Collective action and individual adaptation in natural resource management under the threat of ecosystem change: Insights from economic experiments

Inauguraldissertation
zur Erlangung des akademischen Grades eines Doktors
der Wirtschaftswissenschaften des Fachbereichs Wirtschaftswissenschaften
der Universität Osnabrück

vorgelegt von

Katharina Hembach-Stunden

Osnabrück, 09/2022

Dekan: Prof. Frank Westermann, Ph.D.

Referenten: Prof. Dr. Stefanie Engel

Prof. Dr. Achim Schlüter

Tag der Disputation: 21.09.2022

Supervised by:

STEFANIE ENGEL

(Universität Osnabrück)

Acknowledgements

Firstly, I would like to thank my supervisor Stefanie Engel for supporting me and giving me the freedom to pursue my own research ideas, creating a memorable experience for a PhD. Her guidance has been instrumental in facing the challenges, small and large, that came my way.

Next, I would like to thank my two daily supervisors and co-authors, Aneeque Javaid and Tobias Vorlaufer for taking the time to discuss my research and giving valuable feedback on preliminary versions of this PhD thesis. I would also like to thank Achim Schlüter for providing feedback on the three papers of this thesis and for being my second supervisor.

I would like to acknowledge the Ministry of Science and Culture of Lower Saxony and the Alexander von Humboldt-Foundation, in the framework of the Alexander von Humboldt-Professorship endowed by the German Federal Ministry of Education and Research for their financial support. Both of these funding bodies had no direct influence on the research design or the data collection and analysis strategy.

My special thanks goes to Steffen Beverborg, Fabian Thomas, Dominik Kohl and Imke Lüdeke who provided their support and assistance during the implementation of the laboratory experiment discussed in Chapter 2 in the Laboratory for Economics Research (LaER) at Osnabrueck University. Furthermore, I wish to thank Olaf Bock and his team of the economic laboratory at Hamburg University for their assistance with the implementation of the second experiment discussed in Chapter 3.

I further thank all members of the BEE group for the valuable exchange, fruitful discussions and supporting conversations. Here I mention Adriana Bernal-Escobar, Philipp Gorris, Nick Heinz, Fabian Heitmann, Ann-Kathrin Kößler, Abel-Gautier Kouakou, Bosco Lliso, Juan Felipe Ortiz-Riomalo, Laura Papendorf, Fabian Thomas, Eva-Maria Tolzmann and Caroline van Bers. Furthermore, I thank all members of the Institute of Environmental Systems Research for providing an inspiring workplace and especially my roommate Caroline Lumosi. My special thanks also goes to Matthew Adamson, Frank Hilker and Anthony Sun for their helpful explanations and discussions of ecological dynamics and regime shifts.

I would also like to thank Julia Flesch, Christina Kamm, Linda Laflör and Franziska Meergans for their motivational support during the challenging times of the Covid19-pandemic, which forced us all into the isolation of our homes.

For their continued support, love and care, I thank my family, Dagmar, Fritz, Robert and Karen. I also thank my extended family Carol, Robert, Heather and James for the support from afar. And finally, but most critically: Thank you, James, for your support and love over the last years. Thank you, Nora, for making me laugh with your big smiles. I am grateful for too many things to list here, so I just say thanks for putting up with me and for your help to get this thesis finished.

Summary

Scientific evidence shows that climate change increases the frequency of climate extremes across the globe. These climate extremes exogenously pressure local resource users by causing destruction of natural resources, often affecting ecosystems that already have deteriorated due to overexploitation in the past. The future state of natural resources and entire ecosystems is thereby determined by both exogenous (climate) and endogenous (management by resource users) dynamics. The combination of both the uncertain changing environmental conditions and manmade overexploitation will make the sustainable management of natural resources by local resource users more challenging in the future.

Depending on the underlying ecosystem dynamics, the combination of overexploitation and climate extremes may cause sudden abrupt shifts in natural resources if a resource is driven to its critical threshold (tipping point). These shifts are termed regime shifts. In its most drastic form, a regime shift results in the collapse of the resource with severe economic consequences. Ecological and meteorological warning and forecast systems could potentially warn of approaching regime shifts and climate extremes, thereby motivating the resource users for more sustainable resource management and investments in protective adaptation.

Self-governance of natural resources highly depends on collective action. Resource users need to cooperate and coordinate their resource extraction strategies to keep a resource at a sustainable level of regrowth and to prevent it reaching a critical threshold. Policy makers and ecologists must decide when and how to inform local resource users about the potential threat of crossing critical thresholds. However, critical thresholds are often unknown and ecological early warning signals only provide uncertain threshold knowledge. Knowing if the communication of early imprecise threshold information bears a risk to hamper collective action is thus critical. In addition, in some cases, individual adaptation behaviour determines how far an individual experiences economic losses due to climate extremes. In these cases, the issue is not about collective action, but rather about individuals’ responsiveness to early warnings.

To further understand human behaviour in the light of the aforementioned ecological dynamics, three economic experiments were designed and implemented. The results of these experiments are presented in the three academic papers of this thesis (Chapter 2 to 4):

The first paper, titled *“The interaction of shock experience and threshold knowledge in natural resource management”*, was co-authored by myself with Aneeque Javaid and Stefanie Engel (Chapter 2). This paper addresses the lack of evidence in the literature on the impact of the interaction of exogenously and endogenously driven change in ecosystems on collective action. To analyse this interaction and find the main driver of change in groups’ resource extraction strategies, a novel, (quasi-) continuous-time common-pool resource (CPR) experiment was designed and implemented in the laboratory. The CPR experiment incorporates both dynamics: an unexpected exogenous shock that causes resource scarcity and a critical threshold, at which the resource collapses. The impact of initial resource scarcity on groups’ extraction behaviour is compared to the impact of shock driven scarcity. Furthermore, the effect of shock experience on extraction strategies in the future is assessed. The results indicate that while group members cooperate less when experiencing an exogenous shock to their resource, the knowledge of a critical threshold still motivates successful coordination. However, cooperation amongst group members and efficiency of resource extraction is more sensitive to the resource scarcity itself, than the experience of an exogenous shock. There is no significant effect of shock experience on group’s future extraction strategies.

The second paper, *“Are imprecise early warnings a potential benefit or threat to sustainable resource management?”*, was co-authored by myself with Tobias Vorlaufer and Stefanie Engel (Chapter 3). This paper asks whether an imprecise early threshold warning alters cooperation amongst resource users and analyses if there is a danger of deteriorating individuals’ responses to the certain threshold knowledge by giving an imprecise early warning. On the one hand, imprecise early warnings could raise awareness about the resource’s dynamics and thus, encourage collective action. On the other hand, imprecise early warnings could be taken as a sign of an inevitable upcoming loss of the resource. Thus, resource users increase their individual extraction efforts and collective action fails. To assess the effect of imprecise early warnings on collective action, two additional treatments for the CPR experiment were designed and implemented. The two treatments differ in the degree of uncertainty about the threshold level in the beginning. While groups in both treatments know of the critical threshold, only one treatment receives an imprecise early threshold warning in form of a known threshold range. The experimental results show no effect of such an imprecise early warning on cooperation and coordination amongst group members in comparison to groups who only know of the mere threshold existence.

The third paper, *"False and missed alarms in seasonal forecasts affect individual adaptation choices"*, was again co-authored by myself with Tobias Vorlaufer and Stefanie Engel (Chapter 4).[1] It analyses the effect of varying forecast accuracy on individuals' responsiveness to climate forecast systems. Climate extremes can result in economic losses if individuals are not adequately prepared. The effect of climate forecast systems however, likely depends on their accuracy and individuals' responsiveness to inaccurate climate warnings. An online experiment was designed and implemented to assess individuals' responsiveness to climate forecasts that issue potentially inaccurate warnings about approaching climate extremes leading to the experience of false and/or missed alarms. The results of this experiment indicate that experiencing false alarms more frequently leads to a decrease in individuals' adaptation investments in response to future warnings (so called "cry-wolf-effect"), but has no impact on individuals' responsiveness to the forecast if no warning is issued. In contrast, experiencing missed alarms more frequently leads to an increase in individuals' responsiveness and investment in adaptation regardless whether or not a warning is issued by future forecasts. Individuals who experienced missed alarms more frequently react more sensitive on warnings per se than individuals without this experience. If they receive a warning their adaptation behaviour is less affected by the forecasted probability of the extreme climate event.

[1] A version of this chapter was submitted to the journal Climatic Change in mid-February 2022 and is currently under review.

This thesis extends the understanding of human behaviour in light of changing ecosystem dynamics and provides more information regarding how natural resource management can be improved by using forecast and ecological prediction systems. The improved understanding of interactions between human behaviour and ecosystem change contributes to the exchange between policy makers, social scientists, ecologists and resource users.

Zusammenfassung

Wissenschaftliche Erkenntnisse belegen, dass die Häufigkeit von Klimaextremen weltweit durch den Klimawandel ansteigt. Klimaextreme erhöhen den exogenen Druck auf lokale Ressourcennutzende, da sie zu einer (teilweisen) Zerstörung von natürlichen Ressourcen beitragen. Die betroffenen Ökosysteme sind häufig bereits aufgrund von Übernutzung in der Vergangenheit in einem schlechten Zustand. Der Zustand von Ökosystemen und natürlichen Ressourcen wird somit sowohl von exogenen (Klima) als auch endogenen (Bewirtschaftung durch Ressourcennutzende) Dynamiken bestimmt. Zukünftig wird eine Kombination aus unsicheren, sich verändernden Umweltbedingungen und menschengemachter Übernutzung eine nachhaltige Bewirtschaftung von natürlichen Ressourcen erschweren.

