Beijer Institute OF ECOLOGICAL ECONOMICS | KUNGL. VETENSKAPS-AKADEMIEN THE ROYAL SWEDISH ACADEMY OF SCIENCES

DISCUSSION PAPER

Uncertainty can help protect local commons in the face of climate change

Beijer Discussion Paper Series No. 270

Caroline Schill and Juan Carlos Rocha. 2019.

Beijer Institute of Ecological Economics
The Royal Swedish Academy of Sciences
Lilla Frescativägen 4A
Stockholm
+46 8 673 95 00
beijer@beijer.kva.se
www.beijer.kva.se

Uncertainty can help protect local commons in the face of climate change*

October 12, 2019
Preliminary: Comments welcome.

Caroline Schill[1, 2, a] and Juan Carlos Rocha[1, 2]

[1] The Beijer Institute of Ecological Economics, the Royal Swedish Academy of Sciences
[2] Stockholm Resilience Centre, Stockholm University
[a] Corresponding author, email: caroline.schill@beijer.kva.se

Climate change is likely to trigger abrupt and potentially persistent changes in the structure and function of marine ecosystems[1,2]. Such 'regime shifts'[3] threaten the livelihoods of millions of people reliant on small-scale fisheries[4]. Yet, it is largely unknown how resource users cope with regime shifts, their uncertainty, and impacts. Here, we assess the potential for local collective action to avert uncertain, yet catastrophic, regime shifts. We conducted behavioural economic experiments with small-scale fishers (*n*=256) using a framed, dynamic common-pool resource game to test the effect of different degrees of uncertainty about the presence of climate-induced thresholds on exploitation patterns. Results from four communities in the Colombian Caribbean show that groups facing uncertain thresholds are likely to adapt in the sense that they sustain higher stock levels in order to avoid a regime shift. However, catch inequalities in the game, and community-level attributes appear to mitigate or even eliminate this effect; illustrating the critical role contexts play for behaviour. Our results suggest a more positive outlook regarding the inherent uncertainties of climate change compared to experimental evidence overwhelmingly proposing a negative relationship between uncertainty, collective action, and sustainable resource use[5–8]. Instead, we conclude that in certain circumstances uncertainty can help protect the commons.

Thresholds have been identified in a large variety of systems and scales, including climate[9] and marine ecosystems[10]. Although climate change alone is responsible for many changes in both structure and function of marine ecosystems[1], regime shifts often occur in conjunction with other more local activities such as fishing or agriculture[3]. Examples of marine regime shifts include the collapse of fisheries, hypoxia, and coral transitions[11], and their frequency and intensity is expected to increase with climate change[2]. These regime shifts have been established both theoretically and empirically[12] but the potential for any specific ecosystem to exhibit thresholds is uncertain. Yet, when regime shifts occur they can substantially impact ecosystems son which local communities heavily rely for food or clean water[3].

* This discussion paper, in the same version, is also vailable at SSRN: https://ssrn.com/abstract=3468677 or http://dx.doi.org/10.2139/ssrn.3468677.

Here, we investigate how groups of fishers, heavily relying on local marine ecosystems, respond to different degrees of uncertainty (certainty, risk, and ambiguity) regarding the existence of a climate-induced threshold below which the productivity of their shared fishing ground would drastically reduce. For this purpose, we designed a controlled field experiment in the form of a framed, dynamic common-pool resource (CPR) game. Our insights offer indications about the importance of risk assessments for affected communities[13], and contribute to the sparse empirical literature about decision-making in ambiguous environments[8,14]. Our study extends the emerging body of experimental research on collective action around shared resources in relation to uncertainty, and specifically ambiguity, about critical thresholds and regime shifts motivated by the climate change challenge[7,8,15–17].

Empirical evidence suggests a negative influence of uncertainty on cooperation and the sustainable use of shared resources. Experimental research in static[5,18,19] and dynamic[6] settings has shown that uncertainties around the availability or regeneration of a CPR leads to its unsustainable use; and threshold impact uncertainty[7,16] as well as threshold location uncertainty[7,17] leads to the breakdown of cooperation around public goods. Communication seems to play a critical role. In the absence of large uncertainties about the threshold's location, communication can drastically increase the chances of preventing negative outcomes, as it allows participants to coordinate their actions around the (relatively well-known) threshold location[7,20]. CPR experiments about endogenously driven regime shifts in which communication was allowed, point to the same mechanism: a certain or very likely threshold can mobilize collective action and serve as focal point around which people coordinate their actions to sustain the shared resource[15,21].

These studies, and generally the bulk of behavioural evidence, is based on lab studies with 'WEIRD' participants, i.e. students from Western, Educated, Industrialized, Rich, and Democratic countries[22]. Mounting evidence, however, highlights how behaviour can differ strongly between (sub-)populations and various socio-cultural and ecological contexts[23,24]. It is therefore imperative to conduct field experiments in different communities. To our knowledge, the effect of uncertainty (in the form of risk) about the existence of thresholds on collective action and CPR use patterns has so far only been investigated in the lab[15], and the effect of ambiguous thresholds only in public goods settings, and also only in the lab[8]. We brought our game to the Colombian Caribbean coast and played it with small-scale fishers from four different communities (see *Supplementary Information 1* (*SI*)). All communities rely on fishing as the primary source of livelihood and have been exposed to over-exploitation and marine regime shifts such as fish and mangrove die-offs partially driven by climate change.

In our game, groups of four shared a fishing ground (Figure 1a and *Methods*), with an initial and maximum stock of 50 fish. Over a number of rounds, unknown to the players, each player decided privately how much fish to catch. Communication was allowed. Each fish was worth COP 500 (about € 0.15). The players knew that the more fish they catch, the less would be available in future rounds. At the end of each round, the fish stock regenerated according to its size (Figure 1b) and was announced. In Stage 1 of the game (6 rounds), all groups played the game without a threshold (*baseline*, Figure 1b, lower graph). In Stage 2 (10 rounds), apart from the groups that were randomly allocated to continuously play *baseline* (control), groups were told that a climate event had definitely led, or could lead, to a drastic reduction of the stock's productivity below a certain level (Figure 1b, upper graph). Hence, the threshold was introduced exogenously, but players could avoid (or reverse) the regime shift by keeping the stock above the threshold.

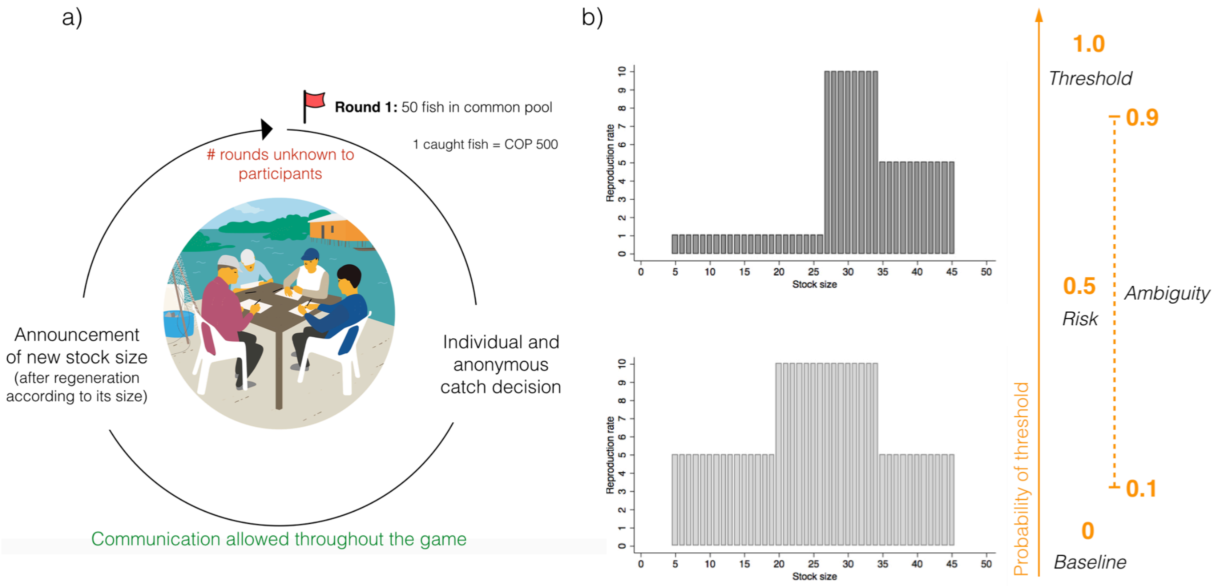

Figure 1. Experimental design. a) Sketch of the set-up of our controlled pen-and-paper field experiment, which was the same for all treatments: groups of four small-scale fishers shared a fishing ground. They were seated around a table, and face-to-face communication was allowed. At the end of each round, the fish stock regenerated according to its size (see stock dynamics in b). The only piece of information the players received was the new stock size at the beginning of each new round. b) The graphs show the relationship between the size of the fish stock and its reproduction as used in the experiment. The lower graph represents the stock dynamics without a threshold (*baseline*, Stage 1 of the game and control in Stage 2). The upper graph represents the stock dynamics with a threshold at a level of 28. The arrow indicates the probability with which a climate event (in any given round in Stage 2 of the game) could lead to a drastic and long-lasting change in the stock dynamics in the form of a threshold (see *Fig. S3* for an overview about when the climate event has happened for each group). The three different treatments (*threshold*, *risk*, and *ambiguity*) and the *baseline* case are indicated in italics. See *SI2.1* for instructions and *Fig. S2* for the visualisations used to communicate the stock dynamics to the players. Illustration in a) by E. Wikander/Azote.

Using three treatments, we tested whether exploitation patterns differ depending on whether there is *certainty* (known probability of 1.0), *risk* (known probability of 0.5) or *ambiguity* (known probability range: 0.1-0.9) about the threshold. In particular, we were interested 1) whether groups sustain higher stock sizes in the face of a (potential) threshold (in order to avoid the potential regime shift) compared to a situation without a threshold (*baseline*), and 2) whether the degree of threshold uncertainty affects overall sustained stock levels. Hence, we focus our analysis on overall sustained stock sizes. Due to non-normally distributed data, we use median stock sizes (across all rounds per group) and rely on non-parametric tests. See *SI3.1* for details on statistical analysis.

