi-S treatment. The model includes one-shot game and all repeated rounds. Individual controls include age, age squared, an indicator variable for comprehension, a risk preference measure, the main type of gear, and crew size. Robust standard errors are clustered at the session level (in parentheses). ***, **, and * indicate significance at the 1,5, and 10% level.

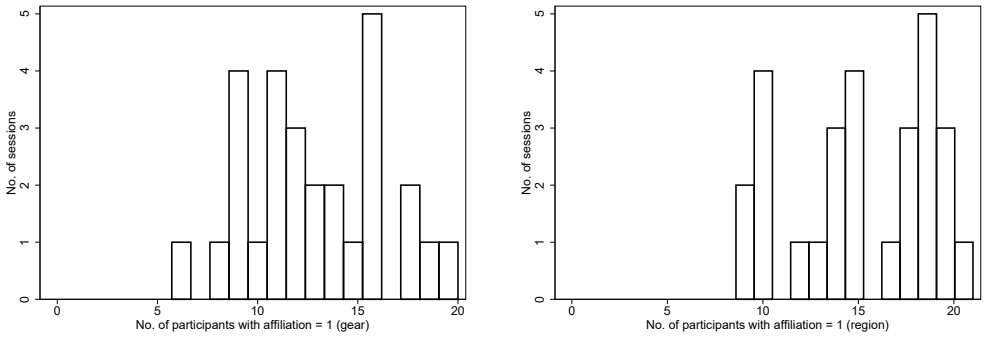
A-3 Robustness Analysis of Social Proximity Result

We now turn to the robustness of the analysis presented in Section 2.6.

Table A-7 depicts the random effects probit regression that underlies the predictive margins shown in Figure 2.5. Results strongly indicate that participants that are close to their peers based on fishing related characteristics exhibit behavior that is in line with the social norm in the respective treatment. That is, participants with high social proximity, on average, cooperate more in the hi-S treatment (baseline treatment in column (1)) and less in the low-S treatment than those that have low social proximity. The effect of social proximity indicates that it is those participants who are close to their peers in the session that drive the overall result. Participants with a low value for social proximity exhibit behavior that is not in line with the social norm.

Table A-8 shows that results are robust to the choice of a linear probability model. Other specifications are unchanged. In the linear model we find that an increase in our social proximity measure by one (see Section 2.3 and below for a detailed description of the index's construction), on average, leads to 10.2 percentage points higher cooperation rates in the hi-S treatment. The coefficient indicates that those with strong social proximity are more likely to conform with the norm of cooperation. Similarly, results display that an increase in the proximity measure by one, on average, leads to 15.9 percentage points lower cooperation rates in the low-S compared to the hi-S treatment. Social proximity does not have a significant effect in either treatment without a sanctioning institution. Results for the empirical expectation measure follow the same pattern. That is, social proximity is associated with higher expectations in the hi-S and lower expectations in the low-S treatment, reflecting the relationship between social proximity and cooperation, and therefore giving a strong indication for the creation of different social norms.

In the following, we take a closer look at how the social proximity index is constructed. Figures 2.2(a) and 2.2(b) depict the distribution of session by the number of participants with the indicator variable for main gear type (Figure 2.2(a)) and region of origin (Figure 2.2(b)) $= 1$.



(a) Main gear

(b) Region of origin

Fig. A-2: Histogram of sessions by number of participants giving the modal response to the respective proximity indicator

For participants' main type of gear, we observe the lowest value in one session where only

Table A-7: Estimates of non-linear model explaining the effect of social proximity

	Cooperation	Empirical Expectations
	(1)	(2)
Social Proximity (SP) = 0	-0.375	-1.081***
	(0.417)	(0.404)
SP = 0 × hi-noS	0.311	1.417***
	(0.544)	(0.524)
SP = 0 × low-noS	0.251	1.331*
	(0.606)	(0.758)
SP = 0 × low-S	0.936	1.548***
	(0.648)	(0.582)
SP = 1	0.520**	0.523***
	(0.218)	(0.155)
SP = 1 × hi-noS	-0.378	-0.468
	(0.433)	(0.318)
SP = 1 × low-noS	-0.778	-0.590
	(0.495)	(0.466)
SP = 1 × low-S	-0.825**	-0.732**
	(0.324)	(0.305)
hi-noS	-0.115	-0.238
	(0.325)	(0.268)
low-noS	0.216	-0.495
	(0.298)	(0.326)
low-S	-0.554**	-0.813***
	(0.255)	(0.283)
Round	-0.025	-0.025
	(0.032)	(0.032)
Round × hi-noS	-0.181***	-0.071*
	(0.049)	(0.036)
Round × low-noS	-0.172***	-0.022
	(0.053)	(0.044)
Round × low-S	-0.018	0.022
	(0.041)	(0.036)
Session fixed effects	Yes	Yes
Individual controls	Yes	Yes
N	4116	4116

Notes: The table reports random effect estimators for an individual-level probit model. The baseline is set to the hi-S treatment. For the social proximity variable, the baseline is set at 0.5. The model includes one-shot game and all repeated rounds. Individual controls include age, age squared, an indicator variable for comprehension, a risk preference measure, and crew size. Robust standard errors are clustered at the session level (in parentheses). ***, **, and * indicate significance at the 1,5, and 10% level.

Table A-8: Estimates of linear model explaining the effect of social proximity

	Cooperation	Empirical Expectations
	(1)	(2)
Social Proximity (SP) = 0	-0.046	-0.175**
	(0.075)	(0.085)
SP = 0 × hi-noS	0.056	0.260**
	(0.097)	(0.107)
SP = 0 × low-noS	0.001	0.219
	(0.119)	(0.142)
SP = 0 × low-S	0.177	0.314***
	(0.111)	(0.109)
SP = 1	0.102***	0.075**
	(0.038)	(0.030)
SP = 1 × hi-noS	-0.068	-0.049
	(0.082)	(0.057)
SP = 1 × low-noS	-0.159*	-0.094
	(0.081)	(0.083)
SP = 1 × low-S	-0.159***	-0.086
	(0.061)	(0.056)
hi-noS	-0.036	-0.048
	(0.057)	(0.047)
low-noS	0.047	-0.107*
	(0.047)	(0.058)
low-S	-0.108**	-0.151***
	(0.044)	(0.055)
Round	-0.005	-0.005
	(0.006)	(0.006)
Round × hi-noS	-0.032***	-0.014**
	(0.008)	(0.006)
Round × low-noS	-0.029***	-0.004
	(0.009)	(0.008)
Round × low-S	-0.003	0.004
	(0.008)	(0.007)
Session fixed effects	Yes	Yes
Individual controls	Yes	Yes
N	4116	4116

Notes: The table reports random effect estimators for an individual-level OLS model. The baseline is set to the hi-S treatment. For the social proximity variable, the baseline is set at 0.5. The model includes one-shot game and all repeated rounds. Individual controls include age, age squared, an indicator variable for comprehension, a risk preference measure, and crew size. Robust standard errors are clustered at the session level (in parentheses). ***, **, and * indicate significance at the 1,5, and 10% level.

six participants have the indicator variable = 1, see Figure 2.2(a). So, in this one session, the modal response to the question regarding main gear type is given by six of 21 participants. In most other sessions the modal response is given by considerably more participants. The distribution is skewed to the left (mean = 13) and offers considerable variation (sd = 3.61). In fact, in 21 out of 28 sessions the modal response is given by the majority (≥ 11) of the 21 participants.

For participants' region of origin, we observe the lowest value in two different sessions, where nine participants have the indicator variable = 1, see Figure 2.2(b). So, in these two sessions, the modal response to the question regarding the region of origin is given by nine out of 21 participants. In most other sessions the modal response is given by more participants. Also, the distribution for region of origin is skewed to the left (mean = 15.43) and offer considerable variation (sd 3.87). In fact, in 22 out of 28 session the modal response is given by the majority (≥ 11) of the 21 participants.

Table A-9 and A-10 report robustness results for the reference network analysis presented in Section 2.6. Regression results for the social proximity measure on cooperation (Table A-9) and empirical expectation (Table A-10) are robust to alternative specifications of the proximity index. Column (1) in both tables, respectively, repeats the regressions from Table A-8 with social proximity interpreted as a numerical variable instead of a categorical variable. Results are virtually unchanged, and to economize the exposition we show results where social proximity is employed as a numerical variable in the rest of the tables. In column (2) the index is disaggregated and gear use as well as region of origin proximity indicators are single covariates. Columns (3) and (4) model social proximity with only one of the indicator variables. In column (5) and (6), different weights are applied to both indicator variables in calculating the index. In column (5), 25% of the index is calculated with the region of origin variable while 75% is explained with gear use. In column (6), the weights are turned around, 75% for gear use and 25% for region of origin.

For the regressions on cooperation and on empirical expectation, different specification of the proximity index yields significant results. The robustness indicates that both gear use and region of origin as a proxy for ethnicity indeed have explanatory power for our measure of social norms. Individuals with high social proximity are more receptive to the social information when sanctions are possible.

Table A-9: The effect of social proximity (SP) on cooperation is robust to various index compositions

	Cooperation					
	(1)	(2)	(3)	(4)	(5)	(6)
low-noS	0.176	0.042	0.124	0.084	0.155	
	(0.116)	(0.060)	(0.100)	(0.070)	(0.114)	
low-S	0.054	-0.077	-0.035	-0.012	0.028	
	(0.087)	(0.066)	(0.068)	(0.076)	(0.078)	
hi-noS	0.016	-0.001	-0.032	0.022	-0.002	
	(0.077)	(0.074)	(0.057)	(0.084)	(0.069)	
SP	0.162**					
	(0.068)					
SP × low-noS	-0.225					
	(0.145)					
SP × low-S	-0.317***					
	(0.106)					
SP × hi-noS	-0.126					
	(0.115)					
SP (100% gear)		0.064	0.073*			
		(0.044)	(0.044)			
SP (100% gear) × low-noS		-0.048	-0.053			
		(0.072)	(0.071)			
SP (100% gear) × low-S		-0.125	-0.125			
		(0.076)	(0.078)			
SP (100% gear) × hi-noS		-0.051	-0.059			
		(0.096)	(0.095)			
SP (100% origin)		0.099		0.106*		
		(0.062)		(0.062)		
SP (100% origin) × low-noS		-0.200*		-0.207*		
		(0.112)		(0.112)		
SP (100% origin) × low-S		-0.196**		-0.196**		
		(0.079)		(0.079)		
SP (100% origin) × hi-noS		-0.074		-0.079		
		(0.077)		(0.075)		
SP (25% origin, 75% gear)					0.121**	
					(0.054)	
SP (25% origin, 75% gear) × low-noS					-0.126	
					(0.103)	
SP (25% origin, 75% gear) × low-S					-0.223**	
					(0.097)	
SP (25% origin, 75% gear) × hi-noS					-0.097	
					(0.112)	
SP (75% origin, 25% gear)						0.148**
						(0.073)
SP (75% origin, 25% gear) × low-noS						-0.258*
						(0.146)
SP (75% origin, 25% gear) × low-S						-0.286***
						(0.098)
SP (75% origin, 25% gear) × hi-noS						-0.112
						(0.094)
Round Trend	Yes	Yes	Yes	Yes	Yes	Yes
Session Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	4116	4116	4116	4116	4116	4116

Notes: The table reports random effect estimators for an individual-level linear probability model. The baseline is set to the hi-S treatment. The model includes one-shot game and all repeated rounds. Round trends are interacted with treatment. Individual controls include age, age squared, an indicator variable for comprehension, a risk preference measure, and crew size. Robust standard errors are clustered at the session level (in parentheses). ***, **, and * indicate significance at the 1,5, and 10% level.

Table A-10: The effect of social proximity (SP) on empirical expectations is robust to various index compositions

	Empirical Expectation					
	(1)	(2)	(3)	(4)	(5)	(6)
low-noS	0.054	0.078	-0.046	0.003	0.004	0.050
	(0.128)	(0.133)	(0.078)	(0.106)	(0.100)	(0.129)
low-S	0.049	0.062	-0.078	-0.018	-0.017	0.040
	(0.095)	(0.097)	(0.074)	(0.083)	(0.087)	(0.093)
hi-noS	0.126	0.126*	0.043	0.051	0.092	0.103
	(0.080)	(0.073)	(0.070)	(0.059)	(0.079)	(0.067)
SP	0.226***					
	(0.080)					
SP × low-noS	-0.275*					
	(0.165)					
SP × low-S	-0.320***					
	(0.101)					
SP × hi-noS	-0.255***					
	(0.097)					
SP (100% gear)		0.091**	0.104***			
		(0.039)	(0.039)			
SP (100% gear) × low-noS		-0.081	-0.090			
		(0.089)	(0.087)			
SP (100% gear) × low-S		-0.095	-0.099			
		(0.069)	(0.072)			
SP (100% gear) × hi-noS		-0.087	-0.104			
		(0.073)	(0.071)			
SP (100% origin)		0.138**		0.146**		
		(0.069)		(0.073)		
SP (100% origin) × low-noS		-0.213*		-0.222*		
		(0.116)		(0.118)		
SP (100% origin) × low-S		-0.231***		-0.240**		
		(0.089)		(0.093)		
SP (100% origin) × hi-noS		-0.178**		-0.188**		
		(0.083)		(0.084)		
SP (25% origin, 75% gear)					0.170***	
					(0.059)	
SP (25% origin, 75% gear) × low-noS					-0.175	
					(0.126)	
SP (25% origin, 75% gear) × low-S					-0.201**	
					(0.090)	
SP (25% origin, 75% gear) × hi-noS					-0.180**	
					(0.088)	
SP (75% origin, 25% gear)						0.207**
						(0.086)
SP (75% origin, 25% gear) × low-noS						-0.289*
						(0.156)
SP (75% origin, 25% gear) × low-S						-0.322***
						(0.107)
SP (75% origin, 25% gear) × hi-noS						-0.250***
						(0.092)
Round Trend	Yes	Yes	Yes	Yes	Yes	Yes
Session Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	4116	4116	4116	4116	4116	4116

Notes: The table reports random effect estimators for an individual-level linear probability model. The baseline is set to the hi-S treatment. The model includes one-shot game and all repeated rounds. Round trends are interacted with treatment. Individual controls include age, age squared, an indicator variable for comprehension, a risk preference measure, and crew size. Robust standard errors are clustered at the session level (in parentheses). ***, **, and * indicate significance at the 1,5, and 10% level.

Chapter 3

Changing Collective Action: Social Information Increases Cooperation of Teams in a Prisoner's Dilemma

Joint work with Florian Diekert.

Abstract: We test whether social information about the cooperative behavior of others increases cooperation in a prisoner's dilemma. Our novelty is that actors are not individuals, but teams. Teams in our experiment are fishing crews from Lake Victoria, Tanzania. We randomize two decision making mechanisms across a social information treatment: teams either decide through majority voting or the dictatorial choice of one member. Since both mechanisms are present at Lake Victoria, we can identify experience with hierarchical or egalitarian decision structures as a driver of behavioral change. Providing information on the behavior of other teams increases average cooperation by 14 and 17 percentage points for egalitarian and hierarchical team decisions, respectively. Further, participants with experience in hierarchical decision structures are particularly responsive to social information.

Keywords: collective action; team behavior; social information; common pool resources; lab-in-the-field experiment

3.1 Introduction

Groups, not individuals, are the key actors in many social dilemmas. For example, decisions on how to combat climate change are taken by countries. Similarly, corporations interact or collude in competitive markets. And further, many natural resources are managed by communities and harvested by groups or teams of individuals. Due to difficulties in monitoring and enforcement, these social dilemmas cannot be solved by formal regulations alone. Therefore, activists, scholars, and policy makers advocate "social norms as solutions" (Nyborg et al., 2016). By now there is ample evidence that norm-based interventions, such as providing information on the behavior of others, can indeed nudge individuals towards socially desirable actions (Farrow et al., 2017; Bergquist et al., 2019). In particular, there is direct (Diekert et al., 2021) and indirect evidence (Lopez et al., 2012) that social information increases cooperation in a social dilemma. However, it is not known whether the provision of social information is effective when groups or teams of individuals make joint decisions. This begs the question, can social information also induce a change in collective actions?

In this paper, we present a prisoner's dilemma experiment with unitary teams that studies the effect of social information on team cooperation.[1] We vary whether teams are informed about the cooperative behavior of other teams in a previous experimental session and test the effectiveness of our intervention on naturally occurring teams. Each team consists of three fishermen that work together as a fishing crew at Lake Victoria, Tanzania. Fisheries at Lake Victoria are a particularly good setting for our research as fishermen work in teams and face the social dilemma of common-pool resource use every day. Moreover, there is a need for informal governance as overfishing threatens the income and food security of more than four million people in the region while formal regulations remain ineffective (Mkumbo and Marshall, 2015; Irvine et al., 2019).

Early contributions on the effect of social information focus on the promotion of environmentally-friendly consumer behavior such as energy use reduction (Allcott, 2011; Costa and Kahn, 2013) and water conservation (Ferraro et al., 2011). While applications have broadened to include studies on, *e.g.*, charitable giving (Croson et al.,

[1] In unitary teams, members have to make a joint decision and receive identical payoffs such that there is no material conflict of interest within teams (Kocher et al., 2020). The focus on unitary teams distinguishes our design from the literature that studies multi-level public good games (Blackwell and McKee, 2003; Buchan et al., 2009; Gallier et al., 2019). Similarly, the literature on group contests (Sheremeta, 2018) studies situations in which team members have an incentives to exert effort in order to win a between-group competition while simultaneously having incentives to free ride on the efforts of other members.

2009; Goeschl et al., 2018) and tax compliance (Hallsworth et al., 2017), social information has been predominantly used in settings where agents (i) decide by themselves and (ii) do not directly interact with each other. The successful use of social information in the public goods game (Fehr and Schurtenberger, 2018; Diekert et al., 2021) and the trust game (Bicchieri et al., 2020) has established that the intervention is also effective in strategic interactions but different levels of decision making have not yet been addressed. We therefore contribute to the understanding of social norms as policy tools by shifting focus to team-level decision making.