Abhängig von der zugrundeliegenden Ökosystemdynamik kann die Kombination aus Übernutzung und Klimaextremen zu plötzlichen Veränderungen des Zustandes einer natürlichen Ressource führen, wenn die Ressource an ihren kritischen Schwellenwert (Kipp-Punkt oder Tipping Point) gebracht wird. Solche abrupten Veränderungen werden als Regimewechsel (Regime Shift) bezeichnet. In seiner drastischsten Form führt ein Regimewechsel zum Zusammenbruch der Ressource mit möglicherweise schwerwiegenden wirtschaftlichen Folgen. Ökologische und meteorologische Warn- und Vorhersagesysteme könnten vor bevorstehenden Regimewechseln und Klimaextremen warnen und so zu einem nachhaltigeren Ressourcenmanagement und zu Investitionen in schützende Anpassungsmaßnahmen motivieren.

Die nachhaltige Nutzung natürlicher Ressourcen hängt in hohem Maße von kollektiv abgestimmtem Handeln ab. Ressourcennutzende müssen kooperieren und ihre Strategien zur Ressourcennutzung koordinieren, um eine Ressource auf einem nachhaltigen Niveau zu erhalten und zu verhindern, dass die Ressource ihren kritischen Schwellenwert erreicht. Politische EntscheidungsträgerInnen und ÖkologenInnen haben zu entscheiden, wann und wie sie lokale Ressourcennutzende über die potentielle Gefahr eines Überschreitens kritischer Schwellenwerte informieren. Kritische Schwellenwerte sind jedoch oft unbekannt und

ökologische Frühwarnsignale liefern nur ungenaues Wissen über solche Schwellenwerte. Es ist daher von entscheidender Bedeutung zu wissen, ob die Kommunikation früher, ungenauer Schwellenwertinformationen das Risiko birgt, kollektives Handeln negativ zu beeinflussen. Darüber hinaus entscheidet in anderen Fällen das individuelle Anpassungsverhalten darüber, inwieweit einzelne Personen wirtschaftliche Verluste aufgrund von Klimaextremen erleiden. In diesen Fällen geht es nicht um kollektives Handeln, sondern vielmehr darum, wie einzelne Individuen auf frühzeitige Warnungen reagieren.

Um menschliches Verhalten angesichts der oben beschriebenen ökologischen Dynamiken besser zu verstehen, wurden drei ökonomische Experimente konzipiert und durchgeführt. Die Ergebnisse dieser Experimente werden in den drei empirischen Artikeln dieser Arbeit vorgestellt (Kapitel 2 bis 4):

Der erste Artikel, *The interaction of shock experience and threshold knowledge in natural resource management*, wurde gemeinsam mit Aneeque Javaid und Stefanie Engel verfasst (Kapitel 2). Diese Studie adressiert die Forschungslücke bezüglich der Auswirkungen der Interaktion von exogen und endogen bedingten Veränderungen in Ökosystemen auf kollektives Handeln. Um diese Interaktion zu analysieren und den Hauptantrieb für Veränderungen in der Strategie zur Ressourcenentnahme von Gruppen zu finden, wurde ein neues, (quasi-) zeitkontinuierliches Common-Pool Resource (CPR)-Experiment konzipiert und im Labor durchgeführt. Das CPR-Experiment beinhaltet zwei Dynamiken: einen unerwarteten exogenen Schock, der eine Ressourcenknappheit verursacht, und einen kritischen Schwellenwert, so dass eine Übernutzung der Ressource zu einem abrupten Zusammenbruch führen kann. Zudem wird analysiert inwiefern sich Ressourcenknappheit an sich im Vergleich zu Ressourcenknappheit, die durch einen Schock hervorgerufen wird auf das Verhalten der Ressourcennutzenden auswirkt. Des weiteren wird der Einfluss des Schockerlebnisses und der daraus resultierenden Ressourcenknappheit auf zukünftige Ressourcennutzung untersucht. Einerseits zeigen die Ergebnisse, dass Gruppenmitglieder weniger kooperieren, wenn sie einen exogenen Schock erleben, der durch die teilweise Zerstörung ihrer Ressource zu Ressourcenknappheit führt. Andererseits scheint das Wissen um den kritischen

Schwellenwert eine Mehrheit der Gruppen trotz Schockerlebnis zu einer erfolgreichen Koordination zu motivieren. Die Studie zeigt zudem, dass die Zusammenarbeit zwischen Gruppenmitgliedern und die Effizienz der Ressourcenentnahme stärker durch Ressourcenknappheit an sich als durch die Erfahrung eines exogenen Schocks beeinflusst werden. Ein signifikanter Einfluss von exogenen Schockerlebnissen in der Vergangenheit auf die zukünftige Ressourcennutzung konnte nicht gefunden werden.

Der zweite Artikel, *Are imprecise early warnings a potential benefit or threat to sustainable resource management?*, wurde gemeinsam mit Tobias Vorlaufer und Stefanie Engel verfasst (Kapitel 3). In dieser Studie wird untersucht, ob eine ungenaue Frühwarnung die Kooperation zwischen Ressourcennutzenden verändert und analysiert, ob die Gefahr besteht, dass eine ungenaue Frühwarnung die Reaktion von Ressourcennutzenden auf das Bekanntwerden des tatsächlichen kritischen Schwellenwerts beeinflusst. Einerseits könnten ungenaue Frühwarnungen das Bewusstsein für die Dynamik der Ressource erhöhen und somit kollektives Handeln fördern. Andererseits ist es aber auch denkbar, dass ungenaue Frühwarnungen als Anzeichen für einen unausweichlich bevorstehenden Verlust der Ressource gewertet werden. Dies könnte dazu führen, dass kollektives Handeln scheitert und Ressourcennutzende ihre individuelle Ressourcenentnahme auf Kosten der Gruppe erhöhen. Um die Auswirkungen ungenauer Frühwarnungen auf kollektives Handeln zu untersuchen, wurden zwei zusätzliche Treatments im Rahmen des bereits beschriebenen CPR-Experiments konzipiert und durchgeführt. Die beiden Treatments unterscheiden sich im Grad der Unsicherheit über den Schwellenwert zu Beginn des Experiments. Während alle Gruppen in den beiden Treatments von der Existenz des kritischen Schwellenwerts wissen, erhalten nur die Gruppen des einen Treatments eine ungenaue Frühwarnung. Die ungenaue Frühwarnung besteht darin, dass mitgeteilt wird, in welchem Größenordnungsbereich der Ressource sich der kritische Schwellenwert befindet. Die Experimentergebnisse lassen darauf schließen, dass eine ungenaue Frühwarnung keine Auswirkungen auf die Kooperation und Koordination zwischen den Gruppenmitgliedern hat. Im Vergleich zu Gruppen, die nur von der bloßen

Existenz des Schwellenwerts wissen, werden keine signifikanten Unterschiede im Verhalten festgestellt.

Der dritte Artikel, *False and missed alarms in seasonal forecasts affect individual adaptation choices*, wurde ebenfalls gemeinsam mit Tobias Vorlaufer und Stefanie Engel verfasst (Kapitel 4). In dieser Studie wird analysiert, wie sich die Erfahrung von ungenauen Klimavorhersagen auf das Anpassungserhalten von Menschen auswirkt. Klimaextreme können zu erheblichen wirtschaftlichen Verlusten führen, wenn Individuen keine angemessenen Vorkehrungen zu ihrem Schutz treffen. Die Wirkung von Klimavorhersagesystemen hängt potentiell jedoch von der Genauigkeit der getroffenen Vorhersagen und der Reaktion der Menschen auf ungenaue Klimawarnungen ab. Es wurde ein Online-Experiment entworfen und durchgeführt, um die Reaktion von Individuen auf ungenaue Klimavorhersagen, die potentiell zu falschen und/oder nicht erfolgten Warnungen vor Klimaextremen führen, zu untersuchen. Die Ergebnisse dieses Experiments deuten darauf hin, dass häufiger auftretende Fehlalarme zu einer Verringerung der Anpassungsinvestitionen von Individuen als Reaktion auf zukünftige Warnungen führen (so genannter „Cry-Wolf-Effekt"), aber keinen Einfluss auf die Reaktion der Individuen auf die Vorhersage haben, wenn keine Warnung ausgegeben wird. Im Gegensatz dazu führt ein häufigeres Auftreten von nicht erfolgten Warnungen zu einem Anstieg der Anpassungsinvestitionen der Individuen, unabhängig davon, ob sie eine warnende oder nicht warnende Klimavorhersage erhalten. Es ist zu beobachten, dass Individuen, die häufig keine Warnung erhalten haben, deutlich auf Warnungen an sich reagieren. Wenn sie eine Warnung erhalten, zeigen sie eine hohe Bereitschaft für Anpassungsinvestitionen. Sie passen die Höhe ihrer Anpassungsinvestitionen nur geringfügig an die in der Warnung enthaltenden Wahrscheinlichkeitsangabe des Eintretens eines Klimaextrems an.

Die vorliegende Dissertation erweitert das Verständnis des menschlichen Verhaltens angesichts Veränderungen in Ökosystemdynamiken und liefert Erkenntnisse, wie die Bewirtschaftung natürlicher Ressourcen durch den Einsatz von ökologischen und klimatischen Vorhersagesystemen verbessert werden kann. Das verbesserte Verständnis der Wechselwirkungen zwischen menschlichem Verhalten und Ökosystemveränderungen trägt zum Austausch zwischen politischen EntscheidungsträgerInnen, SozialwissenschaftlerInnen, ÖkologenInnen und Ressourcennutzenden bei.