Fishers playing the *ambiguity* treatment sustained higher median stock sizes (Figure 2). Comparing each treatment with *baseline* (no threshold) shows that the higher the uncertainty, the more distinct the differences in distribution of median stock sizes. Statistical analysis (set of Mann–Whitney–Wilcoxon rank-sum tests (MWW)) shows that differences in median stock sizes between *baseline* and *threshold* and *baseline* and *risk* are not significant (see *Table S4* for results), but we find significant differences between *baseline* and *ambiguity* (MWW, P=0.0822). *Ambiguity* groups also spend on average significantly more time above a stock size of 28 (MWW; P=0.064), and deplete the stock significantly less (five depletion cases in *baseline*, and none in *ambiguity*; Fisher's exact test (FET), P=0.043) compared to *baseline* groups. (*Table S5* shows comparisons between baseline and the other treatments).

We do not find differences in median stock levels depending on the different degrees of uncertainty (see *Table S6* for test results). Moreover, the degree of uncertainty does not seem to play a role in determining the likelihood of groups crossing the potential threshold at some point during the game (FET; P=0.934), or cases of depletion (FET; P=0.220).

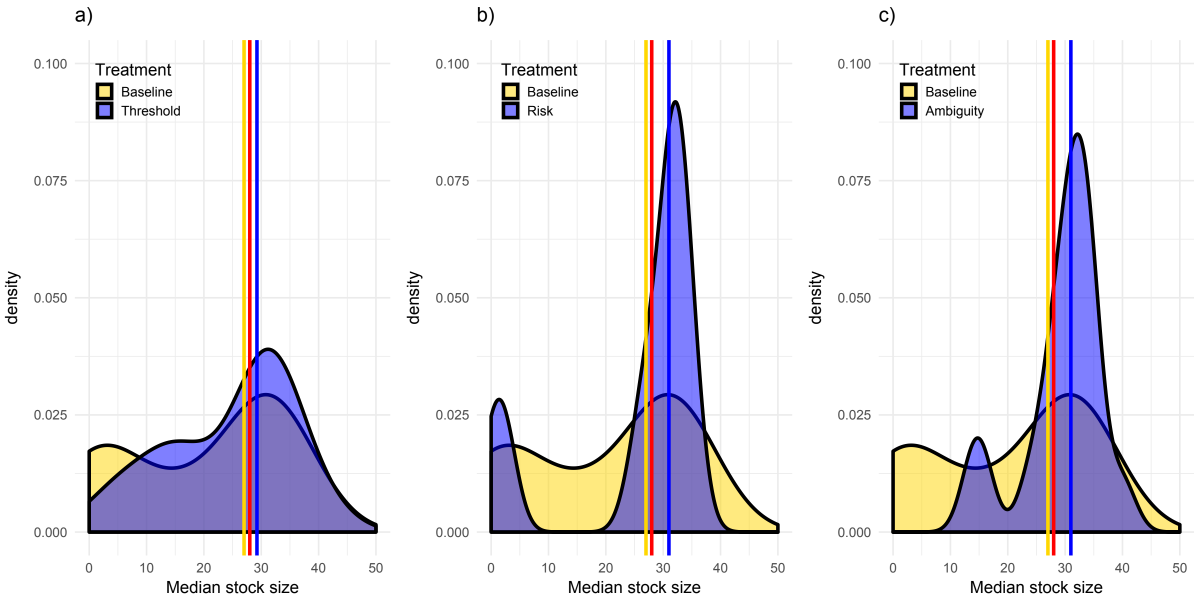

Figure 2. Treatment effects. Comparison of median stock size (after catch, before regeneration) distributions. Individual graphs show the distribution differences between (a) *baseline* and *threshold* treatment, (b) *baseline* and *risk* treatment, and (c) *baseline* and *ambiguity* treatment (n=16 for each distribution). Red lines indicate (potential) threshold. The yellow lines indicate the median stock size for *baseline* (27). The blue line indicates in (a) the median stock size of the *threshold* treatment (29.25), in (b) of the *risk* treatment (31), and in (c) of the *ambiguity* treatment (31). Moving from a) to c) median treatment stock size levels increase and distribution narrows. All observations are from Stage 2.

On the backdrop of growing evidence about the importance of context for behaviour, we compare game outcomes between the four communities (community A-D, Figure 3). Whereas median group stock sizes in Stage 1 (where treatment effects were absent) do not differ between communities (Kruskal-Wallis test (KW), df=3, χ^2=5.417, P=0.144), they do in Stage 2 (KW, df=3; χ^2=15.671, P=0.0013). See *Table S7* and *Table S8* for comparisons for other variables. This difference can be explained by significant differences with respect to how often groups achieved the highest regeneration rate (KW, χ^2_{St2}=24.617, P_{St2}=0.0001), as well as striking differences in cases of depletion (FET, P_{Stage1}=0.181, P_{Stage2}=0.001): 44% of the groups in community B (19% in Stage 1), 25% in community D (6% in Stage 1), and none in community A and B (in neither stage) depleted in Stage 2 (Figure 4). All groups that depleted in Stage 1 also depleted in Stage 2.

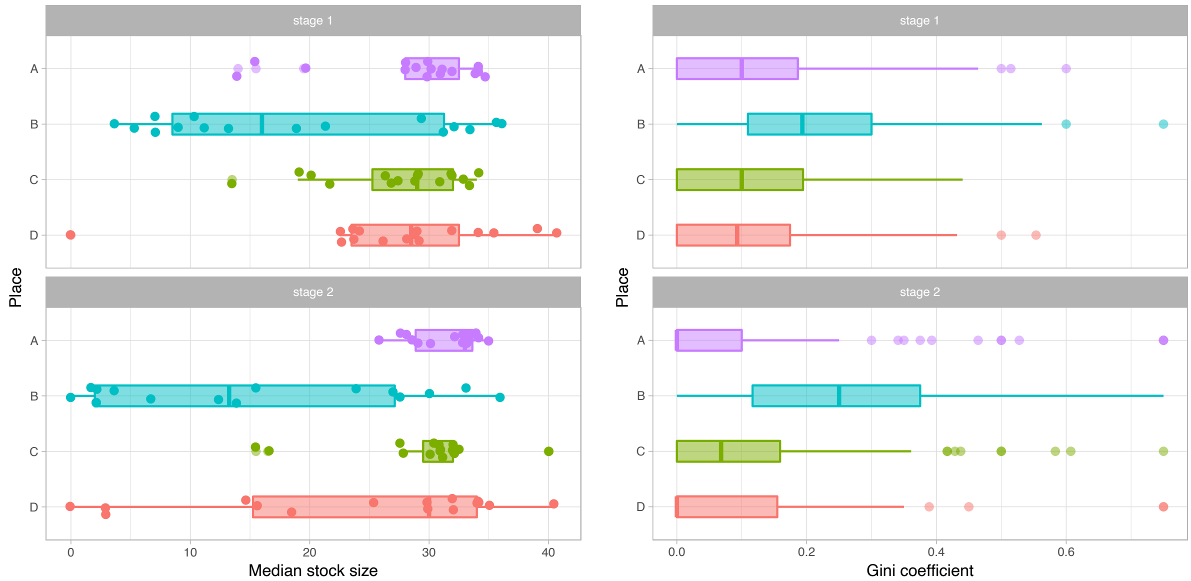

Figure 3. Community differences. Left panels: box plots (*N*=16 per box) of median stock sizes across communities in Stage 1 (upper panel) and Stage 2 (lower panel). Right panel: box plots (*N*=16 per box) of Gini coefficient of group catch (dispersion of total catch within each group) in Stage 1 (upper panel) and Stage 2 (lower panel). The differences between communities was significant for both stages (KW, *df*=3; χ^2_{Stage1}=7.85, P_{Stage1}=0.0492; χ^2_{Stage2}=14.20, P_{Stage2}=0.0026).

It appears from Figure 4 that the treatment seemed to matter more or less for the communities. For example, in community A, no matter the treatment, all groups sustained higher stock levels and in most instances above the threshold. In community D, on the other hand, all *ambiguity* groups stayed above the threshold. Moreover, there were significant community differences in whether a group crossed the (potential) threshold at some point during the game (FET, *P*=0.005). In community B 90% of the groups crossed the potential threshold (Figure 3), in community A it was only 25%, in community C 50%, and in community D 33%.

Community differences in exploitation patterns are likely due to how well groups cooperated (see *SI3.2.1* for other potential explanations). As proxy for cooperation we use the Gini coefficient of group catch (dispersion of total catch within each group), assuming that catch equality is based on players using the opportunity to communicate to agree on collective exploitation levels (*SI3.3*). The Gini coefficient was quite low overall (Stage 1: 0.097±0.093; Stage 2: 0.081±0.098, Figure 3, right panels) and did not differ significantly between both stages (MWW, *P*=0.155), but there were significant differences between communities for both stages (Figure 3).

We find indeed a positive relationship between equal sharing of catch and sustainable resource use (Figure 4). Groups that depleted the stock in Stage 2 had significantly lower Gini coefficients than groups that did not (MWW, *P*=0.008), and *threshold*, *risk* and *ambiguity* groups that sustained the stock above the threshold throughout Stage 2 had significantly lower Gini coefficients than groups that crossed the threshold at some point (MWW, *P*<0.0001). Interestingly, groups with very low Gini coefficients in Stage 1, i.e. with a positive experience, have a significantly lower Gini coefficient in Stage 2 (MWW, P<0.0001) and vice versa (MWW, P<0.0001).