Decision making for teams that interact with other teams differs from decision making for individuals that interact with other individuals. For example, consider a social dilemma. In a social dilemma between teams, there are two decision tasks. First, members of a team have to coordinate among each other to reach a joint decision. Second, the team as a whole interacts with other teams on solving the dilemma. When individuals interact with other individuals, they only face the second task. The behavioral implications of the differences between individual and team decision making are documented by a long-standing literature in economics (Kugler et al., 2012; Kocher et al., 2020) and social psychology (Schopler and Insko, 1992; Wildschut et al., 2007): teams make more rational and selfish decisions. In other words, individual decisions are not necessarily a good predictor of team decisions (Charness and Sutter, 2012). Consequently, findings on the effect of social information on individual decisions cannot be used to infer their effect on team decisions.[2]

In general, comparing the behavior of individuals and teams is of little use for practical purposes as changing who is the relevant actor is rarely a viable policy option. A more relevant policy question is whether social information can at all be used to induce behavioral change when decisions are made by teams. Moreover, policy makers would want to know whether the success of the intervention depends on inherent characteristics of team decision making. Here, we address the fact that teams differ in the mechanism they use to reach a joint decision. In particular, we study the effect of social information for (i) a hierarchical decision structure, and (ii) an egalitarian decision structure. To capture an egalitarian decision structure, we impose that teams use majority voting. That is, all team members state their preferred action and the behavioral response that is favored by the majority of individuals within the

[2]Social information affects both the first task of intra-team coordination and the second task of inter-team cooperation. Both tasks are likely to interact in complex ways. Hence the effect of social information on individual cooperation decision is not a *ceteris paribus* counterfactual that identifies the effect of being in a team, conditional on social information (Manski, 1993).

team is chosen as the team's action. In a hierarchical decision structure, we impose that teams reach a decision through the choice of one team member.

While some examples of egalitarian organizations exist, hierarchical organizations are the more dominant decision making structure in the economy (Coase, 1937; Williamson, 1967). For example, committees have a president, corporations have an executive, and fishing crews have a captain. In contrasting this decision structure with the egalitarian approach of majority voting, we leverage a special feature of our field setting: At Lake Victoria, only about half of all fishing crews have a designated captain who dictates critical production decisions. In the other half, all crew members (including the captain) decide together. In particular, we use the fact that the coordination mechanism in our experiment is imposed in random treatment assignment. Hence, some fishing crews are able to apply their experience with either a hierarchical or an egalitarian decision structure. Other crews are put in the position to determine the team's action in a structure that they are unfamiliar with. This allows us to analyze whether the effect of social information in the experiment depends on participants' real life experience with the respective decision making structure.

Our results suggest that social information induces an increase in cooperation when decisions are made by teams. For both dictatorial decisions and majority voting, cooperation rates with social information are significantly higher than without. Yet, we observe a change in the perceived social norm only for dictatorial decision making. The result is consistent with the related literature, indicating that a behavioral change is mediated by a change in beliefs. With a majority voting mechanism, social information has a negligible effect on beliefs. Moreover, we identify heterogeneous treatment effects with respect to experience. For the dictatorial mechanism, the overall treatment effect of social information is driven by those participants that have real life experience with making decisions in a hierarchical decision structure. In contrast, the overall treatment effect for majority decisions is driven by teams without experience in egalitarian decision structures.

3.2 Literature

We provide unitary teams with information about the behavior of other teams in a social dilemma. To assess whether such a social information message is likely to increase cooperation, we first survey both the related literature on team decision making and the literature on norm-based interventions. What are the drivers of team decisions, do those drivers vary for different team coordination mechanisms, and

what is known about the processes through which social information messages induce behavioral change?

3.2.1 Team Decision Making

While the neo-classical theory of self-interested, rational agents fails to predict individual behavior, it applies remarkably well to team decisions (Kugler et al., 2012; Charness and Sutter, 2012). Teams make patient, time-consistent choices (Shapiro, 2010; Denant-Boemont et al., 2017) and seek efficient outcomes, *i.e.*, they are good at earning high payoffs through cognitive sophistication (Charness and Sutter, 2012) and the avoidance of miscoordination (Feri et al., 2010). Moreover, teams generally transfer small amounts in non-strategic allocation tasks such as the dictator game (Luhan et al., 2009).[3]

Team behavior is more self-interested and rational than individual behavior also in strategic interactions. Experiments in economics (Kagel and McGee, 2016) and social psychology (Schopler and Insko, 1992) show that teams maximize own benefits over making socially optimal decisions in a social dilemma.[4] There are three main motivations why teams are non-cooperative when they interact with other teams: (i) social support for self-interest, (ii) the fear of exploitation by other teams, and (iii) the evasion of responsibility.

First, teams defect in a social dilemma because team members have a preference to benefit the in-group. While choosing a non-cooperative strategy hurts the out-group, it maximizes the in-group payoff. That is, defection benefits oneself and one's team members. It can be rationalized as an act of shared self interest (Insko et al., 1990; Kugler et al., 2012) in which pro-social preferences towards other groups are crowded out by parochial altruism (Charness and Chen, 2020). Such altruism is consistent with a let-down aversion of one's own team (Charness and Holder, 2019) as cooperation is costly not only at the expense of own payoffs but also at the expense of the in-group.

[3]Further research on behavioral regularities in team decisions includes studies on uncertainty preferences, see Kocher et al. (2020) for a review. While Stoner (1961) is the first study to show a so-called "risky shift" in team decisions as a result of polarized attitudinal judgments, evidence with respect to teams' risk preferences remains inconclusive as later studies fail to replicate Stoner's results (Baker et al., 2008) or even show that teams are rather risk averse decision makers (Shupp and Williams, 2008; Masclet et al., 2009).

[4]In other related strategic settings, similar results are found. In the ultimatum game (Bornstein and Yaniv, 1998), the trust game (Kugler et al., 2007), or in a gift-exchange game (Kocher and Sutter, 2007) teams exhibit only limited pro-social behavior. The finding is labeled as "discontinuity effect" in the social psychology literature to describe that team decisions are less pro-social than suggested by the aggregate preferences of all team members (Insko et al., 1988, 1990; Schopler and Insko, 1992; Schopler et al., 1995).

Second, teams defect in a social dilemma because they fear the exploitation by other teams. Defection protects oneself and the in-group against a sucker payoff (Bornstein et al., 2004). Therefore, it becomes necessary as a defensive response when other teams are not trusted to cooperate (Kagel and McGee, 2016). Such a lack of trust in interactions between teams is documented in Kugler et al. (2007) and Song (2009), who find that teams expect other teams to act selfish.

Third, teams defect in a social dilemma because team members evade the responsibility to cooperate. The individual within a team is not identifiable and thus cannot be solely held accountable for a selfish choice (Schopler et al., 1995; Kugler et al., 2012). The resulting diffusion of responsibility facilitates selfish behavior (Charness, 2000). Moreover, the lack of distinct identification allows team members to hide behind a "shield of anonymity". This increases the social distance between teams which is detrimental for cooperation (Bohnet and Frey, 1999).[5]

The motivations of shared self-interest and the fear of exploitation apply to all team decisions, but the possibility to evade responsibility depends on the structure that teams use to make a joint decision (Song, 2009). In a hierarchical decision structure, diffusion of responsibility is not possible. The person that dictates the action for the team is identifiable, and by the converse argument of Charness (2000), should cooperate more. However, Atanasov and Kunreuther (2016) show that team representatives are cautious decision makers that act tough as they worry about the impression they make with their team, unwilling to let down other members (Dufwenberg and Gneezy, 2000). The responsibility for in-group payoffs can therefore also crowd out cooperation between teams (Charness and Jackson, 2009; Humphrey and Renner, 2011).

To summarize, several motivations including shared self-interest among team members, the fear of exploitation by others, and the lack of responsibility drive teams towards self-interested behavior. Consequently, there is a need for tools that focus team decisions on socially optimal strategies.

[5]Comparing the studies by (Cason and Mui, 1997) and Luhan et al. (2009), who study team decisions in a dictator game, highlights the implications of identifiability. While Luhan et al.'s (2009) finding of selfish allocations is in line with the large majority of the literature on team decisions, it contrasts with the results in Cason and Mui (1997), who observe more altruistic decisions. Among the differences between the two studies that may cause those higher transfers is the fact that Cason and Mui (1997) publicly identify single team members when teams are formed and thereby remove the shield of anonymity.

3.2.2 Behavioral Change through Social Information

Behavior and opinion of others are powerful drivers of individual decision making. A large number of successful social information interventions leverage this fact, see Farrow et al. (2017) and Bergquist et al. (2019) for reviews. Studies by Croson et al. (2009) and Bicchieri and Xiao (2009) show how social information activates social norms: After individuals receive credible information about what others do (descriptive norm) or about what others consider to be appropriate (injunctive norm), they update their expectations about others' beliefs and behavior.[6] Thereby, social information changes the perceived social norm. Both injunctive and descriptive information have been successfully used in norm activation. They rely on linking a socially accepted value to a target behavior (Miller and Prentice, 2016). As long as individuals perceive the behavior of others to be a desirable action, or have a preference for conformity, a change in the perceived social norm can cause a change in individual behavior.

A number of theoretical models formalize the intuition that violating social norms introduces discomfort (Sugden, 2000; Kimbrough and Vostroknutov, 2016; Michaeli and Spiro, 2017). The discomfort increases in the mismatch between own actions and the actions of others such that following the norm avoids disutility. By changing the perceived social norm, social information induces such a conformity response. Depending on the behavior in questions, social information is however not limited to activating preferences for conformity as such. For example, a message on others' behavior in a social dilemma conveys useful information on (i) which behavior is desirable, (ii) which behavior may lead to equitable outcomes, and (iii) whether one should fear the exploitation by others. Here, norm activation can also work through norms of fairness and norms of trust.

The credibility of the social information message and the relevance of the reference group are two main principles that have been identified for a successful design of interventions that target social norms (Miller and Prentice, 2016; Bicchieri and Dimant, 2019). Agents may disregard a message about others' behavior or beliefs if it (i) does not come from a trusted source, or (ii) does not draw a comparison to a relevant social group. Moreover, Diekert et al. (2021) show that in strategic settings, the effect of social information increases with social proximity between agents. We

[6] Belief formation is usually tracked through incentivized elicitation of injunctive and descriptive beliefs (Bicchieri, 2017). Yet, dependencies between stated beliefs and behavior complicate the identification of normative or descriptive beliefs as driving mechanisms of behavioral change. While some agents' behavior is driven by beliefs, others may state a certain belief to justify their behavior (Andreoni and Sanchez, 2014). Eymess (2021) studies the interplay of beliefs, behavior, and social norms with a conditional process analysis.

are not aware of a study that tests whether the effect of social information depends on real-life experience, but we note that experience with the decision situation ties in to the discussion on the message's overall relevance to its audience.

3.3 Fishing at Lake Victoria

The Lake Victoria fisheries in East Africa are an important driver of local and regional economies in the three countries that share the lake's resources: Kenya, Tanzania, and Uganda (see Figure 3.1). The income and food security of more than four million people is supported by a common pool resource system that is under pressure from overfishing, pollution, climate change, and rapid population growth (Cowx and Ogutu-Owhayo, 2019; Gichuru et al., 2019; Irvine et al., 2019).[7] As the demand for resources and food from the lake is steadily increasing and formal regulatory structures continue to be dysfunctional, it is urgent to find effective policies that balance the societal needs of both short-term resource exploitation and long-term conservation (Aura et al., 2019).

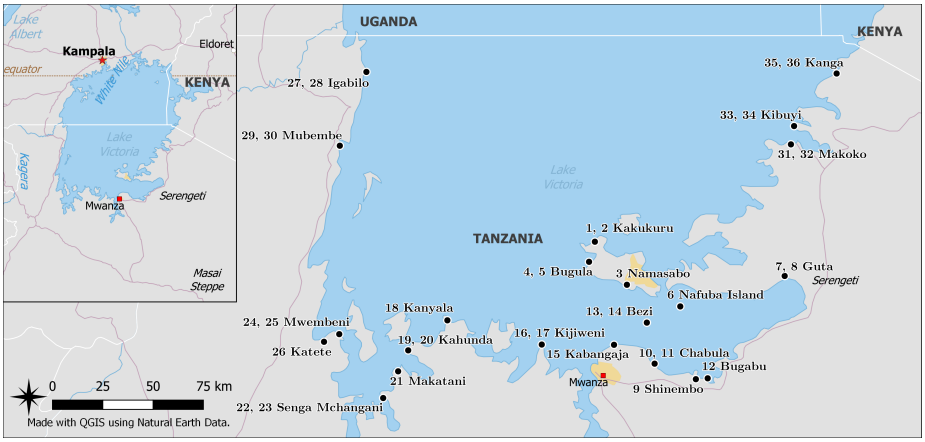


Fig. 3.1: Map of Lake Victoria and visited landing sites

[7] A series of publications by Jeppe Kolding and colleagues (Kolding et al., 2014, 2016, 2019) claims that a pessimistic focus on the problem of overfishing is misguided and stresses that issues of food security and nutrition should be at the forefront of governance. This underlines the fact that the Great African Lakes suffer from a multifaceted problem that is (i) inadequately addressed and (ii) in need of a comprehensive approach for effective and sustainable management.

The Need for Informal Governance

Due to strategic incentives, limited state capacity, and dysfunctional formal institutions, the enforcement of fishing regulations at Lake Victoria is weak. All adjacent countries have passed fisheries regulations to govern issues such as licensing, gear use and the protection of breeding areas. Yet, the violation of regulations is common and attempts to reduce illegal fishing practices are plagued by issues of corruption (Nunan et al., 2018). To help monitor and enforce regulations, landing site level co-management structures known as beach management units (BMU), *i.e.*, elected community representatives, were introduced to the lake in the late 1990s. However, strong norms of kinship compromise the utilization of these co-management structures as formal enforcement mechanism (Etiegni et al., 2017). Simply devolving law enforcement from the national government to elected community representatives has not worked. Especially in situations of economic distress, fishermen at Lake Victoria continue to break regulations.

Resources are mostly contested on a local level since a large part of the fisheries rely on species that populate inshore areas (Taabu-Munyaho et al., 2013). When fishing crews choose to break regulations for their own economic benefits, they especially threaten the livelihoods of others in their own community or in neighboring communities. The social dilemma of common pool resource use can thus be broken down to the local level, generating important implications for policy makers that debate between bottom-up or top-down approaches to regulation. Through locally targeted interventions, social norms may be a promising tool to facilitate cooperation and self-management by resource users in local communities (Ostrom, 2008; Nyborg et al., 2016). Stakeholders that aim to ensure the sustainable use of Lake Victoria's resources in the long-term without jeopardizing the livelihoods of fishermen and their families in the short-term, may look to activate social norms of cooperation in the communities themselves.

Fishermen are Organized in Teams

Most fishermen at Lake Victoria work in teams. Figure 3.2 plots the distribution of crew size in our sample. Only about 3% of fishermen harvest on their own while 8% work in pairs. The fishery is dominated by small fishing crews of three (46%) and four (35%) members, indicating that the resource is contested on the boat level and not between individual fishermen.

Fishing crews at Lake Victoria can be distinguished by the structure through which

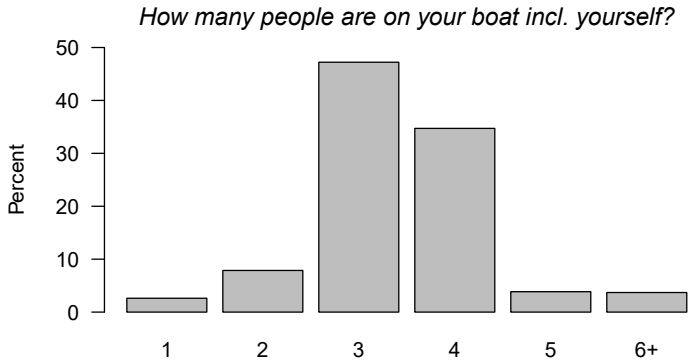


Fig. 3.2: Distribution of crew sizes of fishing crews at Lake Victoria, $N = 648$.

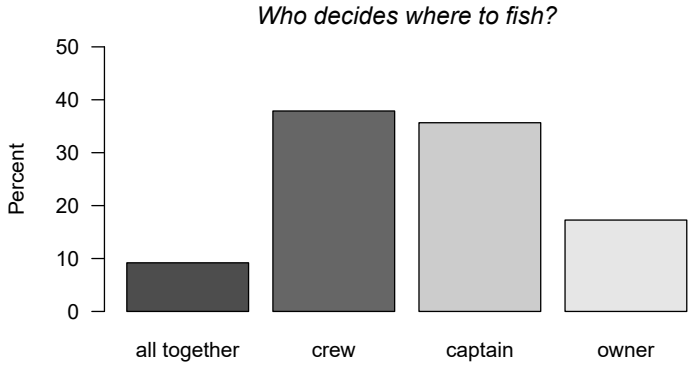


Fig. 3.3: Distribution of decision makers for the location of fishing (top, $N = 631$)

they reach joint decisions. An important daily decision that determines the catch is the location of fishing. The fishing location is either determined by all crew members together or by the crew's captain or boat owner. Figure 3.3 shows the distribution of decision makers for the fishing location in our sample. The two darker colored bars indicate an egalitarian structure in which the decision is either made by all fishermen together (including the owner who often stays ashore) or the crew that goes out for fishing. In contrast, the two lighter colored bars indicate a hierarchical structure in which the decision is either made by the captain or the boat owner. We observe that the distribution between the two forms of decision structures is about equal (47% egalitarian to 53% hierarchical structure). Hence, the data not only suggests that the social dilemma of common pool resource use at Lake Victoria needs to be solved by teams but also that these teams use two different structures to reach a joint decision.

3.4 Experimental Design and Implementation

To model the social dilemma of common pool resource use, we utilize a repeated two-team prisoner's dilemma game with disapproval and incentivized belief elicitation. Three participants play together in a team. Two teams share an account with eight points. Both teams play with a binary choice set, framed as a decision to take four points from the shared account *(defect)* or leave the points in the shared account *(cooperate)*. Moves are made simultaneously. The points remaining in the shared account are increased and then distributed equally. For four points left in the shared account, both teams receive three points, *i.e.*, a marginal per capita return of 0.75. The payoff matrix illustrates that defection is the payoff-dominant strategy while mutual cooperation is the social optimum, see Table 3.1. Points are later exchanged into real money.