Contents

List of abbreviations

BDM	Becker–DeGroot–Marschak method
CPR	Common-pool resource
CQ	Control question
CTRL	Control treatment
DV	Dependent variable
EWS	Early warning signals
FA	False alarm treatment
FET	Fisher’s exact tests
HU	Treatment “High uncertainty”
MA	Missed alarm treatment
MCC	Maximum carrying capacity
MWW	Mann–Whitney–Wilcoxon tests
MSY	Maximum sustainable yield
NEP	New Environmental Paradigm scale
NGE	Normalised group extraction
LU	Treatment “Low uncertainty”
RUS	Resource units per second
SD	Standard deviation
SES	Social-ecological systems
SGE	Socially optimal group extraction
SoPHIE	Software Platform for Human Interaction Experiments
SVO	Social Value Orientation
UK	United Kingdom
WTP	Willingness to pay

Figures

Tables

Chapter 1 Introduction

1.1 Overall motivation

Collective action amongst users of natural resources and individuals' adaptation to environmental conditions is under pressure due to both exogenously and endogenously driven change in ecosystems. Climate change increases the frequency of extreme weather events such as droughts or heavy precipitation, which have adverse effects on ecosystems around the world (IPCC, 2014, 2019). From the perspective of local resource users, this can be seen as an exogenous destruction of natural resources provided by these ecosystems. Compounding the problem, resource users' overexploitation of natural resources often impose an endogenous threat by driving gradual change in an ecosystem's underlying conditions (Scheffer *et al.*, 2001; Polasky, Zeeuw and Wagener, 2011). Often ecosystems respond smoothly to gradual or abrupt change until they reach a *critical threshold,* or *tipping point*. Reaching this critical threshold can cause the ecosystem to suddenly switch towards an unfavourable alternative state, which may, in extreme cases, cause the ecosystem to collapse, along with the resource (Scheffer *et al.*, 2001). Such drastic changes are called *regime shifts* if they result in substantial restructuring of the ecosystem, often with extensive effects on human economies and societies (Biggs, Carpenter and Brock, 2009; Scheffer, 2009; Crépin *et al.*, 2012). Some examples of regime shifts include abrupt alterations in water quality, the collapse of fisheries, and shifts in regional climate (Millenium Ecosystem Assessment 2005).

Advanced warnings for upcoming threats based on changes in the endogenous ecological dynamics of ecosystems or based on forecasts of exogenously driven extreme climate events could help to overcome collective action problems in natural resource management and improve individuals' adaptation behaviour. Ecological early warning signals (EWS) have the potential to warn resource users about impending endogenous threats like regime shifts. EWS is a general description of dynamic patterns in a system's behaviour that precedes regime shifts when a system approaches a critical threshold (Scheffer, 2009; Boettiger, Ross and

Hastings, 2013). One of the best studied examples is a phenomenon called *critical slowing down*, where an ecosystem's response to exogenous disturbances is observed to slow down closer to the critical threshold (Boettiger, Ross and Hastings, 2013). However, ecologists debate whether such EWS come in time to prevent reaching the threshold, since the exact location of a critical threshold is difficult to determine (Biggs, Carpenter and Brock, 2009; Scheffer, 2009). Thus, resource users might face uncertainty about the exact level of a threshold before receiving precise information, which could hamper cooperation amongst themselves and prevent sustainable resource management.

Climate forecasts are increasingly important as a predictive tool to warn of exogenous threats like extreme seasonal climate conditions. Seasonal climate forecasts can support and guide reoccurring investments in climate change adaptation. For example, climate change increases the risk of devastating wildfires even in historically low-risk areas, so reoccurring private investment into fire protection for properties before the annual fire season are important (Dowdy *et al.*, 2019; NSW Rural Fire Service, 2020; Readfearn, 2020). Agriculture is also highly susceptible to climate variability and farmers' preparedness for climate extremes like droughts or periods of heavy precipitation depends on the forecasting they receive (Darbyshire *et al.*, 2020). Evidence indicates that modern prediction models outperform predictions based on historical data (Taylor et al. 2015). However, climate forecasts are still highly uncertain, causing forecasts of extreme climate events to be inaccurate (Zommers, 2012; Taylor, Dessai and Bruine De Bruin, 2015; Taylor, Kox and Johnston, 2018). This forecast inaccuracy can lead to two different erroneous forecast scenarios (Losee and Joslyn, 2018). On the one hand, if a forecast issues a warning, but it is a false alarm, decision makers potentially have invested in unnecessary adaptation measures. On the other hand, if a forecast issues no warning, yet the climate conditions are extreme, decision makers could experience a loss because they did not invest in adaptation. In both scenarios, decision makers who base their adaptation choices on inaccurate forecasts potentially experience a loss of welfare. In the long run, the experience of multiple false or missed alarms may result in scepticism of the forecast system and a lower responsiveness to forecasts, with potentially devastating consequences, such as in the case that no adaptation results in the loss of lives and/or livelihoods.

The focus of my PhD research is on two types of behaviour: the management of renewable natural resources by groups of resource users, and individuals' adaptation to climate extremes. With my research, I extend the understanding of collective action and individual adaptation behaviour in the face of exogenously and endogenously driven ecosystem change. I focus on two types of environmental change and the warnings related to each of these: exogenous shocks caused by climate extremes and endogenously driven regime shifts due to human overexploitation. The three co-authored academic papers of this thesis (Chapters 2 to 4)[2] are based on behavioural economic experiments. Chapters 2 and 3 assess the sustainable management of local, open access common-pool resources (CPRs) with the potential for self-governance considering exogenously and endogenously driven ecosystem changes. Specifically, Chapter 2 focuses on the effect of exogenous shocks on collective action if resource users are aware of critical thresholds. Chapter 3 analyses the impact of threshold uncertainty and if imprecise, early warnings of unknown thresholds could hamper collective action. Chapter 4 assesses the impact of repeated inaccurate seasonal forecasts on individuals' adaptation behaviour. By providing an improved understanding of the interaction between human behaviour and ecological dynamics, this thesis aims to provide a body of research to support local policy makers that have to make the decisions of when and how to inform local resource users about an approaching change in ecosystems.

1.2 Background and literature review

This section gives a broad overview of the different strands of literature that this thesis is based upon. The social-ecological systems (SES) approach allows conceptualizing complex human-

[2] The paper presented in Chapter 2 was co-authored by Aneeque Javaid and Stefanie Engel. The papers in Chapters 3 and 4 were co-authored by Tobias Vorlaufer and Stefanie Engel. I am the first author of all three papers.

environment interactions and is introduced in Section 1.2.1. From a SES perspective, it is necessary to consider the underlying ecological dynamics in addition to social aspects when analysing collective action problems. The relevant ecological dynamics are covered in Section 1.2.2. Section 1.2.3 proceeds with a summary of the relevant literature on collective action facing these ecological dynamics. Section 1.2.4 gives an overview of the literature on ecological early warnings signals and climate forecasts as examples for potential warning systems of approaching endogenous and upcoming exogenous threats respectively.

1.2.1 Common-pool resources and the concept of social-ecological systems

In general, the welfare of resource users depends on the use of resource-providing ecosystems (Scheffer, 2009). Open access CPRs are defined as resource systems with no defined property rights whose resource units are accessible to various resource users and where excludability is difficult (Ostrom, 1990; Poteete, Janssen and Ostrom, 2010). The second defining characteristic of CPRs is that they are rival in consumption, i.e. any resource unit extracted by one user is no longer accessible for other users, such that the common-pool resource system is jointly used, but the resource units themselves are not (Ostrom, 1990). However, if the extraction of resource units is balanced with the regrowth of resource units in the system, CPRs can be managed sustainably (ibid.). Prominent examples of CPRs in the context of natural resources are fishing grounds, groundwater basins, grazing areas and water sources, like streams, lakes, and oceans (ibid.).

The concept of SESs emphasises the reciprocal feedback and interdependence between ecosystems and human society (Folke *et al.*, 2010). Due to the linkage of human behaviour and nature, the management of CPRs does not take place in a static environment. Rather, CPRs are part of SESs (Berkes, Folke and Colding, 1998; Ostrom, 2009; Schlüter, Tavoni and Levin, 2016). SESs include interactions amongst resource users, but also interactions between resource users and the resource, for example through management activities like monitoring and extraction (Schlüter, Tavoni and Levin, 2016). Thus, if characteristics of the ecological system affect the user-resource interaction in SESs, these characteristics likely affect behaviour amongst the resource users as well (ibid.). Furthermore, human influence might be crucial in determining the dynamics of the ecological system, depending on how humans adapt their behaviour to ecological change (Folke *et al.*, 2010; Lade *et al.*, 2013).

1.2.2 Ecological background knowledge

Every ecosystem can experience gradual driven change of its environmental conditions because of changes in the climate, or harvesting of resources by resource users (Scheffer *et al.*, 2001; Scheffer, 2009). Usually ecosystems smoothly adapt to such gradual change in conditions over time (ibid.). However, depending on the underlying ecological dynamics of the ecosystem, gradual change can result in a regime shift to an unfavourable alternative state once a critical threshold is reached (ibid.). Regime shifts often persist because improving the environmental conditions to a point that allows for a return to the original state of the ecosystem tends to be costly or even impossible due to the so-called hysteresis effect (ibid.). Thus, regime shifts of ecosystems usually result in irreversible damages to resources and cause severe costs for human society and economic welfare (Scheffer *et al.*, 2001; Crépin *et al.*, 2012).[3] A common example of a regime shift is the collapse of fisheries that can eradicate the income sources of whole areas, as when the Atlantic Cod fishery collapsed in 1992 (e.g. Crépin et al. 2012 and citations therein). Other examples are shifts from forests to savannahs or the bleaching of coral reefs (Scheffer *et al.*, 2001; Scheffer, 2009).