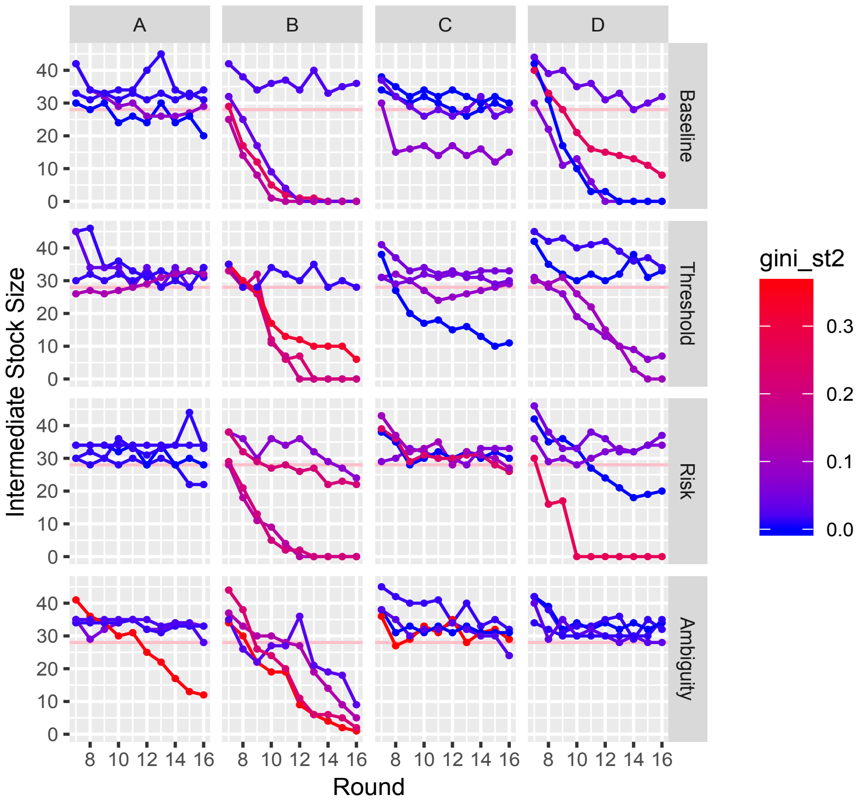

Figure 4. Stock size patterns over time. a) shows time series of intermediate stock size (after catch, before re-growth) over time for each group per treatment (vertical variation) and place (A-D; horizontal variation) in Stage 2. The colour coding indicates the overall Gini coefficient of group catch (Stage 2). The figure shows that in community A 25% of the groups crossed the potential threshold, in community B it was 90%, in community C 50%, and in community D 33%; these differences are significant (Fisher's exact test, P=0.005). The figure also shows that in community A and B no group depleted and in community B 44% of the groups (19% in Stage 1) and 25% in community D (6% in Stage 1). Note, no group in community A and B depleted either in Stage 1 and all groups that depleted in Stage 1 also depleted in Stage 2. See *Fig. S3* for indications whether and when the climate event has happened for all *risk* and *ambiguity* treatment groups.

To bring these results together and to determine average treatment effects given group- and community-level factors, we fitted a multiple linear regression model with 'median stock size' as response variable, while controlling for treatment, community, and group differences (Table 1A, left). The results corroborate that *ambiguity* groups are more likely to sustain higher median stock sizes but that indeed community effects can mitigate or even eliminate the treatment effect. For example, the effect of community B on median stock sizes is likely to be negative, and stronger than the positive effect of *ambiguity*. However, the negative effects of community D, are likely not to be stronger than the effect of *ambiguity*. We also find that *risk* groups sustain higher stock sizes Group differences play a significant role as well; groups that achieved higher median stock sizes in Stage 1 are more likely to do so too in Stage 2.

Table 1. Regression analysis. Estimated average treatment effects based on community, and group differences (median stock size in Stage 1 and/or Gini coefficient in Stage 1, independent of treatment), see S13.4 for details. A) left: model with 'median stock size' as response variable (N=64, R^2=0.656; P<0.0001; see also *Table S9*); right: model with 'percentage of rounds above threshold' as response variable (N=64, R^2=0.566; P<0.0001; see also *Table S10*). B) (*threshold* treatment as 'reference group') left: model with 'median stock size' as response variable (N=48, R^2=0.645; P<0.0001; *Table S11*); right: model with 'percentage of rounds above threshold' as response variable (N=48, R^2=0.553; P<0.0001; *Table S12*).

	A)		**B)**	
	Median stock size	**Percentage of rounds above threshold**	**Median stock size**	**Percentage of rounds above threshold**
Threshold	2.223	0.0912	--	--
	(0.414)	(0.448)	--	--
Risk	5.940^{+}	0.222*	3.142	0.0983
	(0.067)	(0.040)	(0.272)	(0.375)
Ambiguity	7.575**	0.204*	5.043*	0.113
	(0.005)	(0.044)	(0.026)	(0.265)
B	-10.22**	-0.359**	-10.67**	-0.466***
	(0.004)	(0.004)	(0.005)	(0.000)
C	-1.198	-0.0545	-0.943	-0.0478
	(0.463)	(0.514)	(0.642)	(0.640)
D	-6.907**	-0.185*	-4.154^{+}	-0.146
	(0.005)	(0.035)	(0.091)	(0.145)
Median/Mean Stock St 1*	0.685***	0.0148*	0.419^{+}	
	(0.000)	(0.025)	(0.059)	
Gini St 1		-0.874*	-21.49	-1.387**
		(0.044)	(0.270)	(0.002)
Constant	8.063*	0.377^{+}	19.04*	0.949***
	(0.048)	(0.060)	(0.014)	(0.000)
R^2	0.656	0.566	0.645	0.553
Adjusted R^2	0.613	0.502	0.583	0.487
Observations	64	64	48	48

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
* Median Stock St 1 and Gini St 1 are negatively correlated; Pearson's r = -0.54. Mean Stock Stock St 1 was used for the model with 'Percentage of rounds above threshold' as response variable.

To test the effect of the degree of uncertainty on exploitation patterns, while controlling for community and group effects, we fitted in a last step of this analysis two models using *threshold* treatment groups as reference group (i.e. excluded *baseline* groups) with the response variables 'median stock size', and 'percentage of rounds above threshold' (Table 1B). We find that groups confronted with ambiguous thresholds are more likely to sustain higher median stock sizes compared to groups that face thresholds with certainty. *Risk* is not significant. As above, for median stock sizes, community B is likely to have a negative and stronger effect than *ambiguity*, the effect of community D is likely to mitigate the *ambiguity* effect, and group differences are likely to be positive. However, neither *ambiguity* nor *risk* seem to explain the percentage of rounds a group spends above the threshold, indicating the relative insignificance of precise risk assessments.

Our results contrast previous experimental evidence[5–8]. One explanation can be found in our particular game design. Our players do not receive direct feedback about their group members' actions, they only get to know the new stock size at the beginning of each round. Uncertainty masks this feedback even more. Whereas in the *threshold* treatment, there is little doubt in the mind of a player that sudden and unexpected low stocks are the result of others having extracted much, in the *risk* and *ambiguity* treatments, players could instead believe that sudden low stocks are due to the climate event. The latter would not erode trust in group members and the willingness to cooperate, which is commonly conditional[25,26]. Additionally, here, groups can use communication to coordinate their actions. This can be instrumental for sustainable CPR use[15,21]. Both these elements bring our experimental design closer to real settings in small-scale commons dilemmas.

Our results suggest that communicating potential regime shifts and their consequences at the local level can trigger adaptation, however, how strong seems to depend on the community in question. In the Caribbean, this might be especially important in periods of draught or ENSO years when hypoxic and fish-die-off events are more likely to occur[27,28]. We did not find a relationship between the degree of threshold uncertainty and whether or not groups cross the threshold at some point, neither with respect to the percentage of time they spend above the threshold during the game. This suggests to place less focus on determining more precise likelihoods of thresholds but rather focus on identifying what variables have thresholds, and their potential consequences. These insights are particularly important for regions where relevant ecological data is sparse but livelihoods can be severely threatened by regime shifts. In such contexts, it seems more effective to invest in governance arrangements providing collective decision-making fora, as previously suggested by a similar CPR lab experiment[15]. Furthermore, when considering such local governance and adaptation plans, our findings highlight the importance of acknowledging context-dependencies as well as the attributes of the sub-populations in question.

Gaining insights into potential behavioural responses of individuals and communities facing uncertain ecological thresholds is pivotal for dealing with climate change impacts. Our results show that uncertainty around critical climate-induced thresholds can protect shared resources. This is in stark contrast with previous experimental evidence that overwhelmingly suggests a negative relationship between uncertainty, collective action, and sustainable resource use[5–8]. Hence, our study provides hopeful insights given the irreducible uncertainties inherent to climate change, and environmental change more broadly.

Methods

The CPR fishery game was performed with 256 small-scale fishers from four communities (see *Fig. S1* for map) along the Colombian Caribbean coast in February 2016. In each community, we spent two days running four groups (with 4 players each) in parallel in the morning and afternoon. The recruitment was through community leaders that recruited fishers in each community before we arrived. We asked the leaders to try to include fishers from different ages and that use different fishing methods. At the beginning of each session, we briefly introduced ourselves, and provided an overview (including purpose and duration) of the activity. We clarified that the use of money in the game was not a payment for attendance but to make decision-making realistic. Once the participants signed the consent forms, one experimenter explained the game to all 16 participants before they were randomly allocated to four groups with 4 participants each. We made sure that members from the same household were not in the same group and we also tried as well as possible to avoid that fishers from the same fishing crew were in the same group. Before starting with the actual game, each group played several practice rounds to become familiar with the rules and dynamics of the game. We used a rotation scheme for the experimenters to avoid the experimenter effect and made sure that groups could not overhear each other.

In the game, groups (N=64) of four players exploited a common fishing ground with an initial and maximum stock of 50 fish (see *SI2.1* for instructions). Since communication was allowed, non-binding agreements could emerge (e.g., about catch rates), but decisions were private and kept confidential. At the beginning of each round, the new fish stock level (x) was announced. In Stage 1 of the game (*baseline*, Figure 1b, lower graph), the fish stock regenerated with 0 fish for $x \in \{0,1,\ldots,4\} \cup \{45,46,\ldots,50\}$, 5 fish for $x \in \{5,6,\ldots,19\} \cup \{35,36,\ldots,45\}$), and 10 fish (corresponding to the maximum sustainable yield; MSY) for $x \in \{20,21,\ldots,34\}$) (see Figure 1). In Stage 2, groups were randomly assigned to continuously play *baseline* or one of the following three treatments with a renewed stock of 50 fish: *threshold*, *risk*, and *ambiguity* (16 groups per treatment, 4 per community). *Baseline* groups played the exact same game as in Stage 1 (control). In the three treatments, players were told that a climate event had *definitely* led (*threshold*) or *could* lead (*risk* and *ambiguity*) to a permanent reduction of the reproduction rate to 1 fish for stock levels below 28 (Figure 1b, upper graph). For all other stock sizes the reproduction rate remained the same as in Stage 1. The difference between *risk* and *ambiguity* was in whether the players knew the exact probability (0.5 for *risk*) of the event occurring in each round or only the probability range (0.1-0.9 for *ambiguity*). Fish that remained in the common pool at the end of either stage had no monetary value for the players. The game builds on a lab experiment originally designed by Lindahl et al.[21], and further developed by Schill et al.[15]. We modified the original design for the field and to test the research question at hand.