Table 3.1: Payoff matrix

		Team B	
		cooperate	*defect*
Team A	*cooperate*	6,6	3,7
	defect	7,3	4,4

We run a social information treatment that is randomly assigned across sessions. Teams in the social information treatment (SI) are given information about past behavior of other teams in a previous session of the experiment while teams in the no social information treatment (noSI) play the prisoner's dilemma without prior information on others' behavior. By leveraging social comparison, the social information message is designed to affect participants' expectation about the upcoming interaction and activate social norms of cooperative play. The following message is verbally provided to all participants during the instructions of the game:

You are not the first landing site where fishermen participated in this survey. In a previous session, many/most teams left the points in the shared account.[8]

After all participants are fully informed about the game's rules, teams have to

[8]In Swahili, the meaning of both "many" and "most" is expressed by the word "wengi". Hence, the original Swahili message does not imply a strict majority but conveys the general information that cooperation was a common choice by other groups. The truthfulness of the message relies on data from one of the first sessions during the data collection (without social information) in which half the teams cooperated.

decide on an action. Participants privately and simultaneously choose whether they want their team to take four points from the shared account or leave the points in the shared account. To reach a joint decision, teams use an imposed coordination mechanism. The mechanism is varied across sessions such that within sessions all teams make their decisions in the same way. The other coordination mechanism is not mentioned. When teams are exposed to the hierarchical decision making structure, the team's action is determined by implementing the choice of a randomly selected team member. We call this the dictatorial decision mechanism. That is, all team members make a simultaneous choice before knowing whether their decision is implemented as the team's action. When teams are exposed to the egalitarian decision making structure, the team's action is determined by a majority vote. That is, all team members make a simultaneous choice before it is aggregated to a team decision by unanimity or a two-to-one split. We call this the majority voting mechanism. Participants are not informed about the identities of members in the other team. Direct communication or interaction within or across teams is not allowed.

After making their contribution decision, participants have the opportunity to express their disapproval of specific strategies. Each participant has to simultaneously choose one of the following options: (i) to disapprove defection, (ii) to disapprove cooperation, or (iii) to disapprove neither action. All participants are informed about the number of participants disapproving each option during feedback, see below. Disapproval votes are given without knowledge of the choices by other team members, the actions chosen by the other team that they are matched with, or the actions chosen by the other teams in the session.

Finally, participants are informed about the outcome of the prisoner's dilemma game. First, everyone within a team is informed about the choices of their team members and the resulting team action. Second, all participants within a team are informed about the action of the other team in their pairing. They are however not informed about the individual choices that lead to the aggregate decision of the other team. Third, participants are informed about their own team's total payoff from the prisoner's dilemma. No information is given about the outcomes in other team-pairings. Finally, everyone is informed about the number of participants in the session that disapprove of either action and the number of participants that do not disapprove of any action.

The game is played for five rounds. Teams are re-matched into new pairs based on a total stranger matching protocol, *i.e.*, for each new round of the game, teams are randomly matched with another team that they have not played with before and

will not play with afterwards. The composition of participants in a team is fixed over all five rounds. When teams use a dictatorial decision, each round has a new random draw to determine whose decision is chosen as the team's action.

The social information message is designed to affect descriptive beliefs. If participants see informational value in the message provided, descriptive beliefs should be adjusted. Specifically, we elicit participants' descriptive beliefs about others' behavior by asking what they "guess most teams in this session will actually do?". The elicitation of descriptive beliefs uses the same binary choice set as the cooperation decision and is incentivized with one extra point such that participants have no financial incentive to hide their true beliefs. Additionally, we elicit normative beliefs by asking participants what they think is the right thing to do in the given situation. For the elicitation of normative beliefs, participants can respond with a third option through which they can indicate a preference for conditional cooperation (*i.e.*, "do what other teams do"). While normative beliefs are only elicited in round one, we elicit descriptive beliefs in every round.

Additionally, we exploit the fact that the coordination mechanism in the experiment is imposed in random treatment assignment, such that some participants have experience with the decision structure while others have not. In particular, we use information on how the fishing unit that participates in our experiment determines the location for fishing in real life, see Figure 3.3. When we impose the majority voting mechanism, a team is categorized to have experience if the respective fishing crew determines the location of fishing (by themselves or together with the captain/owner). When the location is dictated by either the captain or the owner, the team is categorized as not having experience with an egalitarian decision structure. In sessions where a dictatorial decision determines the team's action, we categorize experience with a hierarchical decision structure on the individual level. Here, there are two dimensions to consider. First, the fisherman has to be part of a team that uses a hierarchical decision structure, and second, the fisherman has to be the one with the decision power. That is, we consider a participant to have experience with a hierarchical decision structure if the fisherman is the captain (owner) and reports that the fishing location is dictated by the captain (owner).

Implementation

The experiment was implemented with fishermen from Lake Victoria, Tanzania. The research trip comprised 36 sessions at 22 landing sites spanning the entire Tanzanian

coastline, see Figure 3.1. In a total of ten sessions (five in SI, five in noSI), we imposed the dictatorial decision mechanism while in 26 sessions (13 in SI, 13 in noSI), we imposed majority voting. Data was collected between March 9th and March 31st 2020.

For each session, six boats were randomly selected from the list of registered fishing vessels at a given landing site. From each boat, we then randomly selected three fishers that were willing to participate in the experiment as a team.[9] In each session, we therefore observe 18 participants in six teams that form three pairs in the prisoner dilemma in every round of the game. Participant characteristics are balanced across almost all relevant observables, see Appendix B-1. For the dictatorial coordination mechanism, the only difference between social information and no social information treatment is with respect to age. For the majority voting mechanism, the sample is unbalanced with respect to crew size and the propensity of fishermen that target Dagaa, proxied by the main type of gear used to catch this species.[10] We control for all unbalanced characteristics in our analysis, see Section 3.6.

A seating arrangement ensured that team members would not sit next to each other. Informed consent was obtained and a detailed explanation of the game's rules was given. In particular, it was highlighted that all decisions have to be made anonymously, that communication is not allowed, and that the points earned during the game directly translate to real money at the end of the experiment. To ensure that rules were understood, we played out test scenarios and assessed comprehension of the scenarios' outcomes with test questions. Responses serve as a measure of understanding in the analysis.

After all repetitions of the prisoner's dilemma game were completed, one round was randomly chosen for payout. The game was calibrated such that participants, independent of treatment, earn an average of approximately 2,700 Tanzanian Shilling (TZS).[11] In combination with an unrelated second experiment and a questionnaire to survey background information, each session lasted about two hours.

[9] Boats in our sample have an average crew size of 3.41 with the overwhelming majority of boats having three or four crew members, see Figure 3.2. That is, most often the random selection of three crew members sampled the whole or almost the whole crew.

[10] The Lake Victoria fisheries are split into targeting two main species for commercial use: (i) Dagaa which is fished at night, and (ii) Nile perch which is fished during the day.

[11] 2,700 TZS translates to approx. 1 Euro. The median daily catch earnings for a fisherman is about 5,000 TZS. All decisions in the experiment were made on tablet computers using the oTree software (Chen et al., 2016). Screenshots of all relevant choice situations (incl. end of the round feedback) and the experimental instructions are available in the supplementary material, published online.

3.5 Hypotheses and Statistical Methods

To the best of our knowledge, there is no study that tests whether a social information message is effective when decisions are made by teams. Furthermore, the literature on team decision making illustrates that we can neither extrapolate individual decisions to team decisions, nor can we hope to find an exhaustive and accurate theory that predicts how social information affects individual decisions in teams and how these aggregate to team actions.

On the one hand, one may reasonably expect that a social information message does work on teams. There is ample evidence that individuals hold significant other-regarding preferences (potentially differentiated by team membership) as well as preferences for conformity. Since a social information message is received and processed by individuals, and there is no reason why individual team members should ignore the message, one may expect that the positive effect of social information on cooperation also transpires in a team decision. On the other hand, one may reasonably expect that social information does not work on teams. The literature documents that teams act selfish and rational suggesting that a social information message that attempts to induce pro-social behavior may be unsuccessful with team decisions. If defection is the dominant strategy independent of belief about others' behavior (like in our experiment), we would then not expect any effect of a social information message.

In sum, there is little guidance on whether social information is successful in inducing pro-social actions or whether self-interest dominates in teams. Maintaining that individuals trust the source of social information and that the message has normative implications, at least for some individuals, there is no reason to believe that the message has a *negative* effect on cooperation. Hence, we formulated (and pre-registered[12]) directed hypotheses that predict a positive effect of social information on cooperation. Since the mechanism that teams use to coordinate on an action may significantly change how the social information message is perceived and responses play out in the intra- and inter-team processes, we discuss the expected effect of social information separately for majority voting and dictatorial decisions.

Majority Voting

We highlight three reasons why a social information message on the cooperative behavior of other teams could increase cooperation when teams use an egalitarian decision

[12]The pre-analysis plan is available at: `https://doi.org/10.1257/rct.5542-1.0`.

structure such as majority voting. First, team members may fear less exploitation by other teams. When other teams are expected to cooperate, they are not achieving gains at the cost of one's own group. Hence, social information decreases the need for in-group protection. Second, team members may be motivated by preferences for equitable outcomes (fairness) across teams. When other teams are expected to cooperate, cooperation of the own team is the fair response. Hence, team members with fairness preferences are more likely to vote for cooperation. Third, team members may be motivated by a preference for conformity. Then, they would experience disutility when the actions of the own team differ from the actions of other teams, or when the own vote differs from the votes of the other team members, or both. If the social information message is understood as a statement on the likely action of other teams, conformity preferences increase the likelihood to vote for cooperation.

These three motivations differ in underlying preferences, but they have the same observable implication for outcomes:

Hypothesis 1 *Average cooperation by teams with majority voting is higher with social information.*

Dictatorial Decisions

For teams with a hierarchical decision structure, the reduced fear of exploitation, preferences for fairness, and preferences for conformity are similarly suggestive of a positive effect of social information on cooperation. Yet, a hierarchical structure differs from an egalitarian structure because the dictator cannot evade the responsibility for the team's outcome, and thereby the responsibility for the outcome of each team member.

On the one hand, responsibility may complement the effect of a social information message. First, if dictators themselves have a preference for fairness or conformity with other teams, they should act in line with the social information message as they have the full responsibility for the team's decision. Second, when dictators believe that the members of their team have a preference for fairness or a preference for conformity with other teams and are unwilling to let their team members down, they would also want to act in line with the intervention. In both cases, the activation of social norms is amplified through responsibility. On the other hand, responsibility may be detrimental to the effect of a social information message. The obligation of representation may induce cautious decision making due to loss aversion and thereby crowd out cooperation. While such an adverse effect of responsibility would not cause

the intervention to backfire, it could force dictators to ignore the message.

Responsibility thus has an ambiguous impact on the effectiveness of social information in a hierarchical decision structure, but a reduced fear of exploitation, preferences for fairness, and preferences for conformity still suggest an increase of average cooperation:

Hypothesis 2 *Average cooperation by teams with a dictatorial choice is higher with social information.*

Experience

Experience with a given decision structure could moderate the effect of the social information intervention. If individuals are familiar with the decision structure, they may be more receptive to the social information message and have a better idea what to expect from other teams. Hence, experience would amplify the effect of the social information message. By implication, the effect may be muted for individuals without experience. They may be occupied with figuring out how to behave given the unfamiliar decision structure, or may wonder how to best interpret the social information message. Consequently, the social information message would not translate into action for individuals without experience.

That said, the effect of the social information message could also be weaker for individuals with experience. If they use internalized responses in familiar decision situations, they may put less weight on the social information message. In contrast, those that find themselves in an unfamiliar decision situation may view the information about others' behavior as especially valuable. These inexperienced individuals may think that they make no mistake when they do what others have done.

While we expected that the extent of familiarity with an hierarchical or egalitarian decision structure matters with respect to the social information treatment, we did not pre-register any hypothesis in this regard. Given the ex-ante ambiguous effect of experience, we treat it as an open question.

Statistical Methods

The main treatment effect of interest is the difference in average team cooperation over all five rounds between treatments with social information (SI) and without social information (noSI), see Hypotheses 1 and 2. In teams that reach a decision through majority voting, all three team member decisions are necessary to determine a team

action such that the outcome of interest is the team's aggregate decision. In teams that use dictatorial choice, each individual team members makes a simultaneous and private decision on behalf of the three person team before a random draw determines whose decision is implemented as the team's action. Hence, each individual decision is analyzed as a team decision.[13] We average the binary cooperation decisions over all five rounds and observe a cooperation rate (in discrete steps of 0.2 including zero and one) for $N = 156$ (78 in SI, 78 in noSI) teams that use majority voting and a cooperation rate for $N = 180$ (90 in SI, 90 in noSI) dictators.[14]

To test Hypotheses 1 and 2, we first present descriptive results and report non-parametric tests. Whenever we compare sample or subsample means to identify a social information treatment effect, we take the bi-modal distribution of our ordinal outcome variable into account and report Wilcoxon Mann-Whitney two-sample tests with exact p-values based on the actual distribution of the test statistic. Then, we support our results with regression analyses. Here, we use fractional probit models to examine the effect of social information on the team cooperation rate and report average marginal effects to ease the interpretation of coefficients. The main model of cooperation rate Y of team t on a set of covariates $\mathbf{x}$ (including a set of controls C) uses a quasi-likelihood estimation with the probit function G (Papke and Wooldridge, 1996) and is specified as follows:

$$E(Y_t|\mathbf{x}) = G(\beta_0 + \beta_1 \textit{Social info} + \beta_2 C_t)$$

We first analyze the model separately for each decision structure. For a pooled data analysis that compares the effect of social information between the egalitarian and hierarchical decision structure, an interaction of social information and decision structure is included. Likewise, we include additional interaction terms when we study heterogeneous treatment effects with respect to experience. In all specifications, standard errors are clustered on the session level to account for idiosyncratic conditions in the experimental setup at each landing site.

[13] Following Selten (1965), the elicitation method of asking everyone to make a decision before randomly determining which decision is carried out allows for an incentive compatible way of gathering data not only on those decisions that were implemented but also on those that were not implemented.

[14] We are not worried about learning effects or other dependencies that occur within team's over time. In fact, a part of the behavioral change induced by social information is likely to transpire through the process of retrospection and learning through the observation of other people's behaviors in repeated interaction. Hence, we choose to average cooperation over time as a measure that is able to capture such a social learning effect.

3.6 Results

We first analyze the main treatment effect of interest: Does a social information message increase average cooperation when decisions are made by teams? Since the effect of the intervention likely depends on the imposed coordination mechanism and in turn on the motivations associated with being a dictator or partaking in a majority decision, we study the main treatment effect for the two coordination mechanisms separately. In Section 3.6.3, we then pool all observations and present how real-life experience with the respective decision structure moderates our results. Finally, we turn to individual outcomes in Section 3.6.4. We analyze how the provision of social information affects descriptive beliefs and discuss to what extent normative beliefs and/or a desire for conformity may explain our results.

3.6.1 The Effect of Social Information on Majority Decisions

Figure 3.4 plots the distribution of cooperation rates for majority decisions by social information treatment. In both treatments, most teams use a strategy of zero cooperation across all rounds. Yet, with social information, the share of teams that cooperate in all five rounds increases. Hence, we observe a positive treatment effect of social information. The average team cooperation rate in the no social information treatment is 33.1% and increases by 13.8 percentage points to an average of 46.9% with social information, $p = .057$, combined $N = 156$.[15] The treatment effect is especially pronounced when limiting observations to contribution decisions in the first round. Here we observe a team cooperation rate of 28.2% without and 47.4% with social information, a difference of 19.2 percentage points that is significant with $p = .020$, combined $N = 156$.

Our results on the treatment difference suggest that the social information message is successful in changing team decisions reached with a majority voting mechanism. The result is supported by a marginally significant coefficient for the social information treatment in a fractional probit regression. For the average marginal treatment effect, see column (1) in Table 3.2. The regression model predicts that social information leads to an increase in team cooperation by, on average, 14.9 percentage points ($p = .060$). We therefore accept Hypothesis 1: Social information increases cooperation by

[15] One can argue that teams are not independent from each other within a session as they are informed on the decision of the other team they played with at the end of each round. As we employed a total stranger re-matching procedure between rounds, we do not share such a concern. Nonetheless, when running a two-sample non-parametric test on the session level, the treatment difference is marginally significant with $p = .087$, combined $N = 26$.

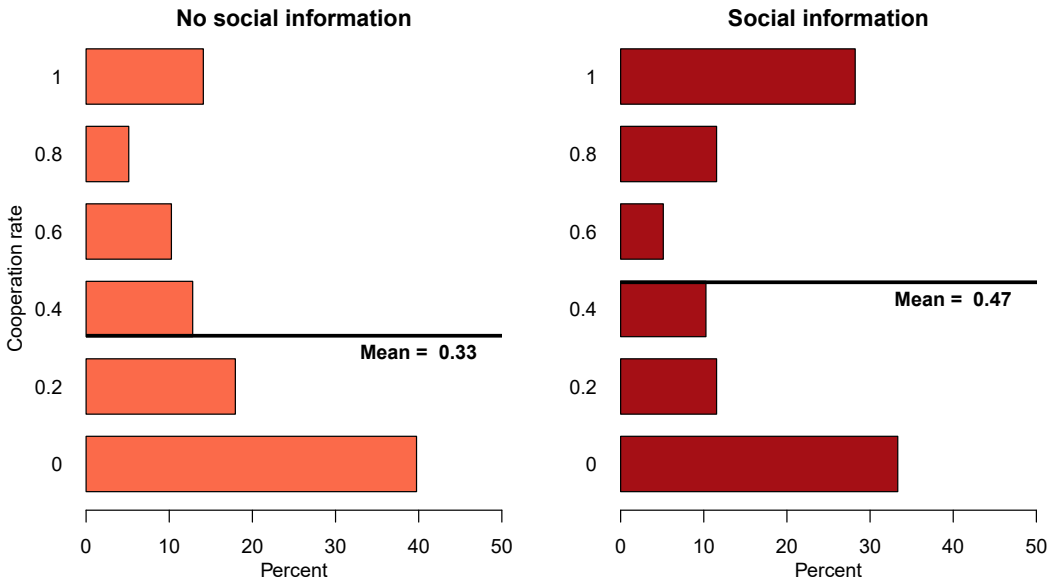


Fig. 3.4: Average team cooperation rate of majority decisions by social information treatment (combined $N = 156$)

teams with majority voting.