Even when gradual changes in environmental conditions may have no obvious effect on an ecosystem's state, they are still likely to weaken the ecosystem's resilience (Scheffer *et al.*, 2001; Scheffer, 2009). Resilience describes the ability of the ecosystem to recover to its original state after a stochastic, exogenous shock (ibid.). Examples of such shocks are extreme weather events like hurricanes, heavy precipitation or droughts, which have a sudden impact on the underlying environmental conditions (Scheffer *et al.*, 2001). An exogenous shock can

[3] Regime shifts can also have positive consequences, for example, if a turbid lake shifts back to a clear water state (Scheffer *et al.*, 2001). However, this thesis focuses on regime shifts with negative impacts on resource users.

cause a regime shift if it is large enough to push the ecosystem across its critical threshold (ibid.). The weaker an ecosystem's resilience, the closer it is to its critical threshold and the easier the ecosystem can be pushed across the threshold by a perturbation, i.e. an exogenous shock (Scheffer *et al.*, 2001; Scheffer, 2009). A regime shift of an ecosystem is often caused by a combination of gradual change in the underlying environmental conditions, such as endogenously driven overexploitation of resources by resource users, and the occurrence of an exogenous shock, for example due to extreme weather events (ibid.).

1.2.3 Collective action in light of ecosystem changes

The management of open-access common-pool resources characterises a social dilemma known as collective action problem where resource users would be better off as a group if they cooperate and agree on a sustainable management strategy (Ostrom, 1990). This collective action problem has frequently been modelled as a prisoner's dilemma game (e.g. Ostrom et al. 1992, Cardenas and Carpenter 2008, Poteete et al. 2010). Each user of the CPR has a strong incentive to extract from the resource at the individual maximum (Poteete, Janssen and Ostrom, 2010). However, the social optimum would be keeping the resource at its level of maximum sustainable yield, which is the welfare maximising strategy in the long run (ibid.). Each resource user would be better off if all resource users would cooperate (Poteete, Janssen and Ostrom, 2010; Barrett, 2016). Yet, every individual resource user that does not extract their maximum fears that the others will free ride and extract the remaining resources (ibid.). Thus, game theory predicts that cooperation will ultimately fail if individuals behave as rational *homo economicus*; free ride on others' conservation efforts and strategically maximise their own personal benefits (Hardin, 1968; Ostrom, 1990, 2006; Poteete, Janssen and Ostrom, 2010; Barrett, 2016). Consequently, overexploitation by non-cooperative resource users will drive the CPR beyond its safe yield level and eventually destroy it (Walker and Gardner, 1992; Ostrom, Gardner and Walker, 1994). This failure of collective action due to the discrepancy between individual and social interests is commonly referred to as the "Tragedy of the Commons" (Hardin, 1968).

The majority of previous studies on collective action and open access CPR management focus on the analysis of different institutional settings and their impact on the social dynamics amongst resource users in CPR settings with a static environment (for a review of studies see e.g. Ostrom et al. 1994, Cardenas et al. 2013, Janssen et al. 2015 and citations therein). Based on observational field data and economic lab experiments, it has been shown that communication amongst resource users, the establishment of credible commitments, monitoring of extraction activities and sanctioning of misbehaviour help to overcome the collective action problem (e.g. Ostrom 1990, Ostrom et al. 1992, 1994, Balliet 2010). Thereby, such informal institutions and the establishment of social norms can lead to successful self-governance of CPRs without relying on external authorities (Ostrom, 1990, 2006; Ostrom, Gardner and Walker, 1994; Janssen, Lindahl and Murphy, 2015; Nyborg *et al.*, 2016).

However, by using static and deterministic CPR representations with given environmental conditions, these previous studies neglect the interdependence of ecological and social systems (Ostrom, 2009; Janssen, 2010; Tavoni and Levin, 2014; Janssen, Lindahl and Murphy, 2015; Schlüter, Tavoni and Levin, 2016). One early exception is the introduction of multiple rounds to the CPR game with path-dependency of the resource development by Walker and Gardner (1992). This was a first step to include ecological complexities. The study found that path-dependency of the resource development does not prevent failure of collective action, and destruction of the resource prevailed (ibid.).

More complex ecological dynamics have only recently been integrated in experimental lab and field studies to analyse their impact on collective action, such as spatial and temporal variation of the resource stock (e.g. Janssen 2010, Janssen et al. 2010, Cardenas et al. 2013), exogenously or endogenously driven resource levels (e.g. Osés-Eraso et al. 2008, Cerutti and Schlüter 2019) and varying resource growth rates (Kimbrough and Vostroknutov, 2015). These studies confirm theoretical predictions that environmental dynamics effect social interaction and thus, self-governance and cooperation amongst resource users (see also Ostrom 2009, McAllister et al. 2011, Prediger et al. 2014, Pfaff et al. 2015, Schlüter et al. 2016). However, the empirical evidence of an effect of exogenously given scarcity on collective action is inconclusive (Nie et al. 2020 and citations therein). Some experimental lab and field studies find that exogenously given resource scarcity has a negative effect on collective action (del Pilar Moreno-Sánchez and Maldonado, 2010; Prediger, Vollan and Herrmann, 2014; Blanco, Lopez and Villamayor-Tomas, 2015; Gatiso, Vollan and Nuppenau, 2015; Pfaff *et al.*, 2015; Cerutti and Schlüter, 2019), while others find the opposite (Rutte, Wilke and Messick, 1987; Gibson, 2001; Osés-Eraso and Viladrich-Grau, 2007; Osés-Eraso, Udina and Viladrich-Grau, 2008; Nie, Yang and Tu, 2020).

Besides resource scarcity, the mere knowledge of a threatening regime shift affects resource management strategies. Theoretical work by Polasky, Zeeuw and Wagener (2011) outlines that knowledge of potential future regime shifts impacts optimal resource management strategies differently depending on (1) the consequence of the shift and (2) the dynamics that drive the shift. The authors argue that if the regime shift is exogenous (independent from management activities) and causes a stock collapse, the potential of the regime shift increases resource users' discount rate and in turn, causes an increase in extraction. Yet, if the regime shift and thus, the collapse of the stock is driven by resource use (endogenous), the effect on extraction is ambiguous (Polasky, Zeeuw and Wagener, 2011). On the one hand, the potential loss of the resource due to the regime shift could increase exploitation, because resource users might try to gain profits before the expected loss of the resource occurs (ibid.). On the other hand, the desire to avoid the regime shift could decrease exploitation (ibid.). Overall, management tends to become more precautionary in light of an endogenously driven regime shift that causes a collapse, but it is questionable whether that is enough to override the stock effect and thus, the incentive to increase resource exploitation (ibid.).

Overall, the presence of a critical threshold changes the nature of the underlying game faced by resource users (Dannenberg and Tavoni, 2016). If a critical threshold is known with certainty, nature itself acts as a sanctioning institution (Barrett and Dannenberg, 2012, 2014a; Barrett, 2013). Thereby, the presence of a certain threshold turns the multi-player prisoners' dilemma game where individuals need to cooperate to maintain the resource into a coordination game (ibid.). Coordination to maintain the resource at a stock level above the threshold is relatively easy if the benefit of avoiding the regime shift is higher than the cost of doing so (Barrett, 2013; Wagener and de Zeeuw, 2021).

The impact of critical thresholds and the threat of endogenously driven regime shifts on collective action has been assessed in recent economic experiments (e.g. Milinski et al. 2008, Barrett and Dannenberg 2014a, Dannenberg et al. 2015, Schill et al. 2015, Lindahl et al. 2016). This literature focuses on two sub-groups of experiments: (1) threshold public good games (see Dannenberg and Tavoni 2016 for a review) and (2) common-pool resource games. Certain threshold knowledge successfully enforces coordination of individuals' contributions in threshold public good games (e.g. Barrett and Dannenberg 2014a) and leads to more efficient resource management with lower rates of regime shifts in CPR games (Lindahl, Crépin and Schill, 2016; Cerutti and Schlüter, 2019). Threshold public good games without any threshold or with an uncertain threshold represent the classical prisoners' dilemma situation whereas the certain threshold treatment represents a coordination game (Barrett and Dannenberg, 2014a). Both, no threshold being present or uncertain knowledge about the threshold lead to insufficient contributions (Barrett and Dannenberg, 2014b, 2014a). Contributions are lower if no threshold is present than if an uncertain threshold is communicated (Barrett and Dannenberg, 2014b). Certain threshold knowledge in CPR games is also found to decrease overexploitation in comparison to no threshold knowledge set-ups if an endogenously driven regime shift causes an unfavourable change in the resource dynamics (Lindahl, Crépin and Schill, 2016; Cerutti and Schlüter, 2019). This finding is in line with the above-mentioned theoretical predictions of resource management facing endogenously driven regime shifts of Polasky, Zeeuw and Wagener (2011).

1.2.4 Early warning signals and climate forecasts as warning systems

Scientific warning systems become increasingly important to guide resource users' extraction strategies and individuals' adaptation investments in light of complex resource dynamics and potential exogenous shocks. This thesis focuses on two potential warning systems: (1) ecological early warning signals to guide sustainable resource management by warning of approaching endogenously driven regime shifts and (2) climate forecasts that may help to optimise individual adaptation behaviour by warning them of upcoming (exogenous) climate extremes.