All decisions were made on individual protocol sheets, collected by an assistant at the end of each round. A fifth experimenter calculated the new stock size for all groups and provided this information to the assistants on a note, which was handed over to the experimenters of each group. The experimenter informed the group about the new stock size at the beginning of each round, which allowed players to deduce information on group members' decisions, for example, whether everyone followed the group agreement in case there was one. Communication was allowed, limited to two minutes before participants made their decision. Each caught fish was worth COP 500 (about € 0.15). The participants had full information about the relationship between the stock size and its reproduction.

The game entailed two stages. All groups played both stages. Stage 1 lasted six rounds and Stage 2 ten rounds, unless a group depleted the stock earlier at which point the game was stopped. The number of rounds was unknown to the players, to avoid the end-of-game-effect. Apart from the changes in the stock dynamics, conditions and rules remained unchanged. In the *risk* and *ambiguity* treatment, players were also told that the climate event is unpredictable; nobody knows whether and when it will happen, including the experimenters and assistants. They knew, however, that the probability of the event occurring in future rounds was constant (see *SI2.3*). Each treatment was run four times in each location.

Experimenters and assistants collected observational data during the game following a structured form on aspects such as whether or not players communicated and what kind of agreements they made, if any. After the game, each player took part in a risk and ambiguity attitude elicitation task[29] (see *SI2.4*) and was interviewed following a structured form (see *SI2.5*). In the meantime, the number of fish caught by each player was calculated, and the respective amount in cash was paid out privately.

In addition to a show-up fee of COP 15,000 (about € 4.20), average individual earnings were COP 19,514 (about € 5.50) with a minimum of COP 3,000 and a maximum of COP 40,000, paid privately at the end of each session. Total average earnings were higher than the median income our participants make on a normal day (see *Table S1*). For comparison, the daily wage in 2016 in this region was about COP 23,000.

Acknowledgements

We would like to thank the fisher communities that participated in our study. The field work would not have been possible without the support of Lina María Saavedra-Díaz, Alisson Soche Forero, Nidia Vanegas Perez, as well as Darlin Botto Barrios, María de los Ángeles Gonzáles Pabón, Jaime González Cueto, Jesús Jiménez Torres, Gloria de León Martínez, and Cristhian Marrugo Marmolejo. In the design and planning phase we received valuable feedback from Therese Lindahl, Anne-Sophie Crépin, Rocio Moreno-Sanchez, Jorge Higinio Maldonado, Juan Camilo Cárdenas, and María Alejandra Vélez. Daniel Ospina, Matías Piaggio, Chandra Kiran, Diego Galafassi, Jean-Baptiste Jouffray, Thomas Hahn, Juan-Camilo Cárdenas, Emily Boyd, Fredrik Carlsson, and Maricela De La Torre Castro provided valuable feedback to earlier drafts. Thank you also to Juan Camilo Cárdenas for providing the instructions for the risk and ambiguity attitudes elicitation task and Oskar Gauffin and Måns Karlsson for statistical consultancy. This research was supported by a project grant from Formas (#211-2013-1120).

References

1. Litzow, M. A., Mueter, F. J. & Hobday, A. J. Reassessing regime shifts in the North Pacific: Incremental climate change and commercial fishing are necessary for explaining decadal-scale biological variability. *Glob. Chang. Biol.* **20**, 38–50 (2014).
2. Cheung, W. W. L., Reygondeau, G. & Frölicher, T. L. Large benefits to marine fisheries of meeting the 1.5°C global warming target. *Science (80-.).* **354**, 1591–1594 (2016).
3. Rocha, J., Yletyinen, J., Biggs, R., Blenckner, T. & Peterson, G. Marine regime shifts: drivers and impacts on ecosystems services. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20130273 (2015).
4. FAO. *The State of World Fisheries and Aquaculture 2016. Contributing to food security and nutrition for all.* (2016).
5. Budescu, D. V., Rapoport, A. & Suleiman, R. Resource dilemmas with environmental uncertainty and asymetric players. *Eur. J. Soc. Psychol.* **20**, 475–487 (1990).
6. Hine, D. W. & Gifford, R. Individual Restraint and Group Efficiency in Commons Dilemmas: The Effects of Two Types of Environmental Uncertainty. *Jounral Appl. Soc. Psychol.* **26**, 993–1009 (1996).
7. Barrett, S. & Dannenberg, A. Climate negotiations under scientific uncertainty. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17372–17376 (2012).
8. Dannenberg, A., Löschel, A., Paolacci, G., Reif, C. & Tavoni, A. On the Provision of Public Goods with Probabilistic and Ambiguous Thresholds. *Environ. Resour. Econ.* **61**, 365–383 (2015).
9. Lenton, T. M. *et al.* Tipping elements in the Earth's climate system. *Proc. Natl. Acad. Sci.* **105**, 1786–93 (2008).
10. Rocha, J. C., Peterson, G. D. & Biggs, R. Regime Shifts in the Anthropocene: Drivers, Risks, and Resilience. *PLoS One* **10**, e0134639 (2015).
11. Biggs, R., Rocha, J. C. & Peterson, G. D. Regime shifts database. *Ecol. Soc.* **in print**, (2018).
12. Scheffer, M. *Critical Transitions in Nature and Society.* (Princeton University Press, 2009).
13. IPCC. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* (Cambridge University Press, 2014).
14. Camerer, C. & Weber, M. Recent Developments in Modeling Preferences: Uncertainty and Ambiguity. *J. Risk Uncertain.* **5**, 325–370 (1992).
15. Schill, C., Lindahl, T. & Crépin, A.-S. Collective action and the risk of ecosystem regime shifts: insights from a laboratory experiment. *Ecol. Soc.* **20**, 48 (2015).
16. Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A. & Marotzke, J. The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proc. Natl. Acad. Sci.* **105**, 2291–2294 (2008).
17. Barrett, S. & Dannenberg, A. Sensitivity of collective action to uncertainty about climate tipping points. *Nat. Clim. Chang.* **4**, 36–39 (2014).
18. Rapoport, A., Budescu, D. V., Suleiman, R. & Weg, E. Social dilemmas with uniformly distributed resources. in *Social dilemmas: Theoretical issues and research findings* (eds. Liebrand, W. B. G., Messick, D. M. & Wilke, H. A. M.) 43–57 (Pergamon Press, 1992).
19. Gustafsson, M., Biel, A. & Gärling, T. Overharvesting of resources of unknown size. *Acta Psychol. (Amst).* **103**, 47–64 (1999).
20. Tavoni, A., Dannenberg, A., Kallis, G. & Löschel, A. Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc. Natl. Acad. Sci.* **108**, 11825–11829 (2011).
21. Lindahl, T., Crépin, A.-S. & Schill, C. Potential Disasters can Turn the Tragedy into Success. *Environ. Resour. Econ.* **65**, 657–676 (2016).
22. Henrich, J., Heine, S. J. & Norenzayan, A. Most people are not WEIRD. *Nature* **466**, 29 (2010).
23. Prediger, S., Vollan, B. & Frölich, M. The impact of culture and ecology on cooperation in a common-pool resource experiment. *Ecol. Econ.* **70**, 1599–1608 (2011).
24. Hoff, K. & Stiglitz, J. E. Striving for balance in economics: Towards a theory of the social determination of behavior. *J. Econ. Behav. Organ.* **126**, 25–57 (2016).
25. Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J. & Sutter, M. Conditional cooperation on three continents. *Econ. Lett.* **101**, 175–178 (2008).
26. Gintis, H., Bowles, S., Boyd, R. & Fehr, E. *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life.* (MIT Press, 2005).
27. Diaz, R. J. & Rosenberg, R. Spreading Dead Zones and Consequences for Marine Ecosystems. *Science (80-.).* **321**, 926–929 (2008).
28. Scheffer, M., Nes, E. H. Van, Holmgren, M. & Hughes, T. Pulse-Driven Loss of Top -Dow Control: The Critical-Rate Hypothesis. *Ecosystems* **11**, 226–237 (2008).
29. Cardenas, J. C. & Carpenter, J. Risk attitudes and economic well-being in Latin America. *J. Dev. Econ.* **103**, 52–61 (2013).

Supplementary Information (SI)

Table of Contents

1. Communities

We ran the experiment in four fishing communities along the Colombian Caribbean coast, see Fig. S1. All communities have been exposed to over-exploitation and marine regime shifts such as fish and mangrove die-offs partially driven by climate change, but to different degrees. All communities rely on fishing as the primary source of livelihood but they differ with regard to where they concentrate their fishing on (Fig. S1): open sea (A), large river delta (B), or within an estuarine lagoon (C and D). This translates to different ecological dynamics and disturbances that the fishers face in everyday life, as well as heterogeneity in fishing styles. Table S1 provides descriptive statistics on a set of characteristics of the participants of each community, derived from post-experimental surveys (see SI2.5).

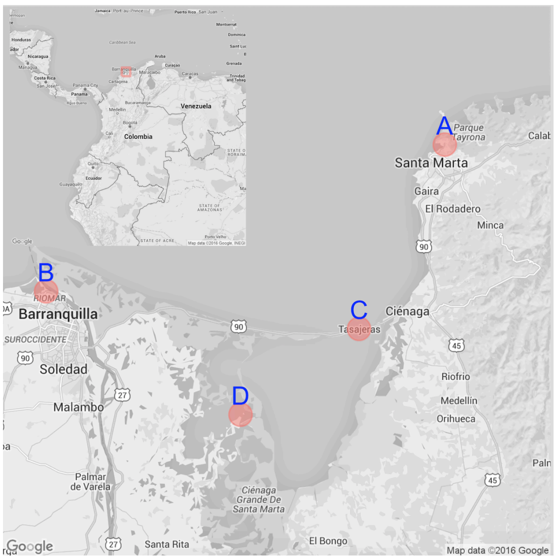

Fig S1. Field work area. Map indicating the location of the four communities along the Colombian Caribbean coast where we conducted the field experiments in February 2016. The labels A-D indicate the order in which we visited the communities, see SI2 for details.

Table S1. Descriptive statistics of characteristics of participants for each community: mean (standard deviation); range (min-max); N. See 2.5 for information on post-experimental survey on which this information is based.