Due to the aggregation of individual votes, majority voting may mechanically lead to a low or high cooperation rate on the team level. To see this effect, suppose the individual propensity to vote for cooperation is p. The probability P that a three-person team cooperates under majority voting when members' votes are independent from each other is then given by $P = p^3 + 3p^2(1-p)$. Because $P < p$ for $p \in [0, 1/2)$, and $P > p$ for $p \in (1/2, 1]$, there is a difference between the individual propensity to vote for cooperation and the resulting cooperation rate of teams (unless $p = 1/2$).

Indeed, we find evidence for such a difference in the data. Without social information, 38.4% of all participants vote to cooperate, which leads to a team cooperation rate of around 30%. In contrast, 47.6% of the participants that are exposed to the social information message vote for cooperation, which leads to a team cooperation of just under 50%. The treatment difference for individual votes is 9.2 percentage points and hence smaller than the 13.8 percentage point treatment difference on the team level. Nevertheless, the treatment difference for individual votes is statistically significant with $p = .032$, combined $N = 468$.[16]

[16] For more information on the majority mechanic, see Appendix B-3.

3.6.2 The Effect of Social Information on Dictatorial Decisions

Figure 3.5 plots the distribution of cooperation rates for dictatorial decisions by social information treatment. Without social information, the most frequent strategy is zero cooperation across all rounds. With social information, the most frequent strategy is full cooperation. In both treatments, mixed strategies are played in less than 50% of all cases. We consequently observe a strong positive treatment effect of social information. The average cooperation rate is 16.6 percentage points higher with social information (54.2%) than without social information (37.6%), a significant increase in cooperation of 44%, $p = .008$, combined $N = 180$.

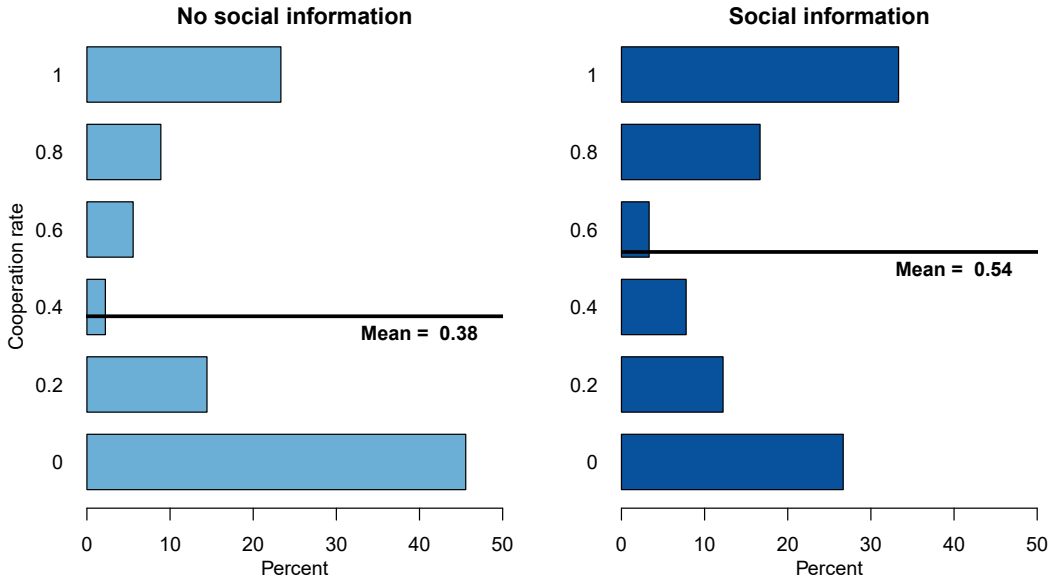


Fig. 3.5: Average cooperation rate of dictatorial choices by social information treatment (combined $N = 180$)

One can argue that the cooperation rate of each dictator is not an independent observation as team members can observe each others' decisions during the end of round feedback. We employ two strategies to alleviate these concerns. First, we average the three dictatorial decisions within each team, leaving us with a combined sample size of $N = 60$. In the corresponding non-parametric test, the treatment difference is significant with $p = .016$.[17] Second, we limit observations to the first

[17]The most conservative approach poses dependencies between teams in the same session as the dictator's behavior is also influenced by the observation of other teams' decisions. We do not share such concerns as a total stranger re-matching procedure of teams into new collectives between rounds and general anonymity during the decision making process rules out direct reciprocity motivations.

round as dictators have not yet observed their team members' choices when making their first decision. With social information, 57.8% of dictators cooperate in the first round while 43.3% cooperate without social information. The treatment difference of 14.5 percentage points is significant with $p = .073$, combined $N = 180$.

Our results on the treatment difference strongly suggest that the social information message is successful in changing team decisions when teams use a hierarchical structure. The result is supported by a significant coefficient for the social information treatment in a fractional probit regression. For the average marginal treatment effect, see column (2) in Table 3.2. Social information leads to an increase in team cooperation by, on average, 18.2 percentage points ($p = .038$). Hence, we accept Hypothesis 2: Social information increases cooperation by teams when decisions are made by a dictator.

In sum, we find that the social information treatment increases cooperation of teams in the prisoner's dilemma, irrespective of the imposed coordination mechanism. In column (3) of Table 3.2, we present average marginal effects of a fractional probit model that compares the dictatorial choice mechanism with the majority voting mechanism. That is, we regress the team cooperation rate, where we consider individual choices when decisions are made by dictators ($N = 180$) and majority voting outcomes otherwise ($N = 156$), on the social information treatment, an indicator for the dictatorial decision mechanism, and an interaction term.[18] The model shows robustness for the social information treatment effect in both coordination mechanisms. While the interaction term is insignificant ($p = .710$), the joint effect of social information and dictatorial decisions is significantly different from zero ($p = .023$). Moreover, we document no differences between majority voting and dictatorial decisions mechanism for team cooperation rate in the baseline (without social information).

3.6.3 How Experience Moderates the Effect of Social Information

The coordination mechanism is imposed in random treatment assignment such that some teams are able to apply their real world experience with the respective decision structure while other teams are put in (for them) unnatural positions to make a decision using an unfamiliar structure. The random assignment enables us to identify whether experience with the decision structure moderates the social information treatment on team cooperation. To isolate the moderating effect of experience from

[18] All models include a set of observable characteristics (see table notes). For completeness, we report all coefficients in Appendix Table B-4.

Table 3.2: Average marginal effects from fractional probit models on team cooperation rate for majority voting and dictatorial decisions

	Team Cooperation Rate			
	Maj. voting	Dict. decision	Pooled	
	(1)	(2)	(3)	(4)
Social info	0.149*	0.182**	0.142*	0.224**
	(0.079)	(0.088)	(0.073)	(0.104)
Dict. decision			0.056	0.122
			(0.084)	(0.093)
Social info × dict. decision			0.041	-0.108
			(0.112)	(0.127)
Experience				0.093
				(0.097)
Dict. decision × experience				-0.337***
				(0.121)
Social info × experience				-0.176
				(0.141)
Social info × dict. decision × experience				0.500***
				(0.186)
Controls	Yes	Yes	Yes	Yes
N	156	180	336	336

Notes: The table reports average marginal effects from team-level (for majority voting) and individual level (for dictatorial decisions) fractional probit regression models on the team cooperation rate. Robust standard errors are clustered at the session level (in parentheses). Margins are calculated at mean values of all covariates. Controls include age, age squared, crew size, an indicator whether the crew mainly targets dagaa, and a measure for comprehension. All controls variables are averaged among the three team members for teams that use majority voting. ***, **, and * indicate significance at the 1,5, and 10% level.

any inherent effects of the respective coordination mechanism, we pool majority and dictatorial decisions. We present average marginal effects of a fractional probit model in column (4) of Table 3.2. The model includes a three-way interaction between social information treatment, imposed coordination mechanism, and experience.

Conditional on no experience with an egalitarian decision structure, social information leads to a significant increase in team cooperation when decisions are made by majority voting ($p = .031$). The treatment effect for dictatorial decisions that are made without experience is slightly smaller, indicated by the negative interaction term (*social information* × *dict. choice*) and is jointly insignificant with $p = .113$.

To study whether real-life experience moderates the social information treatment effect, we shift our attention to the interaction terms with experience. For majority voting, experience with an egalitarian decision structure has a negative, yet insignificant impact on the social information treatment effect ($p = .214$). The coefficient

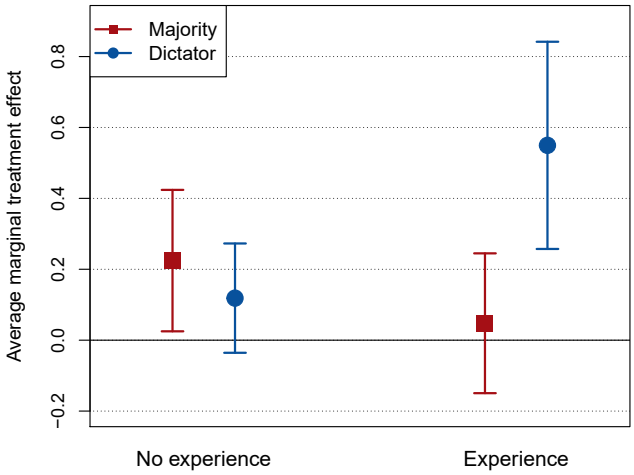


Fig. 3.6: Average marginal effects of social information treatment conditional on experience and imposed coordination mechanism. Whiskers indicate a 95% confidence interval.

point estimates suggest that social information increases team cooperation for teams that come from boats with an egalitarian structure and whose action in the experiment is determined by majority voting by 17.6 percentage points less than for teams whose action in the experiment is determined by majority but who come from boats with an hierarchical structure. The joint effect of social information and experience under majority voting is not significantly different from zero ($p = .618$).

For dictatorial decisions, the social information treatment effect is increased when participants are experienced. Cooperation rates with experience are substantially higher than without experience ($p = .001$). The joint effect indicates that conditional on having experience, social information leads to an increase in cooperation by 55 percentage points ($p < .001$). Also, we find that without social information, decision makers with hierarchical authority in real life cooperate significantly less than their inexperienced counterparts. That is, baseline cooperation is significantly lower ($p = .005$) for those that dictate team decisions in real life.

For an intuitive illustration of the moderating influence of experience, we plot marginal treatment effects in Figure 3.6. For majority voting, the effectiveness of social information is lower with experience. In contrast, for dictatorial decisions we find strong evidence that the effectiveness of social information is higher with experience. Participants who are inexperienced with making decisions on behalf of their team are only weakly affected by the social information treatment while captains and owners

that have the hierarchical authority in real life cooperate substantially more with social information than they do without social information.

The result that the treatment effect is conditional on having experience with the respective decision structure is neither due to the fact that captains cooperate more per se, nor due to the fact that fishermen from boats with egalitarian organization are less cooperative. In Appendix B-2, we present results from regression analyses on individual cooperation rates. For decisions taken under the dictatorial coordination mechanism (see column (1) in Table B-5), we find that being a captain or an owner of a boat with an egalitarian structure is not associated with more cooperative decisions ($p = .627$). Similarly, being crew on a boat with an hierarchical structure does not affect cooperation rates ($p = .874$). In contrast, captains from boats with hierarchical structure are less likely to cooperate than the baseline participant (regular crew members from boats with egalitarian structure) when no social information is provided ($p = .023$). Yet, they cooperate significantly more when exposed to the social information message ($p = .010$). For participants in the majority voting mechanism (column (2) in Table B-5) we see no effects of either being a captain/owner or coming from a boat with hierarchical structure. In other words, captains/owners with authority to make decisions in real life drive the treatment effect for dictatorial decisions.

3.6.4 Individual Level Outcomes

To learn more about the mechanisms by which social information may affect team decisions, we turn to individual level outcomes. The social information message is designed to change behavior through a change in the perceived social norm and it is indeed a common finding that those who receive information about cooperative behavior of others, expect them to cooperate and subsequently cooperate themselves. Hence, we first analyze how the provision of social information affects descriptive beliefs in the first round. We consider the descriptive beliefs elicited in the first round as they are not affected by own decisions or the behavior of others but only by the social information message.

The right plot of Figure 3.7 shows the distribution of descriptive beliefs by social information treatment and decision making mechanism in our sample. For dictatorial decisions, we find that descriptive beliefs are significantly affected by social information. Without social information, 31% of the participants believe that other teams will cooperate. With social information, this share increases by 26 percentage points to

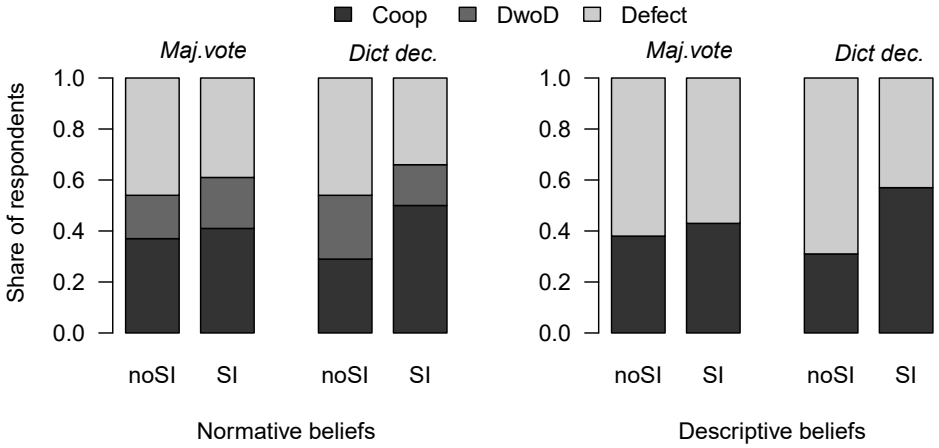


Fig. 3.7: Distribution of normative (left) and descriptive beliefs (right) by decision making mechanism and social information treatment. The dark areas show the share of participants indicating a belief of cooperation. The light areas indicate defection and the medium gray area for normative beliefs indicates the share of "do what others do" responses.

57%. The difference is significant with $p = .001$, combined $N = 180$. To our surprise, we cannot document a significant effect of social information on descriptive beliefs when decisions are determined by majority voting. Without social information, 38% of participants believe that other teams are cooperative. With social information, this share is 43%, an insignificant increase with $p = .300$, combined $N = 468$.

To support these findings, we conduct a mediation analysis and study whether the effect of social information on cooperation is mediated by descriptive beliefs. That is, if social information changes descriptive beliefs and these beliefs are a significant predictor of cooperation decisions, then the treatment runs through a change in the perceived social norm. Table 3.3 presents the average marginal effects of the mediation analysis.

For dictatorial decisions, we find clear evidence that the social information treatment effect is mediated by a change in descriptive beliefs. First, social information significantly increases the likelihood that dictators expect other teams to cooperate by almost thirty percentage points ($p < .001$), see column (1). Second, expecting the other team to cooperate predicts own cooperation ($p < .001$) such that the treatment effect is transmitted to a behavioral response, see column (3).

For majority decisions, we find that social information induces a small and insignificant change in descriptive beliefs ($p = .189$), see column (4). So while descriptive beliefs do drive behavior ($p < .001$, see column (6)), little of that predictive power originates from the social information treatment. In fact, the treatment effect that is

documented by our descriptive results and non-parametric tests (Figure 3.4) appears to be due to a small direct effect of the social information message on behavior (the marginal effect predicts a 8.4 percentage point increase, $p = 0.053$). Those that receive information about the cooperative behavior of other teams have a slightly increased likelihood to subsequently cooperate themselves but they do not necessarily expect other teams to cooperate.

Our mediation analysis sheds some light on the latent mechanism of the behavioral change that is induced by social information when decision are made by teams. While the weak and insignificant treatment effect on descriptive beliefs for majority voting is somewhat surprising, the substantial and clear transmission we observe for dictatorial decisions is in line with our hypothesis and a number of similar results in the literature that analyzes individual decision settings, see *e.g.*, Shang and Croson (2009) or Goeschl et al. (2018).

Table 3.3: Average marginal effects from a mediation analysis of the effect of social information on cooperation through descriptive beliefs (DB) and normative beliefs (NB) for both coordination mechanisms

	Dict. decision			Majority Voting		
	DB	*NB*	*Ind.Coop.Rate*	*DB*	*NB*	*Ind.Coop.Rate*
	(1)	(2)	(3)	(4)	(5)	(6)
Social information	0.275***		0.055	0.065		0.084*
	(0.083)		(0.085)	(0.050)		(0.042)
- NB (cooperate)		0.171**			0.062	
		(0.076)			(0.042)	
- NB (cond.coop.)		0.001			0.001	
		(0.009)			(0.002)	
- NB (defect)		-0.170**			-0.063	
		(0.076)			(0.042)	
DB			0.235***			0.573***
			(0.065)			(0.052)
NB (cooperate)			0.423***			0.187***
			(0.091)			(0.058)
NB (defect)			-0.158***			-0.158***
			(0.051)			(0.051)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	180	180	180	468	468	468

Notes: The table reports average marginal effects from individual level probit models (for descriptive beliefs), ordered probit models (for normative beliefs), and individual level fractional probit models (for cooperation rates). Robust standard errors are clustered at the session level (in parentheses). Margins are calculated at mean values of all covariates. Individual controls include age, age squared, crew size, an indicator whether the crew mainly targets dagaa, and a measure for comprehension. ***, **, and * indicate significance at the 1,5, and 10% level.