Warnings of endogenously driven regime shifts

Firstly, in ecology, early warning signals (EWS) generally describe dynamic patterns that can precede regime shifts (Biggs, Carpenter and Brock, 2009; Scheffer *et al.*, 2012; Boettiger, Ross and Hastings, 2013). Recently, there has been a debate about the potential use of EWS to warn resource users of approaching, endogenously driven regime shifts (ibid.). However, the statistical detection of EWS is difficult and warnings about potential regime shifts include high levels of uncertainty (ibid.). Due to the difficulties of determining exact warning indicators, it is very likely that resource users first receive imprecise warnings without knowing precise specifications. Only with continual and ongoing observation of the dynamic patterns in ecosystems is it possible to determine a specific indicator of a regime shift, such as a critical threshold of a resource (Biggs, Carpenter and Brock, 2009).

Therefore resource users who receive an imprecise early warning of a critical threshold still face high levels of environmental uncertainty regarding the available resource stock. They do not know how much resource they can safely extract without causing a regime shift. Nonetheless, an imprecise early warning reduces the degree of uncertainty about a critical threshold. In general, there are two types of uncertainty defined in the literature, risk and ambiguity (e.g. Sunstein 2010, Aflaki 2013). A critical difference is with the uncertainty of risk, individuals know the underlying probability distribution of events, whereas with the uncertainty of ambiguity, this probability is unknown (ibid.). An imprecise early warning therefore reduces the degree of threshold ambiguity. Resource users gain a vague understanding of the exact threshold, yet the underlying probability distribution remains unknown.

Previous theoretical studies outline two potential responses of resource users to an imprecise early warning that reduces the degree of ambiguity about a critical threshold. On the one hand, resource users could respond to an imprecise threshold warning with more cautious extraction

(Diekert, 2017; Bochet *et al.*, 2019). Receiving an imprecise warning might raise awareness of the threat to cause a regime shift, thus collective action for a cautious extraction strategy could be supported (ibid.). On the other hand, an imprecise threshold warning might enforce overexploitation because resource users feel that the impending regime shift is inevitable (Crépin *et al.*, 2012). The feeling of inevitability could cause a “use-it-or-lose-it” mentality where resource users try to extract as much as they can for themselves and collective action would fail (ibid.).

One strand of the published literature on environmental uncertainty focuses on the impact of uncertainty about the available resource stock on resource management, however the presented evidence is inconclusive. Theoretically, higher levels of uncertainty about the resource size can lower resource users’ extraction efforts (Aflaki, 2013). Not knowing the underlying probability distribution of the resource size (ambiguity scenario) has the potential to make resource users act more cautiously and to decrease resource extraction (ibid.). However, in experimental studies uncertainty rather than certainty about the resource size has been found to increase resource extraction (Budescu, Rapoport and Suleiman, 1990, 1992; Rapoport *et al.*, 1992; Hine and Gifford, 1996; Gustafsson, Biel and Gärling, 1999; Maas *et al.*, 2017).

A second thread of the literature compares certain knowledge of a threshold to different degrees of uncertainty about the level of the threshold. If crossing a critical threshold causes a collapse of the resource, the scenario with uncertainty about the exact level of the threshold is similar to a scenario with uncertainty about the size of the resource stock. In both cases, resource users do not know how much of the resource is available for a safe level of extraction. Previous experimental studies based on threshold public good and CPR games find that collective action fails if the exact level of the threshold is uncertain (Barrett and Dannenberg, 2012, 2014b, 2014a; Brown and Kroll, 2017; Maas *et al.*, 2017). The wider the given range of potential thresholds, the lower the groups’ contributions to the public good, and the more likely collective action will fail (Barrett and Dannenberg, 2014b). Notably, threshold ambiguity causes both coordination and cooperation to fail compared to threshold range knowledge (risk scenario) or certain threshold knowledge (Dannenberg *et al.*, 2015).

Furthermore, theoretical and experimental studies have compared the effect of threshold uncertainty to (1) the effect of uncertainty about the impact of insufficient contributions to a public good (Barrett and Dannenberg, 2012; Barrett, 2016) and (2) the effect of uncertainty about the threshold's presence in CPR games (Schill, Lindahl and Crépin, 2015; Schill and Rocha, 2019). Neither of these two types of uncertainty showed a significant negative effect on collective action (ibid.). Thus, uncertainty about the level of the threshold seems to be more relevant for sustainable resource management and collective action, compared to impact uncertainty and uncertainty about the threshold's presence.

Warnings of exogenous shocks

Exogenous shocks to ecosystems are difficult to predict or control (Scheffer *et al.*, 2001). Climate change increases the frequency of extreme seasons and extreme weather events (IPCC, 2014, 2019). Adaptation is an important tool for the reduction of risks from exogenous shocks caused by extreme climate conditions (ibid.). Due to the ongoing increase in climate variability, it gets increasingly important for governments and individuals to use climate services like seasonal climate forecasts in addition to their climate experiences when making adaptation investments (Bruno Soares, Alexander and Dessai, 2018; Knudson and Guido, 2019; Webber, 2019). In addition to one-time investments, many adaptation investments are reoccurring with the seasons. For example, private homeowners need to invest in protection from wildfires or flooding events before the beginning of each dangerous weather season (Dowdy *et al.*, 2019; NSW Rural Fire Service, 2020; Readfearn, 2020). Farmers' preparedness for climate extremes like droughts or heavy precipitation is also highly dependent on the crop varieties that they plant (Darbyshire *et al.*, 2020). With this in mind, seasonal climate forecasts have great potential to guide reoccurring adaptation investments made by governments, companies and individuals.

However, climate forecasts often suffer from inaccuracy (Zommers, 2012; Taylor, Dessai and Bruine De Bruin, 2015; National Institute of Water and Atmospheric Research (NIWA), 2016; Taylor, Kox and Johnston, 2018) and individuals who base their adaptation investments on forecasts might experience multiple false or missed alarms (Losee and Joslyn, 2018). A *false alarm* is when the forecast issues a warning of an extreme climate with high likelihood (warning forecast), but the climate turns out not to be extreme (ibid.). A *missed alarm* is when the forecast issues no warning (no-warning forecast), but an extreme climate condition occurs (ibid.). Both scenarios have direct costs for decision makers. In the case of a false alarm, they might have invested in unneeded adaptation and be unable to invest elsewhere. In the case of a missed alarm, decision makers may be unprepared for the extreme climate conditions and likely experience a loss of wealth and/or income.

In addition, experiencing multiple false or missed alarms can have negative indirect effects due to a decreased responsiveness to forecasts by decision makers. Firstly, if individuals experience frequent false alarms, the "*cry-wolf effect"* could lead to a lower willingness to invest in adaptation based on a warning forecast in the future (LeClerc and Joslyn, 2015). Secondly, in case of multiple missed alarm experiences, individuals might invest in adaptation even if a future forecast does not issue a warning, thus potentially wasting their resources on unnecessary adaptation measures (ibid.). To avoid negative consequences of false and missed alarm experiences in the long run, these potential indirect negative effects need to be taken into careful consideration by policy makers and agencies who issue climate forecasts.

However the evidence in the literature on these long-term consequences is inconclusive (e.g. Trainor et al. 2015, Lim et al. 2019 and citations therein). Some studies based on a psychological experiment (LeClerc and Joslyn, 2015) and observational data regarding tornado warnings (Simmons and Sutter, 2009; Trainor *et al.*, 2015) find evidence for the cry-wolf effect, whereas other observational studies focusing on behavioural responses to hurricane and tornado warnings do not find evidence for it (Dow and Cutter, 1998; Schultz *et al.*, 2010; Lim, Fisher Liu and Egnoto, 2019).

Furthermore, it is unclear whether false and missed alarms only affect individuals' responses to warning or no-warning forecasts respectively or if they influence individuals' general responses to forecasts. Again, the evidence in previous studies is somewhat mixed. Some studies find evidence that false alarms only affect responses to warning forecasts, and that missed alarms only affect responses to no-warning forecasts (Manzey, Gérard and Wiczorek, 2014; Chancey *et al.*, 2015, 2017). Yet other studies find that both false and missed alarms also have cross-effects on individuals' responses to no-warning and warning forecasts, respectively (LeClerc and Joslyn, 2015; Ripberger *et al.*, 2015; Wiczorek and Meyer, 2016).

1.3 Research questions and hypotheses

Based on the literature review above, Section 1.3 outlines the identified gaps in the literature and presents the arising research questions and associated hypotheses that are addressed in the three empirical papers constituting this thesis (Chapters 2 to 4).

1.3.1 Research question 1

In the two CPR experiments presented in Chapters 2 and 3 (Research questions 1 and 2, respectively), cooperation and coordination are defined as two different behavioural dynamics and analysed separately. *Cooperation* amongst group members is needed to solve the social dilemma and keep the resource at its maximum sustainable yield. Under the assumption of an infinite time horizon, keeping the resource at its maximum sustainable yield would be the social optimum and would result in managing the resource indefinitely at its highest rate of regrowth. However, selfish and myopic individuals have an incentive to maximise their individual payout by free riding on the conservation efforts of others and to choose the maximum extraction for themselves. If cooperation fails and the resource approaches the critical threshold, *coordination* amongst group members is necessary to keep the resource above the critical threshold to avoid the resource's collapse. If the economic consequences of causing a regime shift are extensive, it is in the groups' best interest to prevent reaching the threshold to secure future availability of the resource (Barrett, 2013).