	A	B	C	D	All	Kruskal-Wallis test
Age	32.6 (14.4); 16-74; N=62	53 (12.6); 21-75; N=62	42.2 (14.8); 19-83; N=63	43.3 (13.6); 19-72; N=64	42.8 (15.6); 16-83; N=251	$\chi^2 = 55.7$ $P = 0.0001^{a}$
Formal education[1]	3.1 (0.7); 2-4; N=61	2.1 (0.6); 1-3; N=57	2.5 (0.7); 1-4; N=55	1.9 (0.6); 1-3; N=56	2.4 (0.8); 1-4; N=229	$\chi^2 = 74.5$ $P = 0.0001^{b}$
% of women	5%	8%	0%	0%	3%	$\chi^2 = 9.4$ $P = 0.0244$
Median income (COP) normal day	30,000; N=60	40,000; N=62	25,000; N=64	20,000; N=63	30,000; N=248	$\chi^2 = 30.8$ $P = 0.0001$
Gini coefficient income normal day	0.49; N=63	0.44; N=63	0.44; N=64	0.34; N=64	0.43; N=254	$\chi^2 = 253.0$ $P = 0.0001$
Median income (COP) bad day[2]	0; N=60	5,000; N=62	3,250; N=62	2,000; N=63	2,000; N=247	$\chi = 20.3$ $P = 0.0001$
Median income (COP) good day	100,000; N=60	106,500; N=62	85,000; N=63	50,000; N=63	90,000; N=248	$\chi^2 = 32.2$ $P = 0.0001$
% fishers that fishes in groups	98%; N=62	77%; N=62	82%; N=64	62%; N=63	80%; N=251	$\chi^2 = 26.5$ $P = 0.0001$
Median group size fishing	7; N=61	5; N=48	3; N=52	2; N=38	4; N=199	$\chi^2 = 126.3$ $P = 0.0001$
% fishers that goes with same crew	90%; N=61	98%; N=47	83%; N=53	89%; N=38	90%; N=199	$P > 0.1$
Experienced dramatic change[3]	0.69 (0.47); N=61	0.77 (0.42); N=61	0.83 (0.38); N=64	0.86 (0.35); N=64	0.79 (0.41); N=250	$\chi^2 = 6.3$ $P = 0.0993$
Expect dramatic change[4]	0.88 (0.34); N=24	0.87 (0.36); N=30	0.64 (0.48); N=56	0.82 (0.38); N=57	0.78 (0.41); N=167	$\chi^2 = 9.3$ $P = 0.0259$

Note, one individual each in A and B were excluded for these summary statistics, as it turned out after they have played the game that they are not small-scale fishers. They earn on a normal day more than COP 400.000.

[1] Categorical variable: 1 = none; 2 = basic school education (5 yrs.); 3 = high school (11 yrs.); 4 = university.

[2] To many fishers a bad day means no catch at all, which is the case for 118 fishers in total. For 75% of those this happens several times a week (mostly in A and D).

[3] Binary variable; survey question: Have you experience at any time a sudden, persistent and dramatic change in the amount of fish? We refer here to something more acute than a change due to seasonality where you have noticed that a species in particular has diminished or disappeared for a long period of time.

[4] Binary variable; survey question: Do you expect other sudden and persistent changes in the abundance of fish or other ecosystem aspects in the future (e.g. mangrove, birds)?

[a] According to a set of pairwise Mann–Whitney–Wilcoxon test, participants from A are significantly younger and participants from B are significantly older than participants from all other communities ($P<0.000$ for all tests).

[b] According to a set of pairwise Mann–Whitney–Wilcoxon test, participants from A have significantly higher educational levels than participants from all other communities ($P<0.000$ for all tests); participants from B have significantly lower educational levels than participants from C ($P=0.0011$); participants from C have significantly higher educational levels than participants from D ($P<0.000$).

2. Experimental design and procedure

2.1 Instructions

Available upon request: caroline.schill@beijer.kva.se.

2.2 Visualisation of resource dynamics

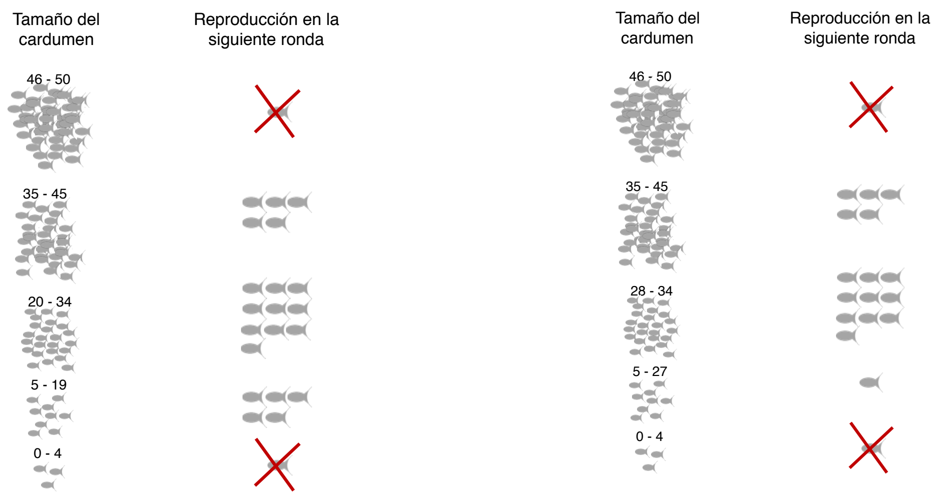

Fig S2. Resource dynamics. Relationship between the size of the fish stock and its reproduction as shown to the participants. (Left) represents the stock dynamics without a threshold (*baseline*). (Right) represents the stock dynamics with a threshold at a stock size of 28. The visualisations for the participants were printed on different coloured sheets to provide a visual difference (*baseline* in green and dynamics with *threshold* in red).

2.3 Visualization and communication of risk and ambiguity treatment

In the *risk* treatment, we showed our participants five green and five red stones. The red stones symbolized the climate event. In front of their eyes, we mixed the stones in an urn and in each round, we drew one stone without revealing its colour. If it was red, the climate event occurred in that round, leading to a change of the reproduction rate below a stock size of 28 until the end of the game (as the event is long-lasting). If the stone was green, the reproduction rate stayed the same (as in *baseline*). No matter the colour of the stone, we continued drawing in each round.

In the *ambiguity* treatment, we followed the same procedure but instead of starting off with 5 stones of both colours, we showed them 10 green and 10 red stones. In front of their eyes, we put 1 red and 1 green stone into the urn. We then mixed the remaining 18 stones in a second urn and added 8 random stones to the first urn without revealing the stones' colour neither to the participants nor the experimenter and assistants. This was achieved by using a non-translucent fabric cloth, so neither experimenters nor participants could see the colours of the stones being moved from one urn to the other.

With this transparent approach to communicating and visualizing the uncertainties involved in the game, we made clear that we did not decide on the event occurring ex-ante, i.e., there were no information asymmetries between the players and experimenters to guarantee credibility and to not affect behaviour (Chow and Sarin 2002). The probabilities were constant over time, i.e., each stone that was picked was thrown back into the urn.

2.4 Risk and ambiguity preferences elicitation task

To elicit risk and ambiguity attitudes of our participants, we made use of the monetarily incentivized risk and ambiguity decision tasks designed by Cardenas and Carpenter (2013). We chose these tasks, as they have been implemented and evaluated extensively in the Latin American context, including people with varying levels of numeracy and literacy. Following the original protocol closely, to measure aversion to *risk*, we asked our participants to choose between six binary lotteries in which the chances of receiving a high payoff are the same as receiving a low one (50-50 chance). To measure aversion to *ambiguity*, we asked them to choose between another set of six lotteries in which the chances of receiving a high versus a low payoff were bound between 1/10 and 9/10 but unknown. Hence, while the probability distribution is known in the risk task, in the ambiguity decision task, it is uncertain. The expected payoff structure is for both tasks the same. To be able to control for order effects, we switched the order of the decision tasks. Half of the sample started with the risk task and the other half with the ambiguity task.

For both decision tasks, we showed our participants a loop consisting of six binary lotteries and asked to choose one to play: $13000|$13000; $10000|$19000; $7000|$25000; $4000|$31000; and $0|$38000. Each lottery was represented by an envelope containing 5 (1-9) high and 5 (1-9) low value payoffs for the risk (ambiguity) task. Clock-wise, the expected payoffs, as well as the variance of the payoffs, increased, i.e., the lotteries became "riskier". Only between the two riskiest lotteries ($2000|$36000 and $0|$38000), the expected payoffs were the same, and only the variance continued to increase. Depending on the participant's choice, one can say something about this individual's level of risk/ambiguity aversion. For example, individuals choosing the safe lottery ($13000|$13000) are extremely risk averse.

Participants did this task right after the CPR fishery game and were told that only one of them would have the chance to earn some extra money (lottery) based on either her risk or ambiguity choice. This is common practice in this type of economic experiments or tasks and has been validated in different studies (Hey and Lee 2005).

Results are shown in Table S2 Overall, for both the risk and ambiguity decision task, the median choice was $ 7,000|$ 25,000. Apart from the median, we also calculated the average choice, following Cardenas and Carpenter (2013), by numbering the lotteries clock-wise from one to six starting with the safe bet. The average choice for the risk and ambiguity task was 3.4, which indicates the average to be the $ 7,000|$ 25,000 lottery for the risk and ambiguity task. This reveals that our participants were neither particularly risk/ambiguity averse nor risk/ambiguity seeking.

Table S2. Risk and ambiguity attitude elicitation task outcomes across communities: mean (standard deviation); median; range (min-max); N.