A key difference between the two coordination mechanisms is that participants cannot evade responsibility when they have to make the team decision by dictatorial choice. Those participants may take the social information as a signal of what should be done. In particular, if they themselves have a preferences for conformity with other teams or they believe that their team members would want their fairness or conformity preferences represented, social information would induce cooperation. Similarly, social information may reduce participant's fear of being exploited and thereby decrease the need to defect for in-group protection. Conversely, the possibility to evade responsibility under majority voting may mean that the information about the behavior of other teams is less relevant for own decisions. In an egalitarian decision structure, participants may care more about benefiting the in-group than conforming with the actions of other teams. Consequently, the social information may be disregarded and does not become a signal for what should be done.

We therefore take a closer look at participants' normative beliefs, *i.e.*, what participants think is the right thing to do. For a distribution of normative beliefs by decision making mechanism and social information treatment, see the left plot of Figure 3.7. Interestingly, we observe a treatment effect of social information on normative beliefs for individuals that make dictatorial decisions, but not for individuals under majority voting. While 29% of dictators hold the belief that cooperation is the morally right thing to do without social information, the share is 50% with social information. The likelihood to answer with the belief that indicates defection and conditional cooperation (*i.e.*, to "do what others do") decreases by ten and twelve percentage points, respectively. A chi-square test for univariate frequency distributions establishes that these differences are significant, $p < .000$ combined $N = 180$. For majority decisions, normative beliefs are only weakly affected, $p = .099$ combined $N = 468$. The share of cooperative beliefs is only increased by four percentage points from 37% without social information to 41% with social information. The likelihood to answer with the belief that indicates defection and conditional cooperation decreases by seven percentage points and increases by three percentage points, respectively.

To identify the role of participants' normative belief in inducing behavioral change, we examine whether the treatment effect of social information is additionally mediated by normative beliefs. We support our non-parametric finding and observe a significant effect of social information on normative beliefs when decisions are made by dictators, see column (2) in Table 3.3. With social information, participants are around 17 percentage points more likely to hold an unconditional preference for cooperation and 17 percentage points less likely to prefer defection than without social information.

The intervention does not affect the likelihood to hold a preference for conditional cooperation. Furthermore, we find evidence that participants' normative belief drives behavior. This indicates that the treatment effect is also mediated by a change in normative beliefs. Compared to those stating a preference for conditional cooperation, unconditionally cooperative dictators cooperate approximately 19 percentage points more and those that state a preference for defection cooperate around 16 percentage points less, see column (3) in Table 3.3. We also conduct a mediation analysis of normative beliefs with respect to majority voting, see columns (5) and (6) in Table 3.3. We find that normative beliefs are not responsive to the social information treatment and can therefore not transmit any treatment effect.

3.7 Discussion

Collective action problems such as climate change, corporate collusion, or community resource management are social dilemmas that need to be solved by teams. These social dilemmas cannot be addressed by formal regulations alone, but require a shift in the social norms that govern behavior. In this paper, we present an experiment to test whether social information increases cooperation when decisions are made by teams. Moreover, our study builds an important bridge between the experimental laboratory and the field, not only because our participants work together in teams in their daily lives as fishermen at Lake Victoria, but also because they make their resource extraction decisions in hierarchical as well as egalitarian structures that match the majority voting and dictatorial decision mechanisms that we impose in the experiment.

The existing literature has – by and large – documented the success of social information on individual behavior. We complement this literature by showing that a social information message can also change collective actions. Interestingly, we find that the social information message is more effective for teams whose action is determined by a dictatorial choice than for teams whose action is determined by majority voting. These results echo with the notion of a greater effectiveness of pro-social incentives when given to individuals (Gatiso et al., 2018). In cases where cooperation may increase social welfare, individuals should be preferred as decision makers (Charness and Sutter, 2012). We add nuance to this discussion by highlighting that the bias towards rational and self-interested behavior in teams can be overcome without dissolving teams as such but by increasing the responsibility for one team member.

Moreover, we contribute to the literature on authority and power in decision making (Fehr et al., 2013) and identify that these attributes may be influential drivers of

behavioral change in team decision making. We find that when team decisions are made by dictatorial choice, agents with real-life authority in hierarchically structured teams are particularly receptive to social information. While these participants are not more cooperative per se, they are particularly responsive to information about the behavior of other teams. Hence, we find that cooperative leadership may develop from paradigms of cooperation, a promising result given that pro-social leaders increase others' cooperation (Jack and Recalde, 2015; Kosfeld and Rustagi, 2015).

As with any experiment, there are limits to the external validity of its findings. Our design abstracts from several features of real-world team decision making. First of all, we allow for only very limited communication among team members. Irrespective of the decision making environment, team members can observe each others' choices, but they cannot freely express the reasons for their decisions. Similarly, while participants can express their disapproval with cooperating or defecting, they cannot publicly speak to convince others in the room that one action was preferable to another.[19]

A second difference with real life is that we impose repeated one-shot interactions with an anonymous perfect stranger matching, such that there are no spillover effects over time. Teams cannot build up a reputation and behavioral patterns of reciprocity or directed altruism cannot consolidate. Also, the absence of payoff linkages across rounds implies that there is no room for history-dependent strategies. Especially for issues of resource management, depleting the resource stock could be used as an effective threat to enforce cooperation, at least in theory (Polasky et al., 2005).

Third, we consider unitary teams. While this is a good approximation for many fisheries where crew members are paid in shares, free-riding incentives within the team and issues of self-selection into teams should be considered in other settings.[20] An analysis on how these intra-team incentives interact with a social information treatment would connect our experiment to the literature on team contests and multi-level public good games.

Obtaining a better understanding of how robustly a social information message

[19] In focus group discussions, fishermen conveyed the sentiment that disapproval has no significance if it is not backed by the entire community. That is, all disapproval signals that are not (almost) unanimous are disregarded. We find suggestive evidence that disapproval is not systematically used. In no treatment and at no iteration of the repeated game, one of the disapproval options was chosen by a meaningful majority of participants (*i.e.*, by at least 60%). How cultural differences may account for the difference in disapproval impacts compared to studies such as Masclet et al. (2003) or Dugar (2013) is beyond the scope of this paper.

[20] Also in our setting at Lake Victoria, there is variation in the payment structure: Fishing crews use different agreements for wage payment or the distribution of catch earnings. Conditions range from proportional catch earnings over fixed daily or monthly wages to more unclear payment structures where the daily revenue is in turn kept by the owner or the crew (Kateka, 2010, and own data).

can increase cooperation in a social dilemma between teams is highly topical. By utilizing naturally occurring teams in a field context, our experiment is a first step but more research is needed to learn about the mechanisms of behavioral change for joint decisions. Here, the experimental tool-kit for laboratory studies is somewhat limited as the simulation of genuine team membership in a student population is nontrivial. While techniques such as inducing minimal groups can serve as a starting point (Tajfel, 1982), researchers should look towards field settings to advance the discipline. For implementation in the field, social information interventions need to be designed carefully. In particular, the credibility of a social information message hinges on a trustworthy source and proper targeting. Well executed, these interventions can become a cost-effective and flexible solution to a social dilemma that operates at the root of behavioral change. As such, social norms are a promising candidate for informal governance that addresses collective action problems at the local level.

Bibliography

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.

Andreoni, J. and Sanchez, A. (2014). Do beliefs justify actions or do actions justify beliefs? an experiment on stated beliefs, revealed beliefs, and social-image manipulation. Technical report, National Bureau of Economic Research.

Atanasov, P. and Kunreuther, H. (2016). Cautious defection: Group representatives cooperate and risk less than individuals. *Journal of Behavioral Decision Making*, 29(4):372–380.

Aura, C. M., Nyamweya, C. S., Njiru, J. M., Musa, S., Ogari, Z., May, L., and Wakwabi, E. (2019). Exploring the demarcation requirements of fish breeding and nursery sites to balance the exploitation, management and conservation needs of lake victoria ecosystem. *Fisheries Management and Ecology*, 26(5):451–459.

Baker, R. J., Laury, S. K., and Williams, A. W. (2008). Comparing small-group and individual behavior in lottery-choice experiments. *Southern Economic Journal*, pages 367–382.

Bergquist, M., Nilsson, A., and Schultz, W. P. (2019). A meta-analysis of field-experiments using social norms to promote pro-environmental behaviors. *Global Environmental Change*, 59:101941.

Bicchieri, C. (2017). *Norms in the Wild – How to Diagnose, Measure, and Change Social Norms*. Oxford University Press.

Bicchieri, C. and Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public choice*, pages 1–22.

Bicchieri, C., Dimant, E., and Xiao, E. (2020). Deviant or Wrong? The Effects of Norm Information on the Efficacy of Punishment. *The Effects of Norm Information on the Efficacy of Punishment (December 1, 2018)*.

Bicchieri, C. and Xiao, E. (2009). Do the Right Thing: But Only if Others Do So. *Journal of Behavioral Decision Making*, 22(2):191–208.

Blackwell, C. and McKee, M. (2003). Only for my own neighborhood?: Preferences and voluntary provision of local and global public goods. *Journal of Economic Behavior & Organization*, 52(1):115 – 131.

Bohnet, I. and Frey, B. S. (1999). Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review*, 89(1):335–339.

Bornstein, G., Kugler, T., and Ziegelmeyer, A. (2004). Individual and group decisions in the centipede game: Are groups more "rational" players? *Journal of Experimental Social Psychology*, 40(5):599–605.

Bornstein, G. and Yaniv, I. (1998). Individual and group behavior in the ultimatum game: are groups more "rational" players? *Experimental Economics*, 1(1):101–108.

Buchan, N. R., Grimalda, G., Wilson, R., Brewer, M., Fatas, E., and Foddy, M. (2009). Globalization and human cooperation. *Proceedings of the National Academy of Sciences*, 106(11):4138–4142.

Cason, T. N. and Mui, V.-L. (1997). A laboratory study of group polarisation in the team dictator game. *The Economic Journal*, 107(444):1465–1483.

Charness, G. (2000). Responsibility and effort in an experimental labor market. *Journal of Economic Behavior & Organization*, 42(3):375–384.

Charness, G. and Chen, Y. (2020). Social identity, group behavior, and teams. *Annual Review of Economics*, 12.

Charness, G. and Holder, P. (2019). Charity in the laboratory: Matching, competition, and group identity. *Management Science*, 65(3):1398–1407.

Charness, G. and Jackson, M. O. (2009). The role of responsibility in strategic risk-taking. *Journal of Economic Behavior & Organization*, 69(3):241–247.

Charness, G. and Sutter, M. (2012). Groups make better self-interested decisions. *Journal of Economic Perspectives*, 26(3):157–76.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree - an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Coase, R. H. (1937). The nature of the firm. *Economica*, 4(16):386–405.

Costa, D. L. and Kahn, M. E. (2013). Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3):680–702.

Cowx, I. G. and Ogutu-Owhayo, R. (2019). Towards sustainable fisheries and aquaculture management in the african great lakes. *Fisheries Management and Ecology*, 26(5):397–405.

Croson, R., Handy, F., and Shang, J. (2009). Keeping up with the Joneses: The Relationship of Perceived Descriptive Social Norms, Social Information, and Charitable Giving. *Nonprofit Management and Leadership*, 19(4):467–489.

Denant-Boemont, L., Diecidue, E., and L'haridon, O. (2017). Patience and time consistency in collective decisions. *Experimental Economics*, 20(1):181–208.

Diekert, F., Eymess, T., Luomba, J., and Waichman, I. (2021). The Creation of Social Norms under Weak Institutions. Heidelberg University, Department of Economics Discussion Paper No. 684.

Dufwenberg, M. and Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior*, 30(2):163–182.

Dugar, S. (2013). Non-monetary incentives and opportunistic behavior: Evidence from a laboratory public good game. *Economic Inquiry*, 51(2):1374–1388.

Etiegni, C., Irvine, K., and Kooy, M. (2017). Playing by whose rules? community norms and fisheries rules in selected beaches within lake victoria (kenya) co-management. *Environment, Development and Sustainability*, 19(4):1557–1575.

Eymess, T. (2021). Dissecting social norms of cooperation: A conditional process analysis. Manuscript, Heidelberg University.

Farrow, K., Grolleau, G., and Ibanez, L. (2017). Social norms and pro-environmental behavior: a review of the evidence. *Ecological Economics*, 140:1–13.

Fehr, E., Herz, H., and Wilkening, T. (2013). The lure of authority: Motivation and incentive effects of power. *American Economic Review*, 103(4):1325–59.

Fehr, E. and Schurtenberger, I. (2018). The dynamics of norm formation and norm decay. *Working paper series/Department of Economics*, (278).

Feri, F., Irlenbusch, B., and Sutter, M. (2010). Efficiency gains from team-based coordination – large-scale experimental evidence. *American Economic Review*, 100(4):1892–1912.

Ferraro, P. J., Miranda, J. J., and Price, M. K. (2011). The persistence of treatment effects with norm-based policy instruments: evidence from a randomized environmental policy experiment. *American Economic Review*, 101(3):318–22.

Gallier, C., Goeschl, T., Kesternich, M., Lohse, J., Reif, C., and Römer, D. (2019). Leveling up? an inter-neighborhood experiment on parochialism and the efficiency of multi-level public goods provision. *Journal of Economic Behavior & Organization*, 164:500 – 517.

Gatiso, T. T., Vollan, B., Vimal, R., and Kühl, H. S. (2018). If possible, incentivize individuals not groups: evidence from lab-in-the-field experiments on forest conservation in rural uganda. *Conservation Letters*, 11(1):e12387.

Gichuru, N., Nyamweya, C., Owili, M., Mboya, D., and Wanyama, R. (2019). Poor management of lake victoria fisheries (kenya); a threat to sustainable fish supplies. *Nature & Faune*, page 38.

Goeschl, T., Kettner, S. E., Lohse, J., and Schwieren, C. (2018). From social information to social norms: Evidence from two experiments on donation behaviour. *Games*, 9(4).

Hallsworth, M., List, J. A., Metcalfe, R. D., and Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148:14 – 31.

Humphrey, S. J. and Renner, E. (2011). The social costs of responsibility. Technical report, CeDEx discussion paper series.

Insko, C. A., Hoyle, R. H., Pinkley, R. L., Hong, G.-Y., Slim, R. M., Dalton, B., Lin, Y.-H. W., Ruffin, P. P., Dardis, G. J., Bernthal, P. R., et al. (1988). Individual-group discontinuity: The role of a consensus rule. *Journal of Experimental Social Psychology*, 24(6):505–519.

Insko, C. A., Schopler, J., Hoyle, R. H., Dardis, G. J., and Graetz, K. A. (1990). Individual-group discontinuity as a function of fear and greed. *Journal of Personality and Social Psychology*, 58(1):68.

Irvine, K., Etiegni, C., and Weyl, O. (2019). Prognosis for long-term sustainable fisheries in the african great lakes. *Fisheries Management and Ecology*, 26(5):413–425.

Jack, B. K. and Recalde, M. P. (2015). Leadership and the voluntary provision of public goods: Field evidence from bolivia. *Journal of Public Economics*, 122:80–93.

Kagel, J. H. and McGee, P. (2016). Team versus individual play in finitely repeated prisoner dilemma games. *American Economic Journal: Microeconomics*, 8(2):253–76.

Kateka, A. G. (2010). *Co-management challenges in the Lake Victoria fisheries: a context approach*. PhD thesis, Acta Universitatis Stockholmiensis.

Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.

Kocher, M. G., Praxmarer, M., and Sutter, M. (2020). Team decision-making. *Handbook of Labor, Human Resources and Population Economics*, pages 1–25.

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

Kolding, J., Medard, M., Mkumbo, O., and van Zwieten, P. (2014). Status, trends and management of the lake victoria fisheries. *Inland fisheries evolution and managementâĂŤcase studies from four continents. FAO Fisheries and Aquaculture Technical Paper*, 579.

Kolding, J., van Zwieten, P., Marttin, F., Smith, S. F., and Poulain, F. (2019). *Freshwater Small Pelagic Fish and Their Fisheries in Major African Lakes and Reservoirs in Relation to Food Security and Nutrition*. Food and Agriculture Organization of the United Nations.

Kolding, J., van Zwieten, P. A., and Mosepele, K. (2016). Where there is water, there is fish. small-scale inland fisheries in africa: Dynamics and importance. *A History of Water: Series III, Volume 3: Water and Food*, page 439.

Kosfeld, M. and Rustagi, D. (2015). Leader punishment and cooperation in groups: Experimental field evidence from commons management in ethiopia. *American Economic Review*, 105(2):747–83.

Kugler, T., Bornstein, G., Kocher, M. G., and Sutter, M. (2007). Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy. *Journal of Economic psychology*, 28(6):646–657.

Kugler, T., Kausel, E. E., and Kocher, M. G. (2012). Are groups more rational than individuals? a review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(4):471–482.

Lopez, M. C., Murphy, J., M., S. J., and Stranlund, J. K. (2012). Comparinng the effectiveness of regulation and pro-social emotions to enhance cooperation: Experimental evidence from fishing communities in colombia. *Economic Inquiry*, 50(1):131–142.

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

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Masclet, D., Colombier, N., Denant-Boemont, L., and Loheac, Y. (2009). Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers. *Journal of Economic Behavior & Organization*, 70(3):470–484.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93(1):366–380.

Michaeli, M. and Spiro, D. (2017). From peer pressure to biased norms. *American Economic Journal: Microeconomics*, 9(1):152–216.

Miller, D. T. and Prentice, D. A. (2016). Changing norms to change behavior. *Annual Review of Psychology*, 67:339–361.

Mkumbo, O. C. and Marshall, B. E. (2015). The Nile perch fishery of Lake Victoria: Current status and management challenges. *Fisheries Management and Ecology*, 22(1):56–63.

Nunan, F., Cepić, D., Yongo, E., Salehe, M., Mbilingi, B., Odongkara, K., Onyango, P., Mlahagwa, E., and Owili, M. (2018). Compliance, corruption and co-management: how corruption fuels illegalities and undermines the legitimacy of fisheries co-management. *International Journal of the Commons*, 12(2).

Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger, W. N., Arrow, K. J., Barrett, S., Carpenter, S., Chapin, F. S., Crépin, A.-S., Daily, G., Ehrlich, P., Folke, C., Jager, W., Kautsky, N., Levin, S. A., Madsen, O. J., Polasky, S., Scheffer, M., Walker, B., Weber, E. U., Wilen, J., Xepapadeas, A., and de Zeeuw, A. (2016). Social norms as solutions. *Science*, 354(6308):42–43.

Ostrom, E. (2008). The challenge of common-pool resources. *Environment: Science and Policy for Sustainable Development*, 50(4):8–21.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–632.

Polasky, S., Tarui, N., Ellis, G., and Mason, C. (2005). Cooperation in the Commons. *Economic Theory*, 29(1):71–88.

Schopler, J. and Insko, C. A. (1992). The discontinuity effect in interpersonal and intergroup relations: Generality and mediation. *European Review of Social Psychology*, 3(1):121–151.

Schopler, J., Insko, C. A., Drigotas, S. M., Wieselquist, J., Pemberton, M. B., and Cox, C. (1995). The role of identifiability in the reduction of interindividual-intergroup discontinuity. *Journal of Experimental Social Psychology*, 31(6):553–574.

Selten, R. (1965). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. Seminar für Mathemat. Wirtschaftsforschung u. Ökonometrie.

Shang, J. and Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal*, 119(540):1422–1439.

Shapiro, J. (2010). Discounting for you, me and we. Technical report, Working paper.

Sheremeta, R. M. (2018). Behavior in group contests: A review of experimental research. *Journal of Economic Surveys*, 32(3):683–704.

Shupp, R. S. and Williams, A. W. (2008). Risk preference differentials of small groups and individuals. *The Economic Journal*, 118(525):258–283.

Song, F. (2009). Intergroup trust and reciprocity in strategic interactions: Effects of group decision-making mechanisms. *Organizational Behavior and Human Decision Processes*, 108(1):164–173.

Stoner, J. A. F. (1961). *A comparison of individual and group decisions involving risk*. PhD thesis, Massachusetts Institute of Technology.

Sugden, R. (2000). The motivating power of expectations. In *Rationality, Rules, and Structure*, pages 103–129. Springer.

Taabu-Munyaho, A., Kayanda, R. J., Everson, I., Grabowski, T. B., and Marteinsdóttir, G. (2013). Distribution and exploitation of nile perch lates niloticus in relation to stratification in lake victoria, east africa. *Journal of Great Lakes Research*, 39(3):466–475.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual review of psychology*, 33(1):1–39.

Wildschut, T., Insko, C. A., and Pinter, B. (2007). Interindividual-intergroup discontinuity as a joint function of acting as a group and interacting with a group. *European Journal of Social Psychology*, 37(2):390–399.

Williamson, O. E. (1967). Hierarchical control and optimum firm size. *Journal of Political Economy*, 75(2):123–138.

Appendix

B-1 Sample Characteristics

Our sample consists of $N = 648$ fishermen from landing sites at Lake Victoria, Tanzania. In Table B-1, we present participant characteristics for both coordination mechanisms with mean comparison tests across social information treatments. Participants are on average in their thirties and we observe only a total of four female participants indicating that the Lake Victoria fisheries are dominated by men. The fishermen in our sample rely heavily on their income from fishing as about 70% of household earnings come from the daily fishing activities. On a range of self-reported preferences (Likert scale from 0 to 5) we observe that fishermen are somewhat risk averse (mean value around 2), and state moderate preferences for altruism (mean value around 3.8). The majority of participants have strong social image concerns (mean value around 3.1). All preferences are bi-modally distributed with peaks at 0 and 5, respectively. 21% of our participants have previously participated in an economic experiment.

Participant characteristics are, for the most part, balanced across social information treatments. The age difference for dictators is not concerning as it is negligible in absolute size.

Table B-1: Participant characteristics with mean comparison tests across social information treatments

	Dictator			Majority		
	SI	no SI	p-val	SI	no SI	p-val
Age	35.28	38.50	.028	35.63	35.83	.827
Female	0	0	-	0.013	0.004	.316
HH income (% fish)	62.00	65.29	.380	70.15	70.85	.757
Risk pref	2.14	1.93	.514	2.05	1.79	.188
Altruism pref	3.71	3.64	.822	3.94	3.87	.704
Social image concern	3.09	3.03	.860	3.19	3.18	.983
Prior participation	0.244	0.250	.930	0.228	0.177	.166
N	90	90		234	234	

Notes: Comparison of individual participant characteristics with mean comparison tests between social information (SI) and no social information (no SI) treatments for both dictatorial and majority coordination mechanism. All displayed test statistics are mean-comparison t-tests with non-adjusted p-values.

Furthermore, we compare sample characteristics across social information treatments for the crews in our data, see Table B-2. In case members of the same crew in our experiment give conflicting answers to any questions that should be answered equivalently for all crew members, we consider the modal response within the team. The fishing at Lake Victoria is dominated by crews of three or four members, and around 30% of crews in our sample mainly target the Dagaa species (which is usually targeted by somewhat larger crews). After adjusting for inconsistent responses, we observe that about 40% of all crews use a hierarchical organization structure where either the captain or boat owner dictates the location for fishing.

Lastly, we report sample characteristics with mean comparison tests for both coordination mechanisms across experience. We find that the sample is mostly balanced across experience with the exception of the household's income share from fishing for dictators and risk preferences as well as prior participation in experiments for majority decisions. All participant

Table B-2: Boat characteristics with mean comparison tests across social information treatments

	Dictator			Majority		
	SI	noSI	p-val	SI	noSI	p-val
Average age	35.28	38.50	.061	35.62	38.83	.859
Crew size	3.20	3.41	.232	3.54	3.88	.030
Main gear (Dagaa)	0.100	0.167	.456	0.282	0.423	.066
Hierarchic organization	0.433	0.300	.292	0.423	0.385	.627
N	30	30		78	78	

Notes: Comparison of boat level characteristics with mean comparison tests across social information treatments between boats that use a participatory approach to decision making and boats that use hierarchic decisions. All displayed test statistics are mean-comparison t-tests with non-adjusted p-values.

characteristics are controlled for in our regression specifications.

Table B-3: Participant characteristics with mean comparison tests across experience

	Dictator			Majority		
	Exp.	no Exp.	p-val	Exp.	no Exp.	p-val
Age	34.14	37.41	.102	35.52	35.92	.667
Female	0	0	-	0.005	0.012	.368
HH income (% fish)	71.62	62.11	.061	71.01	69.97	.617
Risk pref	1.97	2.05	.843	1.73	2.09	.066
Altruism pref	3.75	3.66	.811	3.89	3.91	.893
Social image concern	3.62	3.95	.119	3.13	3.23	.653
Prior participation	0.286	0.240	.608	0.239	0.169	.064
N	29	151		222	246	

Notes: Comparison of individual participant characteristics with mean comparison tests between experienced (Exp.) and inexperienced (no Exp.) participants for both dictatorial and majority coordination mechanism. All displayed test statistics are mean-comparison t-tests with non-adjusted p-values.

B-2 Regression Analysis

In Table B-4, we show the regression coefficients from the fractional probit models on team cooperation rates whose average marginal effects we report in Table 3.2. For completeness, we also report coefficients for all control variables and an additional set of self-reported preferences (including altruism, risk, and social image concerns). All control variables are averaged for teams that use majority decisions. We observe that team cooperation slightly increases with age. Also, those crews that target dagaa cooperate more while social image concerns generally correlate with less cooperative responses under the dictatorial decision mechanism, but not under majority voting.

Table B-4: Regression results from fractional probit models on team cooperation rate for majority voting and dictatorial decisions

	Team Cooperation Rate			
	Dict. dec.	Maj. voting	Pooled	
Social info	0.478**	0.368*	0.375*	0.597**
	(0.219)	(0.206)	(0.197)	(0.280)
Dict. decision			0.161	0.335
			(0.237)	(0.278)
Social info × dict. dec.			0.101	-0.297
			(0.294)	(0.344)
Experience				0.262
				(0.280)
Social info × experience				-0.473
				(0.381)
Dict. dec. × experience				-0.979**
				(0.415)
Social info × dict. dec. × experience				1.716***
				(0.583)
Age	0.118***	0.085	0.103***	0.115***
	(0.045)	(0.077)	(0.040)	(0.044)
Age^2	-0.001**	-0.001	-0.001**	-0.001**
	(0.001)	(0.001)	(0.001)	(0.001)
Crew size	-0.050	-0.060	-0.058	-0.062
	(0.122)	(0.133)	(0.094)	(0.091)
Main gear (Dagaa)	0.307	0.449*	0.347**	0.346**
	(0.235)	(0.248)	(0.171)	(0.169)
Comprehension	-0.501*	-0.431	-0.459*	-0.448*
	(0.284)	(0.444)	(0.240)	(0.240)
Altruism preference	0.074*	0.179**	0.093**	0.084**
	(0.044)	(0.083)	(0.038)	(0.039)
Risk preference	-0.030	0.092	0.000	0.012
	(0.031)	(0.077)	(0.031)	(0.033)
Social image concern	-0.075***	-0.038	-0.069***	-0.078***
	(0.028)	(0.072)	(0.027)	(0.026)
N	180	156	336	336

Notes: The table reports regression coefficient from team level fractional probit regression models on cooperation rate. Robust standard errors are clustered at the session level (in parentheses). All observable characteristics are averaged among the three team members for teams that use majority voting. ***, **, and * indicate significance at the 1,5, and 10% level.

In Table B-5 we furthermore regression results from fractional probit models on individual

cooperation rates for majority voting an dictatorial choice in support of column (4) in Table 3.2. In particular, we dissect our experience measure in (i) whether the individual states to be an captain or owner (CO), (ii) whether the crew uses a hierarchical organizational structure (HO), and (iii) whether the participant states to be the decision making captain or owner in the hierarchically organized crew (CO-HO), *i.e.*, whether the participant reports to have decision making authority. Our results are robust to the interpretation of our heterogeneous treatment effect with respect to experience. That is, for dictatorial decisions, we find that those with experience (the CO-HO participants) cooperate *less* without social information and substantially *more* with social information.

Table B-5: Regression results from fractional probit models on individual cooperation rate for majority voting and dictatorial decisions

	Individual Cooperation Rate		
	Dict. dec.	Maj. voting	Pooled
	(1)	(2)	(3)
Social info (SI)	0.406	0.157	0.170
	(0.282)	(0.162)	(0.162)
Role = captain/owner (CO)	-0.099	-0.082	-0.068
	(0.204)	(0.199)	(0.202)
SI × CO	-0.207	-0.059	-0.072
	(0.383)	(0.245)	(0.244)
Boat with hierarchical organization (HO)	-0.051	-0.148	-0.155
	(0.321)	(0.229)	(0.229)
SI × HO	-0.128	0.447	0.426
	(0.449)	(0.282)	(0.280)
CO on HO boat (CO-HO)	-0.731**	0.331	0.341
	(0.321)	(0.244)	(0.237)
SI × CO-HO	1.534**	-0.229	-0.245
	(0.596)	(0.288)	(0.284)
Dict. dec.			0.091
			(0.230)
SI × dict. dec.			0.236
			(0.305)
Dict. dec.e × CO			-0.077
			(0.323)
SI × dict. dec. × CO			-0.126
			(0.464)
Dict. dec. × HO			0.192
			(0.389)
SI × dict. dec. × HO			-0.635
			(0.504)
Dict. dec. × CO-HO			-1.039***
			(0.350)
SI × dict. dec. × CO-HO			1.683***
			(0.597)
Individual controls	Yes	Yes	Yes
N	180	468	648

Notes: The table reports coefficients from individual level fractional probit regression models on cooperation rate. Robust standard errors are clustered at the session level (in parentheses). Margins are calculated at mean values of all covariates. Controls include age, age squared, crew size, an indicator whether the crew mainly targets dagaa, self-reported altruism, risk preferences and social image concerns, as well as a measure for comprehension. ***, **, and * indicate significance at the 1,5, and 10% level.

B-3 Majority Voting Mechanic

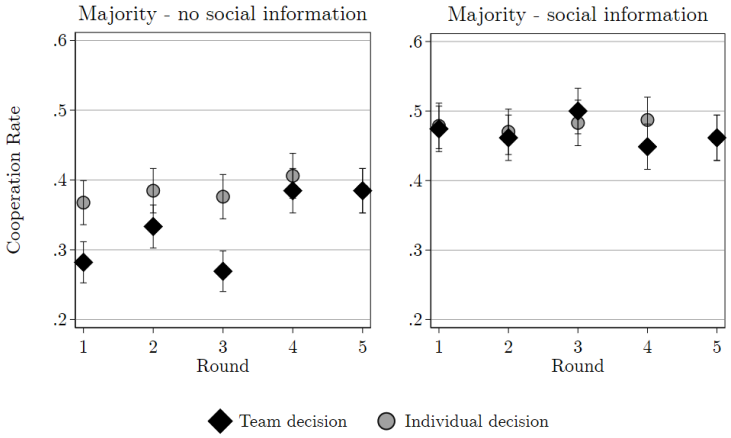


Fig. B-1: Plots of average individual (circles, combined $N = 468$) and average team cooperation rates (diamonds, combined $N = 156$) for the majority decisions without social information (left) and with social information (right) over all five rounds

Figure B-1 illustrates the majority mechanic. As the individual propensity to vote for cooperation in the social information treatment is close to 50%, an average team cooperation rate of close to 50% could, in principle, both be due to the fact that the difference between individual and team cooperation rates is small for values around 50% (see above), or due to the fact that the social information induces a stronger correlation between individual votes. To distinguish these two causes, we compare the multiple correlation coefficient of the observed teams to the correlation coefficient of one thousand hypothetical teams where we randomly grouped three participants. We find that 36% of all hypothetical teams under social information make cooperation decisions that are more correlated than actual teams, see Figure B-2. For the no social information treatment, the corresponding share of hypothetical teams whose votes are more correlated than for the actual teams is 65%, see Figure B-3. Hence, it is not the case that social information increases the correlation among team members' votes.

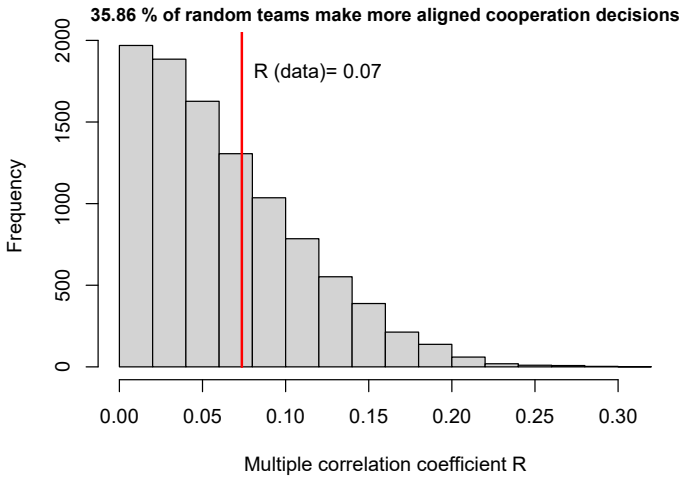


Fig. B-2: Comparison of observed team alignment to the alignment of randomly matched teams – no social information

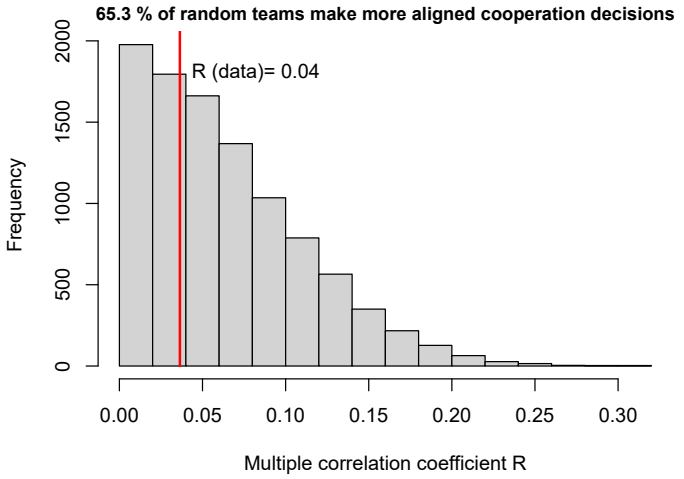


Fig. B-3: Comparison of observed team alignment to the alignment of randomly matched teams – social information

Chapter 4

Captains of Change: The Effect of Social Information on Resource Users in Leadership Positions

Joint work with Philipp Händel.

Abstract: If interventions aim to induce a pro-social change in natural resource use, who should they target? To understand whether resource users in leadership positions can facilitate collective change, we study behavior from two experiments on the effect of social information with fishermen from the Lake Victoria region in Tanzania. We find that vessel owners and to a lesser extent also captains are receptive to a social information message that tries to induce contributions to a public good. In contrast, the intervention fails with regular crew members. We show that the result is not explained by the fact that resource users in leadership positions are more strongly affected by social information per se but rather that the effect depends on the behavior in question and the recipient of the message. Our study has an important policy implication: to activate pro-social behavior that facilitates sustainable resource use, policy makers may target resource users in leadership positions as "captains of change".

Keywords: social information; natural resource management; leadership

4.1 Introduction

Strong leadership is an important social determinant of successful resource management (Gutiérrez et al., 2011; Kosfeld and Rustagi, 2015). Especially in the small-scale fisheries of developing countries, the pairing of co-management and local leadership through key community figures is essential to facilitate pro-social collective action (Bodin and Crona, 2008; Ostrom, 2008; Barnes-Mauthe et al., 2014). The idea is to enable a form of self-governance by involving communities and their local leaders in the process of defining regulations and enforcing compliance in local arenas (Ostrom, 1990). Yet, what is designed as a bottom-up approach to sustainable resource management is often perceived as an illegitimate attempt to retain top-down control (Nielsen et al., 2004; Etiegni et al., 2020). Eventually, corruption and a lack of trust in the devolution of governmental undermine effective co-management and erode compliance (Nunan et al., 2018).