As outlined above, previous experimental studies on resource users' cooperation and coordination behaviour focused on the impact of either resource scarcity due to exogenous shocks (Blanco, Lopez and Villamayor-Tomas, 2015; Cerutti and Schlüter, 2019) or the endogenous threat of crossing a critical threshold, separately (Schill, Lindahl and Crépin, 2015; Lindahl, Crépin and Schill, 2016). However, the potential *interaction* between exogenous and endogenous resource dynamics and its consequences for self-governed resource management has been neglected thus far. This gap is important given that climate change increases the frequency of exogenous shocks due to extreme weather events (IPCC, 2014) while overexploitation of ecosystems weakens their resilience and drives them closer to their critical thresholds (Scheffer *et al.*, 2001). Depending on the impact of the interaction of these ecological dynamics, self-governance of local CPRs might become more challenging and the need for externally introduced institutions to manage the resource sustainably might increase.

Chapter 2 addresses the following research question and hypotheses:

Research Question 1: How does an unexpected exogenous shock resulting in resource scarcity affect both cooperation and coordination behaviour in groups of resource users, when once a known threshold is reached, a regime shift occurs?

Based on previous studies that find an increasing effect of exogenous shock experience on resource extraction (Blanco, Lopez and Villamayor-Tomas, 2015), a decrease in both coordination and cooperation following the exogenous shock is expected. The first hypothesis on this *shock effect* is formulated as follows:

> *Hypothesis 1.1:* With the prior knowledge of a threshold, experiencing an exogenous shock results in both lower *coordination* causing a higher probability of the resource collapsing and lower *cooperation* amongst the resource users as a group.

Two different dynamics are seen as the potential main driver of the shock effect: (1) The mere exposure to resource scarcity itself (*resource size effec*t) and (2) the experience of disruption due to the shock event (*disruption effect*). Thus, the second hypothesis is formulated as follows:

> *Hypothesis 1.2:* The shock induces a decrease in both *coordination* and *cooperation* due to a *disruption effect*. This *disruption effect* is independent of what can be explained by changing resource availability (the *resource size effect*).

Furthermore it is assessed if shock experience and the resulting experience of resource scarcity cause a *spillover effect* on future resource management in a scenario where individuals that experienced a shock to their resource had to migrate to different resources elsewhere. Based on the literature review, the third hypothesis is formulated as follows:

> *Hypothesis 1.3:* There is a negative *spillover effect* of shock experience on future *coordination* and *cooperation*.

1.3.2 Research question 2

Ecological early warning signals (EWS) have been discussed as future policy tools and as potential warning systems of impending regime shifts for resource users (Biggs, Carpenter and Brock, 2009). As highlighted in Section 1.2.4., warnings based on EWS include high levels of ambiguity because the exact determination of details such as critical thresholds is challenging (ibid.). A thorough understanding of the effects of imprecise threshold knowledge on resource management is necessary to guide policymakers and researchers on their decision if and when to use imprecise knowledge to inform the public about critical thresholds. Yet empirical evidence on the effect of imprecise early warnings on both cooperation and coordination amongst resource users is missing and the evidence on the effect of environmental uncertainty on collective action is inconclusive.

Furthermore, previous studies regarding threshold uncertainty neglected the crucial comparison of *different degrees of ambiguity* of the level of the threshold and the *compound* *effect of threshold ambiguity followed by threshold certainty* on both cooperation and coordination behaviour.

Chapter 3 addresses the following research question and hypotheses:

Research Question 2: Does an imprecise early threshold warning, leading to a reduction in the degree of prior threshold ambiguity, affect cooperation as well as coordination behaviour before and after the critical threshold is revealed?

> *Hypothesis 2.1:* Ambiguity about the threshold affects *cooperation* measured as overexploitation of the resource. Groups that receive an imprecise early threshold warning are either (a) *more* likely to overexploit and have a *higher* degree of overexploitation or (b) *less* likely to overexploit and have a *lower* degree of overexploitation than groups that do not receive an early warning.

> *Hypothesis 2.2:* Differences in prior threshold ambiguity affect *coordination*, i.e. the likelihood of the resource collapsing, once the critical threshold is revealed. Groups that receive an imprecise early threshold warning are either (a) *more* or (b) *less* likely to cause a collapse of the resource than groups that do not receive an early warning.

1.3.3 Research question 3

Seasonal climate forecasts are increasingly important to guide individuals' adaptation behaviour in the light of climate change (Bruno Soares, Alexander and Dessai, 2018; Knudson and Guido, 2019; Webber, 2019). Yet, as discussed above, climate forecasts suffer from inaccuracies, which can lead to *false* or *missed alarms*. Previous research suggests that the experience of false and/or missed alarms affects whether individuals are willing to act according to the received forecast (e.g. Manzey, Gérard and Wiczorek, 2014; Chancey *et al.*, 2015; LeClerc and Joslyn, 2015; Wiczorek and Meyer, 2016 and citations therein). However the empirical evidence in these studies on the direction of *main effects* of false alarms on warning forecasts and missed alarms on no-warning forecasts and *vice versa* (*cross-effects*) is inconclusive. Furthermore the reliance on observational and self-reported data in previous studies makes it difficult to control for individuals' actual experiences of false and/or missed alarms (e.g Ripberger *et al.*, 2015; Whitmer *et al.*, 2017; Lim, Fisher Liu and Egnoto, 2019). Overall, a *systematic analysis of how experiencing false and/or missed alarm*s more frequently affect individuals' willingness to invest in adaptation based on probabilistic climate forecasts is missing. Such a systematic analysis, including main and cross-effects, would be crucial for the design and communication of climate forecasts by policy makers and agencies.

This leads to Chapter 4 addressing the following research question and hypotheses:

Research Question 3: If individuals receive probabilistic climate forecasts of unknown accuracy levels resulting in more frequent false and/or missed alarms, what impact does this have on individuals' adaptation investment?

Hypotheses on the main effects:

> *Hypothesis 3.1:* Experiencing *false alarms* more frequently decreases adaptation investments in response to a *warning forecast* ("cry-wolf effect").
>
> *Hypothesis 3.2:* Experiencing *missed alarms* more frequently increases adaptation investments in response to a *no-warning forecast*.

Hypotheses on the cross-effects:

> *Hypothesis 3.3:* Experiencing *false alarms* more frequently increases adaptation investments in response to a *no-warning forecast*.
>
> *Hypothesis 3.4:* Experiencing *missed alarms* more frequently decreases adaptation investments in response to a *warning forecast*.

1.4 Methodology

All three papers of this thesis are based on the statistical analysis of economic experiments. Economic experiments allow to control individuals' decision environment and the present institutions, i.e. the set of available actions (Smith, 1982). In contrast to observational data, the control of treatments' design and a targeted variation of single parameters between treatments allow to estimate causal effects (Smith, 1982; Al-Ubaydli and List, 2015). In addition, confounding factors like participants' socio-economic background, their social and risk preferences can be controlled for in experiments by the implementation of additional surveys and randomisation of participants to treatments.

There are several types of economic experiments that differ in their general set-up. For example, lab experiments are usually computer-based and are commonly run with (university) students as subjects. Artefactual field or lab-in-the-field experiments are similar to lab experiments, but run with professionals as subjects, e.g. farmers or managers (Harrison and List, 2004; Al-Ubaydli and List, 2015). In framed field experiments, the subjects are actors from the field and the field context is part of the experimental set-up, so the decision environment is similar to the actual environment of interest (ibid.). This allows, for example, experiments that target agricultural policies to be run with actual farmers. Online experiments are often survey-based and subjects are recruited from the general population, usually via specialised crowdsourcing platforms. All of these types of experiments have in common that subjects know they are participating in experiments for research (Al-Ubaydli and List, 2015).

1.4.1 Continuous-time common-pool resource experiments in the lab

To address the first two research questions in Chapters 2 and 3, two lab experiments were designed and implemented. By running lab experiments (at both Osnabrueck and Hamburg Universities, Chapters 2 and 3 respectively) we were able to control for the exposure of subjects to alternative ecological scenarios and to use randomisation to balance unobservable variables between treatment and control groups (Harrison and List, 2004; Falk and Heckman, 2009; Al-Ubaydli and List, 2015). The experiments allowed for the identification of causal effects between ecological dynamics and behavioural responses (ibid.). Furthermore, the computerised lab experiments made it possible to run an experiment with moderately complex ecological dynamics, while at the same time controlling and accurately measuring the social experience of subjects in their groups.

The design of the (quasi-) continuous-time common-pool resource (CPR) experiment is based on the general set-up presented in Brandt et al. (2017), and includes more complex ecological dynamics such as exogenous shocks and critical thresholds. The integration of ecological dynamics in the experimental environment increases the relevance of such behavioural experiments and improves the understanding of the impact of ecological dynamics on collective action (Janssen, 2010; Janssen *et al.*, 2010).

In Chapter 2, subjects in all three treatments knew of a critical threshold and that reaching it would cause a catastrophic regime shift, resulting in the collapse of the resource. In addition to the control treatment, we implemented one treatment where subjects experienced a sudden exogenous shock and one treatment with initial resource scarcity. The exogenous shock was unknown to subjects at the beginning of the game and resulted in resource scarcity due to the loss of resource units at a set time during the game. Subjects that faced initial resource scarcity started the CPR game with a lower resource level. In Chapter 3, we implemented two additional treatments that differed in subjects' knowledge about the exact level of the threshold at the beginning of the game. In one treatment, subjects received an imprecise early warning about the threshold by being told the threshold's range upfront. In the other treatment, subjects only knew of the mere existence of the threshold without any further information on its level. Again, the consequence of reaching the threshold, (the collapse of the resource) was communicated to subjects in both treatments from the beginning.

In both CPR experiments (Chapters 2 and 3), subjects were randomly assigned to groups of four and had to manage a shared resource over an unknown amount of time (measured in seconds). The resource's regrowth was based on a logarithmic growth function (Perman *et al.*, 2011; Brandt *et al.*, 2017). Subjects' payout was dependent on the sum of resource units that individuals extracted for themselves.