	Location					
	A	**B**	**C**	**D**	**All**	**Kruskal-Wallis test**
Risk choice	3.0 (1.5); 3; 1-6; N=63	3.4 (1.6); 3; 1-6; N=64	3.1 (1.6); 3; 1-6; N=60	3.8 (1.9); 4; 1-6; N=62	3.4 (1.7); 3; 1-6; N=249	$\chi^2_{risk} = 7.806$ $P_{risk} = 0.0502$[a]
Ambiguity choice	3.2 (1.8); 3; 1-6; N=63	3.3 (1.6); 4; 1-6; N=64	3.0 (1.7); 3; 1-6; N=60	4.1 (1.8); 4; 1-6; N=64	3.4 (1.8); 3 1-6; N=251	$\chi^2_{amb} = 13.188$ $P_{amb} = 0.0042$[b]
	Treatment					
	Baseline	**Threshold**	**Risk**	**Ambiguity**	**All**	**Kruskal-Wallis test**
Risk choice	3.3 (1.6); 3; 1-6; N=63	3.2 (1.6); 3; 1-6; N=60	3.5 (1.6); 3; 1-6; N=64	3.3 (1.8); 3; 1-6; N=62	3.4 (1.7); 3; 1-6; N=249	$\chi^2_{riskT} = 1.426$ $P_{riskT} = 0.7$
Ambiguity choice	3.4 (1.9); 3; 1-6; N=63	3.6 (1.7); 3.5; 1-6; N=60	3.4 (1.7); 3; 1-6; N=64	3.3 (1.8); 3; 1-6; N=64	3.4 (1.8); 3; 1-6; N=251	$\chi^2_{ambT} = 0.904$ $P_{ambT} = 0.8$

[a] According to a set of pairwise Mann–Whitney–Wilcoxon test, participants from A are significantly more risk averse than participants from D (*P*=0.0104); and participants from C are significantly more risk averse than participants from D (*P*=0.0498).

[b] According to a set of pairwise Mann–Whitney–Wilcoxon test, participants from A are significantly more ambiguity averse than participants from D (*P*=0.006); and participants from B are significantly more ambiguity averse than participants from D (*P*=0.0106); and participants from C are significantly more ambiguity averse than participants from D (*P*=0.0010).

We did not find any relationship between the earnings of a participant or the total catch of the groups and the risk or ambiguity choice. Overall, we did not find significant correlations between individuals' total catch and choices in the risk and ambiguity task and neither between group total catch and the groups' average choice in the task. Within each treatment, we only found a weak significant negative correlation between individuals' total catch and the risk choice in the ambiguity treatment (Spearman's correlation test, P=0.0851, r_s = -0.22).

2.5 Post-experimental survey

After the fishery game and the risk/ambiguity task (see above), each participant was interviewed following a structured form. We chose structured interviews to account for illiteracy. The interview consisted of five modules: 1) questions about the game, 2) the participants fishing activities, 3) perceptions on environmental/ecosystem change, including more dramatic changes, 4) cooperation and communication, and 5) demographics and household composition. Available upon request.

3. Statistical Analysis

3.1 General information

Preceding each statistical test, we tested whether the given observations were normally distributed with a Shapiro-Wilk test (Shapiro and Wilk 1965). If the test produced positive results at a 5% significance level, we rejected the assumption of normality, and applied non-parametric tests (Kruskal-Wallis test (KW) for comparisons across more than 2 independent samples and Whitney–Wilcoxon rank-sum tests (MWW), for pair-wise comparisons). For KW, we report in the main text degrees of freedom as well as χ^2 and respective p-values. For MWW, we report p-values only. To compare cases of frequency, we used Fisher's exact test (Kanji 1993). To assess the equality of variances, we applied Levene's test.

Statistical tests are based on group averages or medians as units of observation. If not stated otherwise, reported tests are two-sided.

We analysed the data using Stata 14.2 if not specified differently and we used R for the graphics. We used group averages as unit of observation in all statistical tests and all reported tests are two-sided. See Table S3 for an overview of the variables used for the statistical analysis.

Table S3. Overview varibales used for statisitcal analysis.

	Value range	*Description*	*Motivation*
Overall stock size	[0, 50]	Average stock size (after catch, before regeneration) across all rounds per group per Stage	We use average and median stock sizes as units of observation to take into account stock size dependency between rounds.
	[0, 50]	Median stock size (after catch, before regeneration) across all rounds per group per Stage	
Percentage of rounds above the threshold	[0, 1]	Average percentage of rounds groups sustain a stock level of 28 or above in Stage 2	To assess whether high stock sizes are due to groups spending more time above the (potential) threshold to avoid a (potential) regime shift.
Percentage of rounds above stock size of 20	[0, 1]	Average percentage of rounds groups sustain a stock level of 20 or above in either Stage of the game	To assess whether high stock sizes are due to groups trying to avoid stock sizes below 20 (to avoid low regeneration rates).
Percentage of rounds group achieved highest regeneration rate	[0, 1]	Average percentage of rounds groups achieve the highest regeneration rate in either Stage	Proxy for optimal resource use
Depletion	0 ∨ 1	Group depleted resource before the end of the game	
Threshold crossing	0 ∨ 1	Group crosses the threshold at some point during the game	
Gini coefficient	[0, 1]	Gini coefficient of group catch (dispersion of total catch within each group)	Proxy for cooperation

3.2 Treatment effects

3.2.1 Expectations guiding our analysis

In our game, it is reasonable to assume that players make use of the opportunity to communicate to agree on exploitation levels. For groups that played *baseline*, the optimal collective strategy was to sustain a stock level of 20 throughout the game, i.e. catch 30 fish altogether in round one and then catch ten fish in each following round (highest regeneration rate, Figure 1b lower graph), for as long as all participants believe that the game continues. For the *threshold*, *risk* and *ambiguity* treatment, the optimal collective strategy was to sustain the stock just above the (potential) threshold (stock level of 28). If groups follow this strategy, we would expect that *threshold*, *risk* and *ambiguity* groups sustain overall higher stock levels compared to *baseline*. Assuming that players correctly assign and interpret probabilities and that they are risk-neutral (see *S12.4*), we would furthermore expect that overall stock levels of the *threshold* groups are higher compared to *risk* or *ambiguity*. Whether overall sustained stock sizes for the *ambiguity* treatment will be higher than for *risk* depends on the actual influence of ambiguity on behaviour, what we do not know. However, players might not take the opportunity to communicate, and cooperation might furthermore not automatically lead to optimal resource use, as it was found in similar lab experiments(Schill et al. 2015, Lindahl et al. 2016). In that case, we would nevertheless expect that groups would avoid stock sizes in the lower regeneration rate area. Hence, we would, overall, still expect higher stock sizes for the *threshold*, *risk* and *ambiguity* treatment, compared to *baseline* as well as higher stock sizes for *threshold*, compared to *risk* and *ambiguity*.

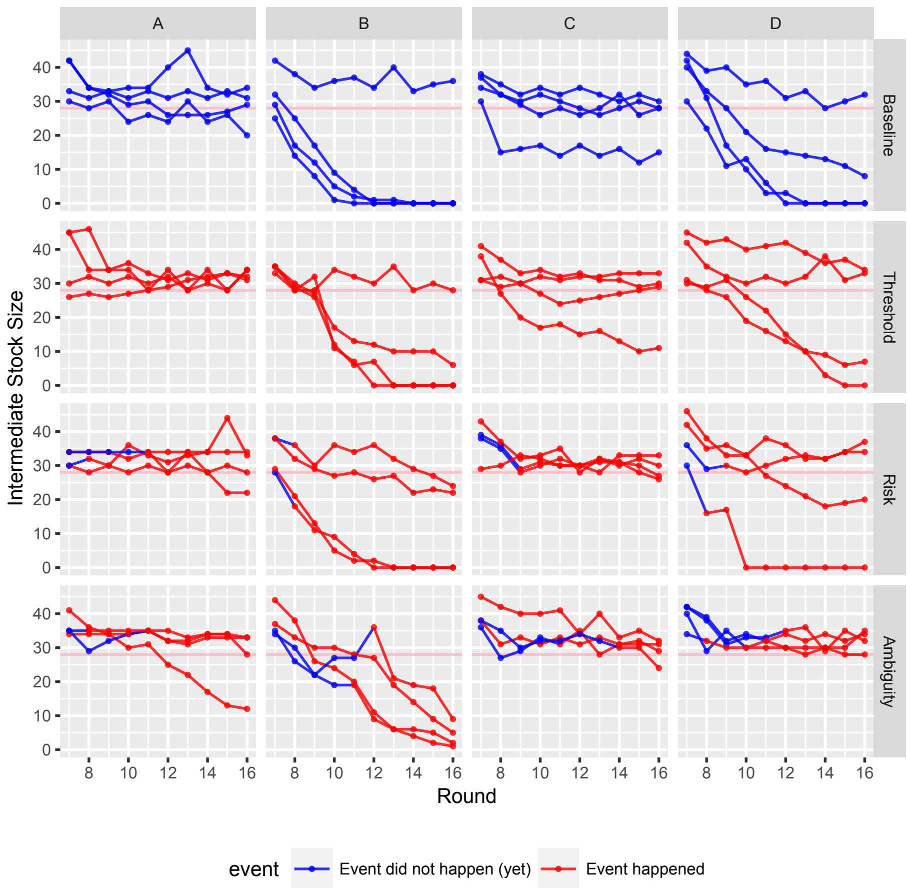

Fig S3. Time series of intermediate stock size (after extraction, before re-generation) for each group across place (A-D) and treatment in Stage 2 of the game. A red data point for any given stock size and round indicates that the climate event has happened and the stock will regenerate in that given round according to the resource dynamics with threshold (see Figure 1b). The pink horizontal line indicates the potential threshold.

Apart from overall sustained stock size, we might also be interested in average catch, as the variable that reflects behaviour directly. However, first, as we measure stock size after catch and before regeneration, our variable of interest (stock size) represents behaviour very well. Second, given our particular experimental design in which the resource dynamics is represented by a discrete logistic growth function (see Fig S1), it is not very informative to look at catch. Lower catch can result from being precautionary (keep stock levels above 34), but also from having a more aggressive exploitation strategy, i.e. sustained stock sizes in the lower resource growth range (below a stock size of 20 in the *baseline*, for example). Furthermore, as the game is played over several rounds, we would not expect to find measurable differences in average catch between the treatments. For example, assuming groups would follow the collective optimal strategy, catch would only differ in the first round (to reduce the stock to 28 rather than 20), but in each following round, the total catch would be 10, according to the MSY. In other words, in our game higher stock sizes can be sustained without compromising catches.

3.1.2 Pair-wise non-parametric hypothesis tests

Table S4. Results of pair-wise Mann–Whitney–Wilcoxon rank-sum tests comparing median stock sizes across treatments.

	Baseline	***Threshold***	***Risk***	***Ambiguity***
Baseline		$P = 0.365$	$P = 0.326$	$P = 0.0822$
Threshold			$P = 0.623$	$P = 0.199$
Risk				$P = 0.509$

Note: there are no significant differences with regard to median stock size across all treatments (Kruskal-Wallis test, df=3; χ^2=3.55, P=0.312). Differences between *baseline* and *ambiguity* are not robust to multiple comparison corrections, using Dunn's test of multiple comparisons using rank sums.