In search of different approaches to fisheries management, policy makers may engage with resource users that fill inherent leadership positions. Critical production decisions such as the location for fishing, gear use, and overall harvesting efforts are made by fishermen with authority over the fishing crew (captains) or those that employ the crew and own the means of production (vessel and gear owners). In their role as captains and owners, these fishermen directly determine both resource exploitation and conservation efforts (Thorlindsson, 1988; Vázquez-Rowe and Tyedmers, 2013) and form an important bedrock of leadership in fishing communities. Should interventions that aim to induce a collective change in resource use target them as "captains of change"?

Cooperative resource management can be facilitated by activating social norms via a social information intervention (Farrow et al., 2017; Bergquist et al., 2019). In particular, Diekert et al. (2021) find that the combination of social information and a social sanctioning institution successfully creates a social norm of cooperation in a prisoner's dilemma experiment with fishermen at Lake Victoria. Diekert and Eymess (2021) revisit the same communities in Tanzania and show that social information is similarly effective when cooperation decisions are made by teams or by a random team leader. In their exploratory analysis, they describe an interesting finding. The effect of social information for decisions by a team leader is driven by those participants that have leadership experience, *i.e.*, by owners and captains. Their finding suggests that social information is more effective with resource users in leadership positions.

We put the finding by Diekert and Eymess (2021) to the test and examine whether

resource users in leadership positions are indeed more strongly affected by a social information intervention. For this purpose, we use two other experimental data sets with fishermen from Lake Victoria. First, we re-analyze the social information intervention on individual contributions to a public good studied in Diekert et al. (2021) and test whether their successful treatment is also driven by resource users in leadership positions. Second, we examine a social information intervention on investment decisions under risk studied in Dannenberg et al. (2021). Again, we test for a heterogeneous treatment effect with respect to role in the fisheries. Our analysis therefore provides insights into (i) whether the finding in Diekert and Eymess (2021) replicates for other social information interventions, and (ii) whether the effect of social information on resource users in leadership positions depends on the behavioral context of the intervention.

Our study adds to an understanding of leadership in natural resource management. Crona and Bodin (2010) argue that the opinion leadership of vessel and gear owners may act as a barrier to collective action. Due to their high investments into the fishery, they are reluctant to favor a decrease in resource use that would diminish their short-term returns (Clark, 2006). Their resistance shapes the opinion of others in the community and thereby stymies desirable collective action. In general, local leaders are key to transform public perceptions towards sustainable resource. Both through leading by example (Jack and Recalde, 2015) and through efficient enforcement (Kosfeld and Rustagi, 2015), pro-social leaders may crowd-in cooperation by others. Yet, leaders are not homogeneous with respect to the degree of leadership they offer and how prone they are to behavioral change. Some may be considered as so-called "big men", prestige-based leaders in small-scale communities who gain a following through recognition of their skill, knowledge, and wealth accumulation. They often behave pro-socially and are notably generous (Henrich et al., 2015). Others may distrust co-management structures and reject any form of exogenous incentives that resemble regulatory instruments (Cowx et al., 2003; Nunan et al., 2018). More empirical data on the individual level is necessary to conclude which resource users act as facilitators and barriers of collective change.

Our main result confirms the exploratory finding by Diekert and Eymess (2021): resource users in leadership positions are exceptionally receptive to a social information message that tries to induce a change in contributions to a public good. The effect is especially pronounced for fishermen in the role of vessel and gear owners. In contrast, the intervention fails with regular crew members. The result is not explained by the fact that those fishermen who work as captains and owners are more strongly

affected by social information per se. When the treatment emphasizes risky investment behavior, crew members and owners are responsive to the treatment while captains are affected to a lesser extent. Our results have important policy implications. To activate pro-social norms that facilitate sustainable resource governance, policy makers should target resource users in leadership positions as facilitators of collective change.

4.2 Materials and Methods

4.2.1 The Lake Victoria Fisheries in Tanzania

The small-scale, artisanal fisheries at Lake Victoria in Tanzania are part of a socio-ecological resource system that supports the livelihood and food security for millions of people in East Africa (Cowx and Ogutu-Owhayo, 2019; Irvine et al., 2019). The fisheries are open-access and governed through a co-management structure, the so-called beach management units (BMU). BMUs are elected community representatives who are tasked to help monitor and enforce compliance with official fishing regulations at the local level. Yet, this form of leadership through elected representation has not been successful in preventing frequent violations and unsustainable resource use (Eggert and Lokina, 2010).

Reasons for the inefficiency of BMUs are manifold. Co-management is plagued by issues of corruption (Nunan et al., 2018) and a perceived illegitimacy of regulations (Cepić and Nunan, 2017). Also, kinship ties in the traditionally close-knit communities prevent elected leaders from punishing non-compliant behavior as many violators are relatives or neighbors (Etiegni et al., 2017). A different perspective on inefficiencies in the co-management structure is provided by Bodin and Crona (2008). In a network analysis of a Kenyan fishing community, the authors find that the beach chairman is not among the most influential individuals, *i.e.*, those turned to for social support or knowledge sharing.[1] If resource users do not recognize the authority of BMU officials (Etiegni et al., 2020), their leadership is compromised.

Yet, some form of leadership is essential to good resource governance (Gutiérrez et al., 2011). In general, leadership may take the form of a leader that acts first, and others following suit (Arbak and Villeval, 2013; Cappelen et al., 2016), a leader that has authority over others (Song, 2009), or a leader that moderates discussions (Gatiso et al., 2018). Also, a leader may be the person that coordinates with additional

[1] Kenya runs a co-management structure comparable to the BMU structure in Tanzania. Here, the beach chairman heads an elected committee of fisherman in a fishing village (Bodin and Crona, 2008).

stakeholders such as (non-)governmental organizations to ensure community benefits (Krishna, 2002; Bodin and Crona, 2008). In the small-scale fishing communities at Lake Victoria, some of these responsibilities are taken up by resource users that fill inherent leadership. In their roles as captains and owners, these fishermen exercise authority and power in decision making on a daily basis. If well-designed policies can successfully induce a norm of cooperative behavior with these leaders, the positive impacts on resource management may spread throughout communities.

4.2.2 Data

We study the effect of social information with fishermen from the Lake Victoria region in Tanzania in two different behavioral contexts. First, we use a data set on contributions to a public good in a three-person prisoner's dilemma ($N = 287$) that was collected in 2018 to study the creation of social norms (Diekert et al., 2021). The prisoner's dilemma anonymously matches three players who each play with a binary choice set. Participants can either contribute their full endowment or nothing and repeat the game for a total of seven decisions under a total-stranger re-matching protocol. The marginal per capita return is 0.5 such that a zero contribution is the payoff-dominant strategy while full contribution is the social optimum. Before contribution decisions in the first round, participants either receive a low or high social information treatment, varied across experimental sessions. In the low information treatment, participants are provided with the information that many previous participants in the study did not contribute to the public good. In contrast, the message in the high information treatment emphasizes that many previous study participants did contribute. The outcome of interest is the individual rate of contributions to the public good over all seven iterations of the game. For further insights on how the social information treatment affects behavior, see Diekert et al. (2021).[2]

Second, we use a data set on investment decisions under risk ($N = 425$) that was collected to study the effects of social information and luck (Dannenberg et al., 2021). The experiment was conducted in a second field trip to mostly the same Tanzanian fishing villages in 2020. Participants are endowed with 1,000 TZS and can choose how much of the endowment to invest and how much to keep. With a 50% chance,

[2]The original design is extensively studied in Diekert et al. (2021) and includes a treatment with a sanctioning institution. After contribution decisions in each round, participants enter a coordinated sanctioning stage within the three-person group that is based on the exclusion of an extra point. The exclusion does not alter the payoff-dominant strategy. Based on their initial finding that social information is only effective when combined with a sanctioning mechanism, we restrict our analysis to the treatment condition with a sanctioning institution.

the investment is tripled but otherwise lost. The non-invested amount is kept by the participant. Here, social information is provided in the form of a visual and verbal queue indicating either low or high average investment decisions by previous participants in the study.[3] In both data sets, social information is varied across experimental sessions.

4.2.3 Participant Characteristics

Participants self-report their role on board the fishing boat they currently work on.[4] Figures 4.1 and 4.2 display the role distribution by social information treatment in each sample. In both samples, the majority of participants are regular crew members (ca. 55% to 60%) while the rest is roughly balanced between owners and captains (ca. 20% each). The imbalanced distribution is expected as a fishing unit generally only has one captain and one owner but several crew members. While there is no explicit data on whether some fishermen fill more than one role, it is common that owners of small-scale fishing units also act as the captain, especially when they own a small number of vessels.

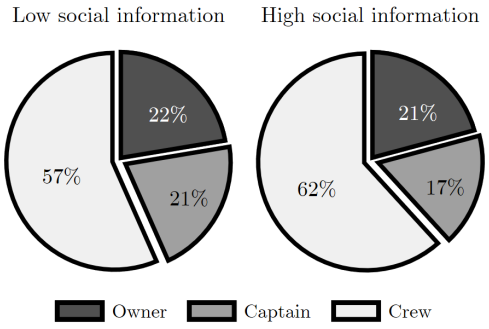


Fig. 4.1: Role distribution by social information for contributions to public good (N=287)

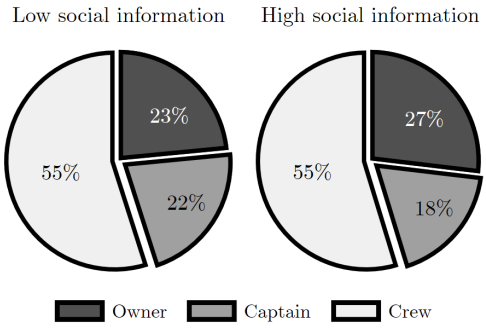


Fig. 4.2: Role distribution by social information for investment under risk (N=425)

Table 4.1 reports participants characteristics by role and social information treatment for both samples. In both samples, owners are older than captains and crew

[3] Dannenberg et al. (2021) extensively study the original design of the intervention which includes a treatment condition without social information. To preserve comparability with the prisoner's dilemma in Diekert et al. (2021) (who only study low and high social information), the treatment condition without social information is dropped from the analysis. Screenshots of all relevant choice situations as well as the experimental instructions for the two underlying experiments are available in the supplementary material, published online.

[4] 14 participants (seven in each sample) chose to not report their role and are therefore dropped from the analysis.

Table 4.1: Participant characteristics by role and social information treatment for each sample

	Contribution to public good			Investment under risk		
Social information	*low*	*high*	*p*-value	*low*	*high*	*p*-value
Owner	N=32	N=30		N=50	N=57	
- Age	42.2	42.0	.864	39.0	40.4	.416
- Tenure (> 10 years)	.47	.43	.981	.40	.46	.698
- Revenue (fish $\geq$ 75%)	.41	.40	1.000	.54	.44	.297
Captain	N=30	N=25		N=46	N=39	
- Age	34.1	35.6	.594	34.3	37.9	.055
- Tenure (> 10 years)	.27	.40	.447	.24	.56	.004
- Revenue (fish $\geq$ 75%)	.30	.32	1.000	.37	.54	.133
Crew	N=81	N=89		N=116	N=117	
- Age	33.7	37.2	.021	33.2	37.0	.003
- Tenure (> 10 years)	.31	.26	.578	.22	.41	.003
- Revenue (fish $\geq$ 75%)	.42	.22	.010	.41	.48	.157

Note: All balance tests are two-sample Wilcoxon Mann-Whitney mean comparison tests with exact p-values.

members. By implication, they report a higher tenure with about 40% to 45% of all owners indicating to be in the fisheries for at least ten years. In contrast, around 35% of captains and 30% of crew members report at least one decade of tenure. We further look at fishermen's reliance on fishing as an income source. Across all roles, around 40% of the participants earn at least 75% of their income from fishing, indicating a generally strong reliance on shared resources.

Participant characteristics are mostly balanced across social information treatments. Exceptions include a higher age and tenure for captains and owners in the high social information treatment for the 2020 sample on investment decisions under risk and a stronger reliance on fishing income for crew members in the low social information sample. Both tenure and the income share from fishing will be controlled for in the analysis.

4.2.4 Empirical Strategy

We study whether there is a heterogeneous treatment effect by role in the fisheries for the two social information interventions described above. For the data on contributions to a public good, the outcome variable of interest y_{i1} is the individual rate of contributions over the repeated game. To account for the proportional outcome, we use a fractional model with a probit function G that bounds the right hand side of the equation between zero and one (Papke and Wooldridge, 1996). With individual

controls C_i and tenure t as additional covariates, the model has the following form:

$$y_{i1} = G(\alpha_0 + \alpha_1 SI + \alpha_2 R_i + \alpha_3 SI \times R_i + \alpha_4 t_i + \alpha_5 SI \times t_i + \alpha_6 C_i)$$

Here, SI denotes the social information treatment (low vs high). Note that the treatment is interacted with the role R of fisherman i to compare the effectiveness of social information between owners, captains, and crew.

For the data on investment decisions under risk, we use a linear OLS regression model:

$$y_{i2} = \beta_0 + \beta_1 SI + \beta_2 R_i + \beta_3 SI \times R_i + \beta_4 t_i + \beta_5 SI \times t_i + \beta_6 C_i + u_{i2}$$

Again, the model poses an interaction between social information and a fisherman's role to test for heterogeneous treatment effects.

The main coefficients of interest are α_1, α_2, and α_3 as well as β_1, β_2, and β_3. They describe the treatment effect of social information on outcome y_{i1} and y_{i2} and the interaction of the treatment with different roles in the fisheries. In both models, we include an indicator variable for high tenure t (more than ten years) and an interaction of tenure with social information. We do so to control for possible confounding factors with respect to a leadership effect in the responsiveness to the intervention. That is, traditional leadership is often tied to experience (Henrich et al., 2015) and should be separated from the main leadership effect of interest, *i.e.*, leadership through the role of being an owner or a captain. Furthermore, both statistical models include individual controls C. Those include further treatment conditions of the original experimental design, the crew's size, the main fishing gear used, and an indicator variable for comprehension of the respective game's rules.

4.3 Results

4.3.1 Contributions to a Public Good

First, we analyze contributions to a public good in the data collected in 2018. The left graph in Figure 4.3 plots the average rate of contributions by role for low and high social information treatment. Owners contribute at an average rate of 41.1% with low and 59% with high social information. Contribution rates for captains are 39% with low and 54.9% with high social information while crew members contribute at a rate of 44.4% and 51.8%. The descriptive results suggest that the treatment is more

effective with owners and captains than it is for regular crew members.

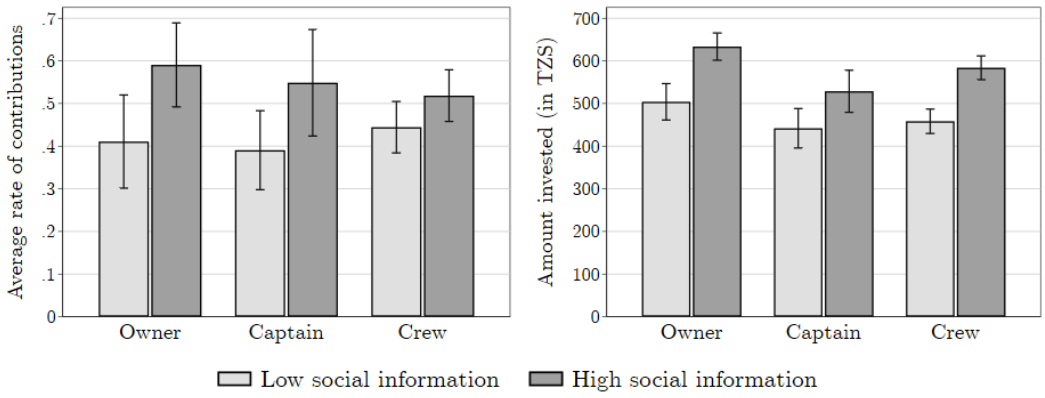


Fig. 4.3: Average rate of contributions to the public good (left graph) and investment decisions under risk (right graph) by role and social information treatment (low and high). Bars indicate ±1 standard error.

To study the treatment effect by role more closely, we report marginal effects from a fractional probit model on the contribution rate that interacts the social information treatment with a fisherman's role, see column (1) in Table 4.2. Crew members serve as the baseline. The model predicts that high social information induces only an insignificant increase in contributions of, on average, 5.2 percentage points ($p = .539$) with crew members. Behavioral predictions change for other roles. Owners contribute, on average, 11.5 percentage points less ($p = .130$) than crew members with low social information. However, the model indicates that the difference between low and high social information for owners is significantly larger than for crew members, the contribution rate increase is 16.9 percentage points larger ($p = .035$). Lastly, captains contribute, on average, 4.7 percentage points less than crew members with low social information ($p = .579$). Here, the difference between low and high social information is insignificantly larger than for crew members, captains' increase in contributions is 10.8 percentage points larger ($p = .439$).

Also, we find a relationship between tenure and social information treatment effect on contributions to a public good. Those fishermen that have more than a decade experience contribute significantly more with low social information than their counterparts with low tenure. The low share of high tenured captains in the low social information treatment may therefore introduce a weak downward bias in outcomes for captains with low social information. With high social information, tenure does not affect contributions (indicated by the significantly negative interaction term).

Table 4.2: Marginal social information treatment effects by role

	Contribution rate	*Investment*
	(1)	(2)
Intercept (low social info)	0.459***	541.08***
	(0.061)	(57.36)
High social info	0.052	146.15***
	(0.085)	(34.15)
Role = owner	-0.115	35.45
	(0.076)	(48.86)
Role = captain	-0.047	-13.27
	(0.085)	(34.25)
High social info × owner	0.169**	17.10
	(0.080)	(56.17)
High social info × captain	0.108	-37.98
	(0.140)	(54.93)
Tenure (> 10 years)	0.180***	59.33**
	(0.063)	(25.77)
High social info × Tenure (> 10)	-0.134**	-64.94*
	(0.066)	(34.37)
Individual controls	Yes	Yes
N	287	425

Note: The table reports average marginal effects from a fractional probit model in column (1). Individual controls include crew size, the main fishing gear, a Gneezy and Potters (1997) risk elicitation measure, an indicator for comprehension, and whether the participant has prior experience in economic experiments. In column (2), the table reports OLS regression coefficients. Here, individual controls include those of column (1) (except for the risk elicitation) and two further treatment conditions (a. high or low endowment, and b. luck or no luck prior to the investment decision). In both models, robust standard errors are clustered on the session level (in parentheses). ***, **, and * indicate significance at the 1,5, and 10% level.