It was decided to use a (quasi-) continuous-time CPR game that incorporated ecological complexities to mimic the dynamic decision environment that actors in the field are exposed to (Janssen *et al.*, 2010). Dynamic decision making is present if context and time matter (Brehmer, 1992). In dynamic decision making, the decision maker must consider the consequences of their present decisions, as this will affect future decisions as well. Furthermore, the decision environment develops not only due to external influences, but also due to the decision makers' intrinsic actions (ibid.). A continuous-time experimental design allows for such dynamic decision making and has multiple advantages over a round based design. In a continuous-time CPR experiment the resource develops dynamically over time and asynchronous, strategic interaction amongst group members is possible (Pettit *et al.*, 2014). Furthermore, long-term interaction between subjects can be achieved in a relatively short time period because time continuity fastens the process of adjustment dynamics (ibid). However, due to technical limitations, the time unit of the updating process of the resource and the extraction development in the two CPR experiments was restricted to seconds, and therefore the CPR experiments are technically *quasi*-continuous-time (Bigoni *et al.*, 2015). Nonetheless, subjects were able to change their individual extraction choice any second and could dynamically respond to both other resource user's extraction behaviour and the development of the resource. Further details about the experimental design and instructions are presented in Chapters 2 and 3.

Due to the complexity of the general experimental set-up, several pilot sessions were run during the design phase of the experiment. The understanding of the instructions was assessed using subjects' feedback regarding the handling of the experiment for both the pilot and experimental sessions with an extensive post-experimental questionnaire. Based on the feedback given to these questionnaires, it was evident that the design was well understood (see Chapters 2 and 3 for details). Together, this implies that this CPR experimental design offers a valid tool to add to the understanding of the effect of ecological dynamics on collective action.

1.4.2 Online experiment

To answer Research Question 3 (Chapter 4), and systematically assess how accurate or inaccurate forecast experiences affect resource user's adaptation investments, participants needed to experience a series of false and/or missed alarms. However, as multiple false or missed alarms are difficult to observe in the real world, an online experiment was used. Implementing such an economic experiment instead of relying on observational data allowed us to control for participants' experiences with inaccurate forecast systems.

The experimental design is based on an approach with "no deception", such that false and missed alarms were implemented by using probabilistic forecast systems that varied in their level of accuracy. In the control treatment, the issued forecast probability warning (or lack of warning) of an approaching extreme climate condition was always accurately based on the underlying risk to face extreme climate conditions. In the false alarm treatment, the forecasted probability for an upcoming extreme climate condition was set to be higher than the true underlying risk. Similarly, in the missed alarm treatment, the forecasted probability for an upcoming extreme climate condition was set to be lower than the true underlying risk. With this implementation, subjects in all treatments still had the chance to experience both accurate forecasts and false or missed alarms, but depending on the treatment the likelihood of these three outcomes was different. Participants knew of the varying levels of accuracy of the forecast systems, but did not know the specifics of the forecast system that issued the forecasts in their case. Chapter 4 outlines the details of the experimental design and the instructions.

The probabilistic nature of the experimental set-up called for higher numbers of observations than achievable by running a lab experiment. Therefore, the experiment was made available online instead of limited to a lab. Generally online experiments are a more cost-effective way to achieve higher power than lab experiments, and include a more diverse background of participants (Peer et al. 2017, Palan and Schitter 2018 and citations therein).

Due to these advantages, online studies are increasingly popular in the social sciences, including the field of experimental economics (Bohannon, 2016). The number of published papers in the social sciences using Amazon Mechanical Turk (Mturk) increased from 61 in 2011 to 1,120 in 2015 according to Google Scholar (ibid.). Mturk has been the dominant crowdsourcing platform so far, but it was never designed to support scientific research (ibid.). Due to weaknesses in the design of the service and a lack of guidelines, experimental data generated via Mturk is potentially biased (ibid.). For this study it was decided to use the crowdsourcing provider Prolific (Prolific, 2021), which is explicitly designed for research. Prolific has clear guidelines and rules for participants and researchers, which distinguishes it from Mturk (Palan and Schitter, 2018). For example, by having clear rules regarding minimum payments and the treatment of submissions, participants on Prolific are less likely to try to please the researcher to make sure that they are paid than participants on Mturk (ibid.). Thus, an experimenter demand effect that could bias the experimental results is less likely (ibid.). Furthermore, Prolific provides a detailed pre-screening function, which gives researchers more control over participants' backgrounds (ibid.). The pre-screening function was especially important for the presented experiment in Chapter 4 to limit the variation of subjects' previous experiences with extreme weather events. For this purpose, the pre-screening function was used to recruit only residents of the United Kingdom as participants.

1.4.3 Statistical methods

Non-parametric and parametric statistical methods were used to analyse the data collected by the implementation of the three experiments. The data analysis of the CPR experiments in Chapters 2 and 3 was based on group observations, whereas the analysis of the online experiment was based on individual observations. The choice of the statistical methods depended on the sample size and the assumption of normality of the data distributions (Moffatt, 2021). Due to the small numbers of group observations in Chapters 2 and 3,[4] the main tests

[4] About 20 group observations per treatment in Chapter 2 and 45 group observations per treatment in Chapter 3.

for significant differences between the treatment and control groups were based on non-parametric tests (ibid.). Regression models were added as robustness checks in the appendices. Besides the robustness check of the treatment effects, the regression models offered the opportunity to assess the influence of additional explanatory variables measuring, for example the effect of participants' socio-economic background on both cooperation and coordination behaviour.

In Chapter 4, parametric tests, i.e. regression models, were used for the main analysis of treatment effects. Given that there are 2,000 participants in the three treatments of the online experiment, there is a sufficiently large enough sample (significantly more than 30 observations) to rely on these parametrical methods (Moffatt, 2021). We assessed the effect of the different treatments and of additional explanatory variables like individuals' socio-economic background on the dependent variable, which in this case was individuals' willingness to invest in adaptation. In all three experiments the additional explanatory variables were elicited by implementing post-experimental questionnaires. Details of the data analysis can be found in Chapters 2 to 4.

Given the ongoing concern regarding the quality of scientific research and the discussion of inflated findings due to "data mining" and "p-hacking" (amongst others Coffman and Niederle 2015, Olken 2015, Camerer et al. 2016, 2018), we decided to pre-register the experiments conducted for Chapters 3 and 4 via the online platform "AsPredicted.org" (Wharton Credibility Lab, 2017). The terms "Data mining" and "p-hacking" refer to the practice of cherry-picking statistically significant results to report in papers, leading to a publication bias of only statistically significant results (Olken, 2015). Pre-registration of the experimental hypotheses and the planned data analysis before examining the data is seen as a commitment device that improves the transparency of research (Coffman and Niederle, 2015; Olken, 2015). Researchers are then motivated to report all their analysis matching their pre-registered analysis plans, including statistically insignificant results and to state explicitly which parts of the analysis are exploratory (ibid.). Thus, a combination of pre-registration and replication of studies should help to fight inflation of p-values in published work in the future (e.g. Coffman and Niederle 2015, Olken 2015, Camerer et al. 2016, 2018). The pre-registration documents are included as appendices in Chapters 3 and 4.

1.5 Results and contributions

This thesis attempts to contribute to the body of knowledge regarding human behaviour, that guides policy makers, social scientists, ecologists and resource users to respond in a meaningful way to climate-change related crises. This is done by assessing the combined effect of exogenously and endogenously driven change of natural resources, as well as the effect of imprecise early warnings of thresholds and conducting a systematic analysis of the impact of inaccurate forecast systems on adaptation investments.

With the support of my supervisors, I designed and implemented two novel (quasi-) continuous-time CPR experiments to assess in how far self-governance and the management of renewable natural resources are influenced by exogenously and endogenously driven environmental change (Chapters 2 and 3). The experiments contribute to the methodology of continuous-time lab experiments (e.g. Oprea et al. 2014, Pettit et al. 2014, Bigoni et al. 2015) and especially to the use of experiments focusing on the implementation of environmental complexity in continuous-time CPR experiments (Janssen, 2010; Janssen *et al.*, 2010; Brandt *et al.*, 2017; Cerutti and Schlüter, 2019). Integrating more complex environmental dynamics in lab experiments is seen as a valuable tool to test the generalisability of outcomes (Janssen et al. 2010). By taking complexity that is observed in the field back to the lab to study the dynamics in detail, lab experiments can support the analysis of processes that are observed in many social-ecological systems (ibid).

The work presented in this thesis is, to the best of my knowledge, the first to combine an exogenous shock that causes resource scarcity with the threat of an endogenously driven regime shift (Chapter 2). Overall, precise knowledge of local resource dynamics including critical thresholds seems to support sustainable self-governance of natural resources in light of negative, exogenous shocks driven by climate change. Comparing groups that do and do not experience an exogenous shock, we find evidence that the experience of an exogenous shock decreases cooperation, but does not impact coordination. The number of regime shifts does not increase significantly if groups experience resource scarcity due to the exogenous shock. Thus, knowing of the threat to cause a catastrophic regime shift once a critical threshold is reached seems to prevent a negative effect of shock experience on coordination, but not on cooperation.

However, the reason for resource scarcity matters. Groups that experience resource scarcity due to an exogenous shock after a period of resource abundance are more cooperative and extract the scarce resource more cautiously, compared to groups that experience resource scarcity from the start. Thus, previous experience of resource abundance seems to have a persisting positive effect on cooperation. For coordination, the corresponding difference is not statistically significant. Furthermore, there is no statistically significant evidence that previous shock experience influences future cooperation or coordination once resource users are confronted with an abundant resource again.