Table S5. Results of pair-wise Mann–Whitney–Wilcoxon rank-sum tests comparing variables of Table S3.

Baseline compared to	*Threshold*	*Risk*
Average stock size	P=0.396	P=0.451
Median stock size	P=0.365	P=0.326
Variance of average stock size	$W0$=2.43; P=0.13	$W0$=2.156; P=0.152
Percentage of rounds above stock level of 20	P=0.308	P=0.274
Percentage of rounds above stock level of 28	P=0.256	P=0.22
Percentage of rounds with highest regeneration rate	P=0.864	P=0.731
Depletion cases	P=0.685	P=0.685

Table S6. Results of pair-wise Mann–Whitney–Wilcoxon rank-sum tests comparing variables of Table S3.

	Threshold compared to *Risk*	*Threshold* compared to *Ambiguity*	*Risk* compared to *Ambiguity*
Average stock size	P=0.777	P=0.266	P=0.584
Median stock size	P=0.623	P=0.199	P=0.509
Variance of average stock size	W0=0.107; P=0.746	W0=5.055; P=0.032	W0=1.567; P=0.22
Percentage of rounds above stock level of 20	P=0.646	P=0.396	P=0.887
Percentage of rounds above stock level of 28	P=0.952	P=0.498	P=0.431
Percentage of rounds with highest regeneration rate	P=0.555	P=0.506	P=0.894
Depletion cases	P=1.000	P=0.226	P=0.226

Whether threshold was crossed at some point during the game	P=1.000	P=1.000	P=0.724

Table S7. Results of pair-wise Whitney–Wilcoxon rank-sum tests across communities in Stage 1.

Median stock size	χ^2=5.417; P=0.144
Variance of average stock size	W0=4.46; P=0.007
Percentage of rounds above stock level of 28	χ^2=7.21; P=0.066
Percentage of rounds with highest regeneration rate	χ^2=11.023; P=0.012
Depletion cases	P=0.181

Table S8. Results of pair-wise Whitney–Wilcoxon rank-sum tests across communities in Stage 2.

Median stock size	χ^2=15.674; P=0.001
Variance of average stock size	W0=12.297; P<0.0001
Percentage of rounds above stock level of 28	χ^2=15.524; P=0.0014
Percentage of rounds with highest regeneration rate	χ^2=24.617; P=0.0001
Depletion cases	P=0.001

3.1.3 Threshold crossing

Cases of crossing the (potential) threshold: Some groups managed to reverse the shift (Fig S2). In the *threshold* treatment, 2/8 groups that crossed the threshold reversed the shift. In the *risk* treatment, 1/9 groups that crossed the threshold reversed the shift (the climate event had happened in all nine instances). In the *ambiguity* treatment, for 3/7 groups that crossed the potential threshold the climate event had not yet occurred, with 2/3 of these groups recovering the stock. For the 4/7 groups in which the climate event did happen none recovered the stock (see Fig S3). In the *baseline* treatment, no single group recovered the higher regeneration rate once below a stock size 20.

3.2 Community differences

3.2.1 *Hypotheses for explaining community differences*

We found pronounced differences in exploitation patterns between communities: groups that use the fishery in the game more sustainably often have higher educational levels, come from communities that rely on fishing styles that require strong coordination efforts, and in which feedbacks between fishing efforts and stock size in reality are less masked. Here we present a set of hypotheses to explain the observed community differences.

Education levels and age could influence the game outcomes. In fact, whereas average education levels of participants from community A and C are high and average age is low, participants from community B have significantly lower education levels and are significantly older on average (see Table S1). This could explain why in comparison to community B, community A and C sustain on average higher stock sizes. Age and education could also influence group Gini coefficients. Higher education levels can increase individuals' understanding about the optimal exploitation level, as well as their propensity to make use of communication for discussing and setting collective exploitation levels. We find a positive relationship between the number of rounds in which groups achieved the MSY and the average education levels in the group, as well as a negative relationship between Gini coefficients and average education (see SI 3.3).

Another explanation for why community A did so well overall could be found in the type of *fishing style* they practice. Survey data reveals that they all fish in groups, which are not only typically larger (see Table S2), but also have a particular fishing style ('Chinchorro'; beach seine net with bag) which requires a high level of group coordination. This could carry over to the game and positively affect the likelihood of players working together. Previous experimental studies showed that direct payoff of cooperation in real life can explain game outcomes(Henrich et al. 2005, Prediger et al. 2011, Gneezy et al. 2016).

The significant differences between community A and B in relation to average stock levels, cases of exploitation beyond the potential threshold, and depletion could also be explained by *differences of ecological conditions the fishers face in their everyday life*. Located in proximity to a major port and along the Magdalena River, Colombia's main river, community B is exposed to high rates of pulse disturbances from the shipping industry, harbour facilities, and sediments and pollutants coming from the river. In particular, large projects for maintenance of the harbour facilities in the past, such as dredging and embankment construction, have heavily affected its local ecosystem. As a consequence, the impact of fishing on stock dynamics may be entirely masked, or at least less obvious, in comparison to the other communities. In contrast, in community A fishers are likely to benefit from the proximity of a marine protected area, which can buffer climatic and anthropogenic disturbances on fish stocks. The magnitude and frequency of exogenous disturbances that fishers from community B experience might affect the players' planning horizon resulting in less sustainable exploitation patterns in order to secure catch now rather than risk losing it later. This aligns with theoretical work on optimal management in the face of exogenously driven regime shifts (Polasky et al. 2011) as well as evidence from CPR games testing the effect of exogenous degradation on use of a shared resource (Blanco et al. 2015) in the same region.

Communities C and D fish mostly inside the lagoon complex that has been historically exposed to regime shift dynamics, such as hypoxic events, fish and shrimp collapses, and mangrove forest die-offs (Vilardy et al. 2011). *Feedbacks between fishing effort and stock levels* might be stronger in comparison to community B, which is also due to the enclosed fishing zone as well as less affected by dynamics of the Magdalena River. Both communities are located in the river's floodplain. However, community C is the least affected, which could explain the significant difference in average stock levels and cases of depletion between both communities. Indeed, previous experimental evidence on endogenously driven regime shifts suggests that negative feedbacks between collective action for sustainable resource use and ecosystem degradation in reality could explain CPR game outcomes (Prediger et al. 2011). These results stem from a grazing game played with farmers from pastoral systems where bare soil patches are strong indicators of resource degradation. While degradation indicators in fisheries are not observable to the same extent, events of fish die-offs, which community C and D experienced in the past, could serve as such.

3.2.2 *Difference between Stage 1 and Stage 2 within communities*

Pair-wise comparisons within communities between Stage 1 and Stage 2 show no significant differences for median stock sizes (MWW, P=0.860). However, there are significant differences with respect to how often groups achieve the highest regeneration rate (MWW, P=0.0033). It appears that groups that did well in the sense of achieving high regeneration rates in Stage 1, did even better in Stage 2. In community A and C, in Stage 2, groups achieve significantly more often a higher regeneration rate compared to Stage 1 (MWW, P_A=0.0001; P_C=0.0036). However, we cannot say whether this improvement was due to the treatment or because of learning effects due to a limited sample size.

3.3 Equal sharing (cooperation)

Gini coefficient as proxy for cooperation

We use the Gini coefficient of total group catch as proxy for cooperation, measuring the dispersion of total group catch within each group. Low Gini coefficients imply that catches and respective earnings are distributed equally among group members. In order to reach low values, it is reasonable to assume that players coordinated their play by making use of the opportunity to communicate to formulate and agree on exploitation levels that everyone followed. Only 3/64 did not communicate at all according to the observational data that the experimenters and the assistants collected. This assumption is supported by the fact that there is a significant positive correlation between average Gini coefficient per round in Stage 1/Stage 2 and overall Gini coefficient in Stage 1/Stage 2 (*Pearson's* r_{Stage1} = 0.89, P < 0.0001; *Pearson's* r_{Stage2} = 0.82, P < 0.0001; Fig S4).

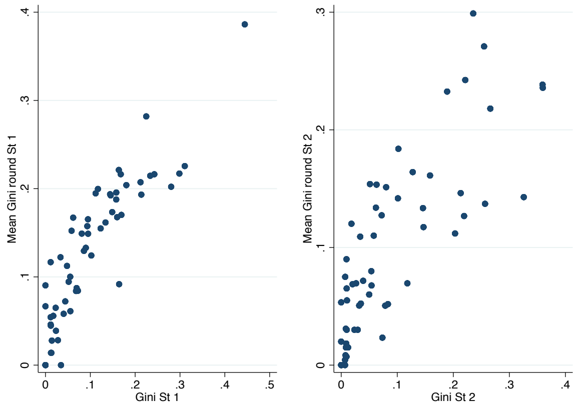

Fig S4. Scatterplot of Gini coefficients averaged across rounds for each group and stage and the overall Gini coefficient measured at the end of either stage; Stage 1 (left); Stage 2 (right).

Some notable patterns regarding the Gini coefficient

- Gini coefficients were overall quite low; in fact, 9 groups in Stage 1 and 11 groups in Stage 2 had a Gini coefficient = 0. However, we also have 7 groups in Stage 1 and 13 groups in Stage 2 in which only one or two players had one single fish more than the other group members. 16/64 groups in Stage 1 and 24/64 groups in Stage 2 achieved a Gini coefficient < 0.015 for the entire stage. A Gini < 0.015 includes group that shared the total harvest equally not within rounds but across rounds. E.g., dividing up the MSY of 10 between 4 does not work evenly, therefore, groups might split it the following: 2 players take 3 and 2 players take out 2. Doing that has the intention to share catch equally but as they do not know when the game ends, they can end up with this slight unequal catch in the end anyway.
- There are no significant changes in Gini going from Stage 1 to Stage 2, neither by using treatments (Kruskal-Wallis, $df = 3$, $\chi^2 = 1.304$, $P = 0.728$) or location (Kruskal-Wallis, $df = 3$, $\chi^2 = 6.235$, $P = 0.1007$) as grouping variable, with the difference in Gini between both stages as response variable.
- There is a negative significant correlation between average education level in the group and overall Gini coefficient measured at the end of Stage 1 (*Pearson's* $r = -0.277$, $P < 0.0001$).
- There is a positive significant correlation between average age in the group and overall Gini coefficient measured at the end of Stage 1 (*Pearson's* $r = 0.298$, $P < 0.0001$).
- There is a positive significant correlation between average education level in the group and number of rounds in which a group achieved the MSY (*Pearson's* $r = 0.53$, $P = 0.000$).