4.3.2 Investment Decisions under Risk

Second, we analyze investment amounts in a decision under risk from the data collected in 2020. The right graph in Figure 4.3 plots the average amount invested by role for low and high social information. Out of 1,000 TZS, owners invest an average of 504 TZS with low and 633 TZS with high social information, an increase of 119 TZS. Captains invest an average of 441 TZS with low and 529 TZS with high social information (88 TZS increase) while crew members invest an average of 458 TZS (low) and 584 TZS (high), an increase of 126 TZS. We study the treatment effect by role in an OLS regression model on the amount invested in the risky option, see column (3) in Table 4.2. The model confirms that social information is effective for all fishermen. The predicted increase in investment for crew members is, on average, 146 TZS and is not significantly different for both owners and captains. Again, we

find a relationship between tenure and social information treatment. With low social information, fishermen with high tenure invest more than those with low tenure.

In summary, we show that the finding by Diekert and Eymess (2021) replicates for contributions to a public good. Social information induces a behavioral change only with resource users in leadership positions. The treatment effect is especially pronounced for fishermen in the role of vessel and gear owners while the intervention fails with regular crew members. In contrast, we observe a change in investment decisions under risk for fishermen in all roles. That is, owners, captains, and crew members alike are significantly influenced by social information. Jointly, our findings indicate that the effect of social information depends (i) on the behavioral context and (ii) on who is the recipient of the message.

4.3.3 Pooled Data Exercise

To study an interaction between social information intervention and the behavioral context, we pool the sample on contributions to a public good and the sample on investment decisions under risk. Outcomes in each sample are standardized to enable a direct comparison between the two behavioral contexts.[5] The standardized outcome measures whether a participant contributes (invests) more or less than the respective sample mean, see Appendix Figure C-1 for a distribution of the data. Importantly, the sample mean is not a baseline in which participants are not provided with social information. Yet, it is a reasonable assumption that a successful low information treatment would induce lower average outcomes while a successful high information treatment would induce higher average outcomes. Therefore, a difference in subsample averages (*e.g.*, by role) may indicate a bias in responses to the treatment.

To study whether the effect of social information depends on the behavioral context, we include interactions of the behavioral context (contribution to public good vs. investment under risk) with social information and a fisherman's role.

For an easy interpretation of the regression results, Figure 4.4 plots linear predictions of the standardized outcome by social information treatment and a fisherman's

[5]The 2018 sample on contributions to a public good and the 2020 sample on investment decisions were collected in the same communities. Some participants are re-sampled and have therefore been subject to a social information message in both data sets. These participants are dropped from the analysis when data sets are pooled, yielding a sample size of $N = 658$. Standardization follows the procedure suggested by Kling et al. (2007) and normalizes outcome Y in each of the two samples k according to $Y_k^* = (Y_k - \mu_k)/\sigma_k$ with μ_k as the sample mean and σ_k as the sample standard deviation. Hence, standardized outcomes in each sample have a mean of zero and a standard deviation of one. For distributions of the standardized outcome in each separate sample and the pooled sample, see Appendix C-2.

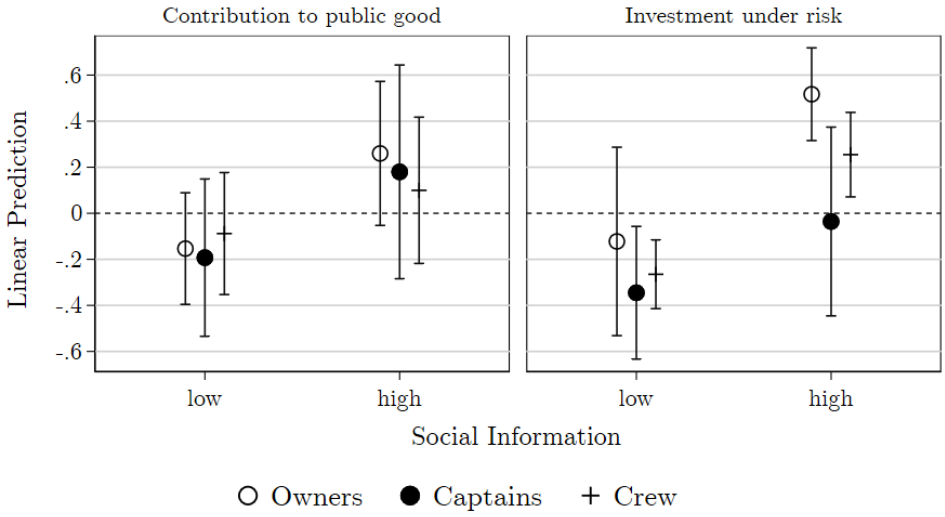


Fig. 4.4: Average marginal effects by social information treatment and role for contributions to a public good (left plot) and investment decisions under risk (right plot). Bars indicate a 95% confidence interval.

role for both behavioral contexts.[6] Results support the conjecture that the effect of social information depends on the behavioral context and the recipient of the message. For owners, social information is effective in both contexts, indicated by the substantial differences between linear predictions for low and high information. The relationship is less pronounced for captains but there appears to be no differences in the effectiveness of social information between the two contexts. For crew members, we observe a difference between behavioral contexts. They are not responsive when social information attempts to induce a change in contributions to a public good. In contrast, we observe significant behavioral change for investment decisions under risk.

4.4 Discussion

We find evidence that supports the exploratory result by Diekert and Eymess (2021): resource users in leadership positions, especially vessel and gear owners, change their contributions to a public good when they receive social information about the contribution behavior by others. Those that are not in leadership positions, *i.e.*, regular crew members, are unaffected by the same intervention. Although the lack of a baseline treatment without social information prevents a more nuanced interpretation of

[6] For the traditional tabular presentation, see the regression coefficients in Appendix Table C-1.

whether a norm intervention is more effective in promoting or dissuading contributions to a public good, our study has important policy implications. Policy makers may target owners and to a lesser extent also captains as facilitators of collective change in fisheries management.

Furthermore, we show that the heterogeneous treatment effect for leadership experience identified by Diekert and Eymess (2021) depends on the behavioral context. In an unrelated second social information intervention, the leadership result does not replicate. In particular, when social information attempts to induce a change in investment decisions under risk, fishermen in all roles are affected.

Our findings and the studies by Diekert et al. (2021) and Diekert and Eymess (2021) suggest that the activation or creation of social norms is a promising approach for natural resource governance. As a policy tool, social norms are cost-effective and work in true bottom-up fashion. They provide communities with much needed agency in resource governance and may help to restore trust in local co-management structures that are undermined by issues of corruption and non-compliance (Nunan et al., 2018). To initiate such normative change, policies should target resource users in leadership positions. Importantly, if pro-social leadership can be activated for good resource governance, benefits may spread. That is, policy makers should try leverage the fact that leaders that lead by example may crowd-in contributions by others (Jack and Recalde, 2015) and that pro-social leaders are able to maintain contributions through efficient punishment (Kosfeld and Rustagi, 2015; Kosfeld, 2020).

Further research is needed to understand the forces that enable communities to facilitate sustainable resource management and how these efforts interact with local leadership. With respect to social norms as policy tools, researchers and practitioners alike should field-test norm activation through well-designed interventions. Our findings and the results in Diekert and Eymess (2021) suggest that resource users in leadership positions should remain in positions of power to leverage their propensity for behavioral change. For example, policy makers may target production inputs such as legal gear use, a decision in which owners often have exclusive authority. Other avenues for future research involve network analyses to understand whether captains and owners can function as opinion leaders in communities that are traditionally led by a few influential individuals (Crona and Bodin, 2010). At the same time, one may study why crew members neglect the social information designed to nudge them towards contributing to a public good. Ultimately, resource governance depends on the development of true forms for participatory resource management that neither neglect societal nor ecological needs.

Bibliography

Arbak, E. and Villeval, M.-C. (2013). Voluntary leadership: motivation and influence. *Social Choice and Welfare*, 40(3):635–662.

Barnes-Mauthe, M., Gray, S. A., Arita, S., Lynham, J., and Leung, P. (2014). What determines social capital in a social–ecological system? insights from a network perspective. *Environmental Management*, 55(2):392–410.

Bergquist, M., Nilsson, A., and Schultz, W. P. (2019). A meta-analysis of field-experiments using social norms to promote pro-environmental behaviors. *Global Environmental Change*, 59:101941.

Bodin, Ö. and Crona, B. I. (2008). Management of natural resources at the community level: exploring the role of social capital and leadership in a rural fishing community. *World Development*, 36(12):2763–2779.

Cappelen, A. W., Reme, B.-A., Sørensen, E. Ø., and Tungodden, B. (2016). Leadership and incentives. *Management Science*, 62(7):1944–1953.

Cepić, D. and Nunan, F. (2017). Justifying non-compliance: The morality of illegalities in small scale fisheries of lake victoria, east africa. *Marine Policy*, 86:104–110.

Clark, C. W. (2006). *The worldwide crisis in fisheries: economic models and human behavior*. Cambridge University Press.

Cowx, I. G. and Ogutu-Owhayo, R. (2019). Towards sustainable fisheries and aquaculture management in the african great lakes. *Fisheries Management and Ecology*, 26(5):397–405.

Cowx, I. G., van der Knaap, M., Muhoozi, L., and Othina, A. (2003). Improving fishery catch statistics for lake victoria. *Aquatic Ecosystem Health & Management*, 6(3):299–310.

Crona, B. and Bodin, Ö. (2010). Power asymmetries in small-scale fisheries: a barrier to governance transformability? *Ecology and Society*, 15(4).

Dannenberg, A., Diekert, F., and Händel, P. (2021). The effects of social information and luck on risk behavior of small-scale fishers. Manuscript, University of Kassel.

Diekert, F. and Eymess, T. (2021). Changing collective action: Social information increases cooperation of teams in a prisoner's dilemma. Manuscript, Heidelberg University.

Diekert, F., Eymess, T., Luomba, J., and Waichman, I. (2021). The Creation of Social Norms under Weak Institutions. Heidelberg University, Department of Economics Discussion Paper No. 684.

Eggert, H. and Lokina, R. B. (2010). Regulatory compliance in lake victoria fisheries. *Environment and Development Economics*, 15(2):197–217.

Etiegni, C., Irvine, K., and Kooy, M. (2017). Playing by whose rules? community norms and fisheries rules in selected beaches within lake victoria (kenya) co-management. *Environment, Development and Sustainability*, 19(4):1557–1575.

Etiegni, C. A., Irvine, K., and Kooy, M. (2020). Participatory governance in lake victoria (kenya) fisheries: whose voices are heard? *Maritime Studies*, pages 1–19.

Farrow, K., Grolleau, G., and Ibanez, L. (2017). Social norms and pro-environmental behavior: a review of the evidence. *Ecological Economics*, 140:1–13.

Gatiso, T. T., Vollan, B., Vimal, R., and Kühl, H. S. (2018). If possible, incentivize individuals not groups: evidence from lab-in-the-field experiments on forest conservation in rural uganda. *Conservation Letters*, 11(1):e12387.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645.

Gutiérrez, N. L., Hilborn, R., and Defeo, O. (2011). Leadership, social capital and incentives promote successful fisheries. *Nature*, 470(7334):386–389.

Henrich, J., Chudek, M., and Boyd, R. (2015). The big man mechanism: how prestige fosters cooperation and creates prosocial leaders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1683):20150013.

Irvine, K., Etiegni, C., and Weyl, O. (2019). Prognosis for long-term sustainable fisheries in the african great lakes. *Fisheries Management and Ecology*, 26(5):413–425.

Jack, B. K. and Recalde, M. P. (2015). Leadership and the voluntary provision of public goods: Field evidence from bolivia. *Journal of Public Economics*, 122:80–93.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.

Kosfeld, M. (2020). The role of leaders in inducing and maintaining cooperation: The cc strategy. *The Leadership Quarterly*, 31(3):101292.

Kosfeld, M. and Rustagi, D. (2015). Leader punishment and cooperation in groups: Experimental field evidence from commons management in ethiopia. *American Economic Review*, 105(2):747–83.

Krishna, A. (2002). *Active social capital: Tracing the roots of development and democracy*. Columbia University Press.

Nielsen, J. R., Degnbol, P., Viswanathan, K. K., Ahmed, M., Hara, M., and Abdullah, N. M. R. (2004). Fisheries co-management – an institutional innovation? lessons from south east asia and southern africa. *Marine Policy*, 28(2):151–160.

Nunan, F., Cepić, D., Yongo, E., Salehe, M., Mbilingi, B., Odongkara, K., Onyango, P., Mlahagwa, E., and Owili, M. (2018). Compliance, corruption and co-management: how corruption fuels illegalities and undermines the legitimacy of fisheries co-management. *International Journal of the Commons*, 12(2).

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge university press.

Ostrom, E. (2008). The challenge of common-pool resources. *Environment: Science and Policy for Sustainable Development*, 50(4):8–21.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–632.

Song, F. (2009). Intergroup trust and reciprocity in strategic interactions: Effects of group decision-making mechanisms. *Organizational Behavior and Human Decision Processes*, 108(1):164–173.

Thorlindsson, T. (1988). The skipper effect in the icelandic herring fishery. *Human Organization*, pages 199–212.

Vázquez-Rowe, I. and Tyedmers, P. (2013). Identifying the importance of the "skipper effect" within sources of measured inefficiency in fisheries through data envelopment analysis (dea). *Marine Policy*, 38:387–396.

Appendix

C-1 Regression Analysis

Table C-1 reports regression coefficients for the results plotted in Figure 4.4. The dependent variable is the standardized outcome which allows a direct comparison of the social information treatment effect between the two behavioral contexts. Crew members serve as the baseline category for a fisherman's role and the contribution to a public good is the baseline context. In general, the model is not able to pick up statistically significant differences between behavioral contexts. Yet, at certain points, the regression coefficients are substantial in size to warrant interpretation.

The direction and sign of the role and social information coefficients for the contribution game indicate that the treatment is more effective with both owners and captains than it is with regular crew members. Also, the large coefficients for high social information in the invest game indicates that regular crew members are more effected by the treatment in the investment context than they are in the contribution context. Lastly, the substantial and negative interaction term for high social information and captains in the investment game suggest that they are less strongly affected than crew members and owners.

Table C-1: Marginal social information treatment effects by role

	Standardized outcome
	(1)
Intercept (low social info)	-0.223
	(0.136)
High social info	0.290
	(0.221)
Role = owner	-0.065
	(0.165)
Role = captain	-0.105
	(0.153)
High social info × owner	0.225
	(0.184)
High social info × captain	0.185
	(0.331)
Context = invest	-0.177
	(0.144)
High social info × invest	0.332
	(0.228)
Owner × invest	0.208
	(0.265)
Captain × invest	0.024
	(0.209)
High social info × owner × invest	-0.106
	(0.313)
High social info × captain × invest	-0.395
	(0.416)
Tenure (> 10 years)	0.335***
	(0.107)
High social info × Tenure (> 10)	-0.287**
	(0.133)
Comprehension	-0.028
	(0.082)
N	658

Note: The table reports OLS regression results. Standard errors are clustered on the session level (in parentheses). ***, **, and * indicate significance at the 1,5, and 10% level.

C-2 Distribution Plots of Standardized Outcome

Figure C-1 plots the distribution of standardized outcomes for the 2018 sample on contributions to a public good (top left), the 2020 sample on investment decisions under risk (top right), and the pooled sample (bottom left). The standardization of the 2018 sample retains the indication that the data has a bi-modal distribution. That is, participants either contributed in all or none of the iterations of the repeated prisoner's dilemma. In contrast, the standardized investment sample is normally distributed.

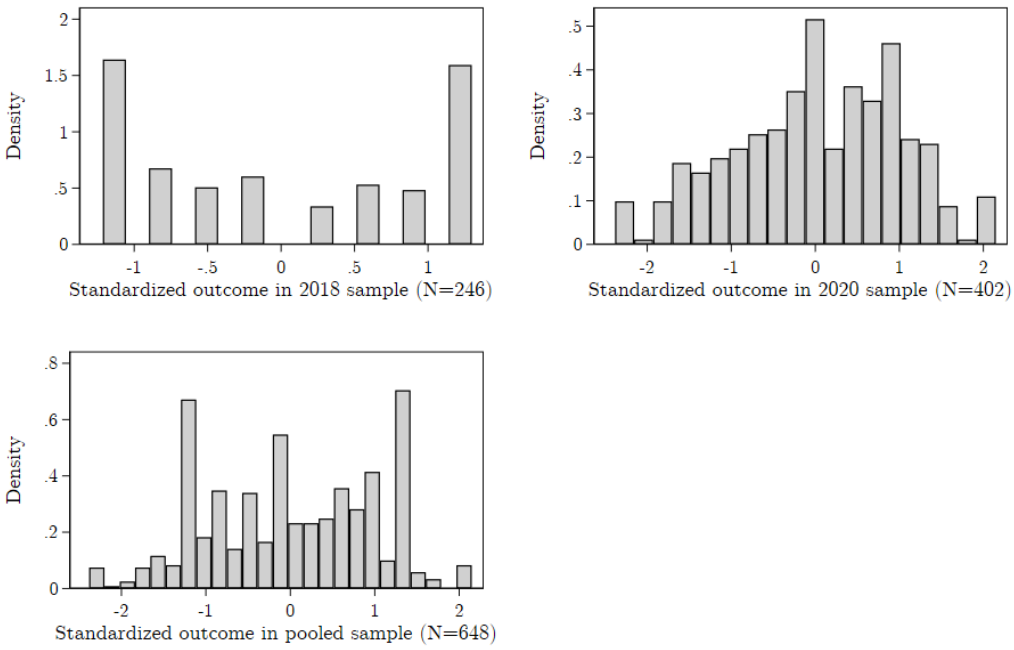


Fig. C-1: Distribution of standardized outcomes by sample and for the pooled data set