3.4 Regression Models

We regard Stage 1 of the game as 'training stage'. This way, we could also gather observations unaffected by treatment. These observations could reduce noise in the regression model and potentially let us see the treatment effects more clearly, as they allow us to take group differences (independent of treatment) into account. We include median stock level from Stage 1 in the model to account for group differences, as there is a seemingly linear relationship between the average stock level in Stage 1 and 2.

To account for community and participant differences, we include as well the factor variable community in the model. We also included in some models the Gini coefficient of total group catch of Stage 1 ('Gini Stage 1') as a proxy for cooperation (see section 3.2 below). We use the Gini coefficient of Stage 1 rather than Stage 2, because the Gini coefficient from Stage 2 might be affected by the treatments, Gini Stage 1 and Gini Stage 2 are highly positively correlated (Pearson $r = 0.8$, $P<0.0001$). Since 'Gini Stage 1' and median stock level Stage 1 are clearly negatively correlated (Pearson's $r = -0.54$, $P<0.0001$), we should not be surprised if only one of these Stage 1 covariates will turn out to be significant. Table S9 shows the result of three different model variations. Model 2 has the lowest AIC and the one reported in the paper.

We also fitted a model with lowest order interactions between our explanatory variables. An F-test showed no significant difference between the model with and without interactions; thus, we concluded that the model with the whole dataset, without interactions should be used, since we found no support of reasonable interaction effects occurring.

Table S9. Regression models with median stock size of Stage 2 as response variable.

	Model 1	**Model 2**	**Model 3**
Threshold	2.567	2.223	4.863
	(0.378)	(0.414)	(0.200)
Risk	6.296^{+}	5.940^{+}	6.800^{+}
	(0.066)	(0.067)	(0.082)
Ambiguity	7.783^{**}	7.575^{**}	9.541^{**}
	(0.005)	(0.005)	(0.004)
B	-10.10^{**}	-10.22^{**}	-14.24^{***}
	(0.005)	(0.004)	(0.000)
C	-1.298	-1.198	-2.058
	(0.448)	(0.463)	(0.279)
D	-6.834^{**}	-6.907^{**}	-7.178^{*}
	(0.007)	(0.005)	(0.023)
Median Stock St 1	0.620^{**}	0.685^{***}	
	(0.001)	(0.000)	
Gini St 1	-12.35		-41.05^{**}
	(0.397)		(0.004)
Constant	10.67^{+}	8.063^{*}	29.48^{***}
	(0.077)	(0.048)	(0.000)
R^2	0.663	0.656	0.508
Adjusted R^2	0.614	0.613	0.447
AIC	442.0	**441.3**	464.1
Observations	64	64	64

p-values in parentheses
Median Stock St1 and Gini St 1 are negatively correlated; Pearson's r = -0.54.
$^{+} p < 0.10$, $^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$
Model diagnostics are available upon request. We tested the model regarding the assumptions of normally distributed residuals and heteroscedasticity.
To account for heteroscedasticity, we use the so called HC3-estimator, a robust sandwich type estimator suggested by Elfron (1982), following the suggestion of Long and Ervin (2000).

Table S9. Regression models with 'percentage of rounds above threshold' (Stage 2) as response variable (*N* = 64).

	Model 1	**Model 2**	**Model 3**	**Model 4**
Threshold	0.0912	0.0686	0.135	0.0763
	(0.448)	(0.549)	(0.294)	(0.535)
Risk	0.222^{*}	0.195^{+}	0.236^{*}	0.204^{+}
	(0.040)	(0.057)	(0.043)	(0.064)
Ambiguity	0.204^{*}	0.187^{+}	0.248^{*}	0.193^{+}

	(0.044)	(0.068)	(0.019)	(0.074)
B	-0.359**	-0.370**	-0.442***	-0.369**
	(0.004)	(0.005)	(0.000)	(0.002)
C	-0.0545	-0.0492	-0.0658	-0.0301
	(0.514)	(0.562)	(0.460)	(0.730)
D	-0.185*	-0.191*	-0.189+	-0.129
	(0.035)	(0.035)	(0.053)	(0.157)
Mean Stock St 1	0.0148*	0.0200***		
	(0.025)	(0.001)		
Gini St 1	-0.874*		-1.500***	-0.985*
	(0.044)		(0.000)	(0.035)
% rnds. stock >= 28 (St 1)				0.302*
				(0.027)
Constant	0.377+	0.175	0.834***	0.600***
	(0.060)	(0.301)	(0.000)	(0.000)
R^2	0.566	0.533	0.497	0.560
Adjusted R^2	0.502	0.474	0.434	0.496
AIC	**19.11**	21.80	26.47	19.96
Observations	64	64	64	64

P-values in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Model diagnostics are available upon request. We tested the model regarding the assumptions of normally distributed residuals and heteroscedasticity. To account for heteroscedasticity, we use the so called HC3-estimator, a robust sandwich type estimator suggested by Elfron (1982), following the suggestion of Long and Ervin (2000).
Note, 'average % of rounds stock >= 28 St 1' and 'Gini St 1' are negatively correlated (Pearson $r = -0.477$, $P = 0.0001$). Table S6 shows the result of four different model variations. Model 1 has the lowest AIC and is the one reported in the paper.

Regression models with threshold treatments only

Table S11. Regression models with median stock size of Stage 2 as response variable, threshold treatments only (*N*=48).

	Model 1	Model 2	Model 3
Risk	3.142	3.241	1.980
	(0.272)	(0.239)	(0.548)
Ambiguity	5.043*	5.295*	4.661+
	(0.026)	(0.014)	(0.060)
B	-10.67**	-10.94**	-12.60**
	(0.005)	(0.004)	(0.002)
C	-0.943	-1.029	-1.005
	(0.642)	(0.543)	(0.671)
D	-4.154+	-4.211+	-4.359
	(0.091)	(0.076)	(0.147)
Median Stock St 1	0.419+	0.562**	
	(0.059)	(0.003)	
Gini St 1	-21.49		-43.08*
	(0.270)		(0.016)
Constant	19.04*	13.12*	33.16***

	(0.014)	(0.011)	(0.000)
***R*²**	0.645	0.618	0.566
Adjusted *R*²	0.583	0.562	0.502
AIC	**325.3**	326.8	333.0
Observations	48	48	48

p-values in parentheses

Median Stock St1 and Gini St 1 are negatively correlated; Pearson's r = -0.54.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Model 1 has the lowest AIC and is the one reported in the paper.

Table S12. Regression models with 'percentage of rounds above threshold' (Stage 2) as response variable, threshold treatment only (*N*=48).

	Model 1	Model 2	Model 3	Model 4
Risk	0.112	0.112	0.0983	0.110
	(0.316)	(0.311)	(0.375)	(0.322)
Ambiguity	0.114	0.120	0.113	0.115
	(0.265)	(0.255)	(0.265)	(0.262)
B	-0.440***	-0.456**	-0.466***	-0.440***
	(0.000)	(0.002)	(0.000)	(0.000)
C	-0.0492	-0.0561	-0.0478	-0.0441
	(0.622)	(0.554)	(0.640)	(0.661)
D	-0.144	-0.149	-0.146	-0.122
	(0.132)	(0.154)	(0.145)	(0.230)
Mean Stock St 1	0.00697	0.0149*		
	(0.262)	(0.025)		
Gini St 1	-1.064*		-1.387**	-1.136^{+}
	(0.036)		(0.002)	(0.051)
% rnds. stock >= 28 (St 1)				0.135
				(0.317)
Constant	0.716**	0.398^{+}	0.949***	0.822***
	(0.004)	(0.063)	(0.000)	(0.000)
***R*²**	0.567	0.512	0.553	0.565
Adjusted *R*²	0.492	0.440	0.487	0.489
AIC	11.38	15.20	**11.00**	11.61
Observations	48	48	48	48

p-values in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Model 3 has the lowest AIC and is the one reported in the paper.

References (for SI only)

Blanco, E., M. C. Lopez, and S. Villamayor-Tomas. 2015. Exogenous degradation in the commons: Field experimental evidence. *Ecological Economics* 120:430–439.

Cardenas, J. C., and J. Carpenter. 2013. Risk attitudes and economic well-being in Latin America. *Journal of Development Economics* 103:52–61.

Chow, C. C., and R. K. Sarin. 2002. Known, unknown, and unknowable uncertainties. *Theory and Decision* 52:127–138.

Elfron, B. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Gneezy, U., A. Leibbrandt, and J. A. List. 2016. Ode to the Sea: Workplace Organizations and Norms of Cooperation. *The Economic Journal* 126(595):1856–1883.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, N.

S. Henrich, K. Hill, F. Gil-White, M. Gurven, F. W. Marlowe, J. Q. Patton, and D. Tracer. 2005. "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *The Behavioral and Brain Sciences* 28(6):795–855.

Hey, J. D., and J. Lee. 2005. Do Subjects Separate (or Are They Sophisticated)? *Experimental Economics* 8(3):233–265.

Kanji, G. K. 1993. *100 Statistical Tests*. SAGE Publications, London, UK.

Lindahl, T., A.-S. Crépin, and C. Schill. 2016. Potential Disasters can Turn the Tragedy into Success. *Environmental and Resource Economics* 65(3):657–676.

Long, J. S., and L. H. Ervin. 2000. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician* 54(3):217–224.

Polasky, S., A. de Zeeuw, and F. Wagener. 2011. Optimal management with potential regime shifts. *Journal of Environmental Economics and Management* 62(2):229–240.

Prediger, S., B. Vollan, and M. Frölich. 2011. The impact of culture and ecology on cooperation in a common-pool resource experiment. *Ecological Economics* 70(9):1599–1608.

Schill, C., T. Lindahl, and A.-S. Crépin. 2015. Collective action and the risk of ecosystem regime shifts: insights from a laboratory experiment. *Ecology and Society* 20(1):48.

Shapiro, S. S., and M. B. Wilk. 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* 52(3/4):591–611.

Vilardy, S. P., J. A. González, B. Martín-López, and C. Montes. 2011. Relationships between hydrological regime and ecosystem services supply in a Caribbean coastal wetland: a social-ecological approach. *Hydrological Sciences Journal* 56(8):1423–1435.