Taken together, these results indicate that an increase in the frequency of extreme weather events due to progressing climate change does not necessarily lead to a higher rate of ecosystem collapses caused by overexploitation. If resource users are aware of ecological dynamics, especially the concept of critical thresholds and regime shifts, and if consequences of a shift are extensive, i.e. a collapse of a resource, resource users are likely to coordinate to maintain that resource. However, an increase in exogenous shocks is found to decrease cooperation amongst resource users, which potentially leads to less efficient resource management. After the shock experience, resource users cooperate less and are therefore less likely to keep a resource at its level of maximum sustainable yield. However resource management strategies should aim to recover resources back to abundant levels to prevent the deterioration of cooperation and coordination in the future. The experimental results indicate that ongoing exposure to resource scarcity, such that scarcity becomes the only known state for resource users, hampers collective action. In contrast, the prior experience of resource abundance seems to support more cautious resource extraction strategies once an exogenous shock causes scarcity. Educating resource users about ecological dynamics and supporting the recovery of resources after exogenous shocks not only improves the immediate state of the resource, but also the ecosystems' resilience by supporting resource users' cooperation following future shock events.

> **Conclusion 1: Experiencing an unexpected exogenous shock resulting in resource scarcity results in less cooperation amongst resource users. To support sustainable self-organised resource management by resource users, it is essential to give them the opportunity to experience resource abundance. Educating resource users about the consequences of critical thresholds and regime shifts is crucial. Lessons learned could be to concentrate development aid to confined areas to enable resource users to experience resource abundance.**

In Chapter 3, two more treatments based on the (quasi-) continuous-time CPR game were included to assess the effect of imprecise early warnings about a threshold on collective action. Previous studies focusing on CPR management have assessed the effect of uncertainty about the threshold's presence (Schill, Lindahl and Crépin, 2015; Rocha *et al.*, 2020), and different levels of uncertainty about the level of a threshold (Maas *et al.*, 2017) in standard round based CPR games. A shortcoming of these studies is they miss the comparison of different levels of ambiguity about the level of the threshold. Resource users in the field are likely to know that there is a critical threshold, yet detailed information about the exact level are often missing or fluctuating due to climate change.

The experiment presented in Chapter 3 furthers the understanding of the effect of different levels of ambiguity about the level of the critical threshold on both cooperation and coordination amongst resource users. We find that receiving an imprecise early warning about the threshold in the form of a threshold range knowledge does not cause a difference in cooperation amongst group members compared to a scenario where only the mere presence of the threshold is known. There is also no evidence that such an imprecise early warning makes a difference in coordination once the precise level of the threshold is revealed. Thus, the presented experimental results suggest that there is neither harm nor benefit in the communication of vague, imprecise knowledge about the range of critical thresholds by researchers and policy makers. We see a weak tendency that lower levels of threshold ambiguity in the beginning lead to more sustainable resource management over time. However, further research is needed to determine the level of precision that is needed to make early warnings a useful tool to support collective action early on.

> **Conclusion 2: Early communication of vague, imprecise knowledge about the range of critical thresholds does not seem to harm nor foster cooperation as well as coordination.**

Chapter 4 contributes to the literature regarding the use of seasonal climate forecasts and their design. As outlined above, previous evidence of the effects of experiencing false and/or missed alarms more frequently on individuals' responsiveness to warnings is inconclusive. By systematically analysing the effects of false and missed alarms on responses to warning and no-warning forecasts, this thesis and papers therein contribute to an improved understanding of the use of probabilistic climate forecasts. We find that the experience of multiple false alarms decreases individuals' responsiveness to future warnings ("cry-wolf effect"), and that the experience of multiple missed alarms decreases individuals' willingness to rely on no-warning forecasts. However, experiencing missed alarms increases individuals' responsiveness to warning forecasts and they are more willing to invest in adaptation than those who experienced mostly accurate forecasts. Individuals that were exposed to a predominantly false alarm prone forecast system in the past, and received no-warning forecasts in the future, displayed no difference in their adaptation behaviour compared to individuals who had only experienced accurate forecasts. However, these treatment effects are relatively small in relation to the effects of the forecasted probabilities on individuals' adaptation investment choices. Even if individuals experience false or missed alarms more frequently, they still respond to differences in the forecasted probabilities and invest less or more if the forecasted probability is lower or higher, respectively.

Therefore, in the long run, the costs of forecast inaccuracies in the sense that individuals spend too little on adaptation might be limited if climate forecasts incorporate probabilities of the alternative forecasted scenarios. Communicating probabilities allows individuals to judge the level of uncertainty for themselves (LeClerc and Joslyn, 2015). Even with a history of false alarms and a reluctance to respond to warning forecasts, individuals still invest more if the forecasted probability of an extreme climate is very high.

However, missed alarm experience causes individuals to strongly respond to a warning forecast *per se*, without taking forecast probabilities adequately into consideration. This may lead to overshooting adaptation investments in the event that they receive a warning of an extreme upcoming season with a low forecasted probability. Nevertheless, in the case of very high forecasted probabilities of extreme upcoming seasons, individuals' adaptation behaviour with missed alarm experience is similar to the behaviour of individuals with a history of accurate or false alarm-prone forecast systems.

However, case specific assessments of forecast designs are necessary if policy makers and researchers want to ensure a beneficial use of such forecasts because responses likely depend on the stakes that are at risk and the costs of adaptation.

> **Conclusion 3: If individuals experience multiple false alarms, they will be reluctant to invest in adaptation in response to future warnings of climate extremes. Thus, their level of adaptation will likely be insufficient. In contrast, missed alarm experience likely causes individuals to overshoot their investments in adaptation since they invest irrespective of the issued forecast.**

Overall, despite the progress achieved by the findings from the lab as well as the online experiments, we are still only beginning to understand the systematic impact of ecological dynamics on human behaviour in socio-ecological systems.

1.6 Methodological limitations and directions for future research

The assessment of the impacts of ecological dynamics and early warnings on collective action, and of the forecast systems' effect on adaptation behaviour in the lab and online, provides a valuable initial step towards understanding general behavioural patterns. However, it should be acknowledged that the composition of the subject samples and the constructed experimental environments limit the generalisability of the outcomes of this thesis (Levitt and List, 2007).

Firstly, students' homogenous socio-economic background in lab experiments can be both an advantage and disadvantage. The students participating in the studies described in Chapters 2 and 3 are part of a western, educated, industrialised, rich and democratic (WEIRD) population, and are from the same university, so are likely to have similar previous experiences with extreme weather events and ecosystem changes (Harrison and List, 2004; Levitt and List, 2007; Henrich, Heine and Norenzayan, 2010; Al-Ubaydli and List, 2015). This is a common phenomenon and applies to many similar studies (ibid.). Recruiting subjects with a similar background makes it easier to control any confounding factors of their behavioural response to the implemented resource dynamics in the CPR experiments, whereas field experiments with more heterogeneous subjects potentially have a more diverse background and accompanying experiences with ecosystem changes. The homogenous nature of students from a single source limits the generalisability of experimental outcomes from the lab (also referred to as "low external validity") and insights from these lab experiments will not fully translate to real world systems (Harrison and List, 2004; Levitt and List, 2007; Henrich, Heine and Norenzayan, 2010; Al-Ubaydli and List, 2015).

It should be noted, that in contrast to the lab experiments run with students at Osnabrueck and Hamburg Universities (Chapters 2 and 3), the recruited subjects for the online experiment (Chapter 4) were from the general population of the United Kingdom. The subjects from the online experiment therefore have a more diverse socio-economic background and are likely more representative of the general population than students are. Nevertheless, subjects are still from a WEIRD population.

Secondly, even though decision making in the experiment is incentive-compatible in the sense that monetary incentives imply real economic consequences for subjects (Falk and Heckman, 2009), they make their decisions in a constructed environment (Levitt and List, 2007). Therefore, outcomes from both lab and online experiments have limited generalisability (ibid.). To further this work, I suggest that the findings of this thesis be both complemented and verified with lab-in-the-field and/or field experiments and studies based on data from natural settings (ibid.). For example, the implications of imprecise early warnings and the impact of forecast designs on individuals' responsiveness could be analysed with real actors in (lab-in-the) field settings to improve policy recommendations and designs. Overall, lab or online experiments, field experiments, survey data, and standard econometric methods should be seen as complements to improve knowledge, rather than substitutes (Levitt and List, 2007; Falk and Heckman, 2009).

Thirdly, the generalisability of experimental outcomes is limited by the self-selection of subjects into participation. All subjects for the three experiments were recruited from preregistered subject pools to which individuals sign up voluntarily.[5] Thus, all participants were interested and motivated to take part in research, which potentially biases the experimental outcomes because volunteers might have different preferences and motivations than non-volunteers (Levitt and List, 2007). Furthermore, the sampling bias could be further enhanced in online experiments because less wealthy individuals might be lacking the technical means to participate in such experiments, inducing differences in the economic background of

[5] The participants of the three experiments were recruited from three different subject pools with the help of three different organisational software tools. The experiment for Chapter 2 was implemented at the Laboratory for Experimental Research at the Westerberg Campus of Osnabrueck University and participants were recruited with the software Orsee (Greiner, 2015). The experiment presented in Chapter 3 was implemented at the WISO Experimental Lab of Hamburg University and participants were recruited with the organisational software tool hroot (Bock, Baetge and Nicklisch, 2014). The participants for the online experiment were recruited via the crowdsourcing platform Prolific (Prolific, 2021).

participants and non-participants (Palmer and Strickland, 2016). We implemented post-experimental questionnaires and gathered information on subjects' socio-economic background to identify which types of individuals self-selected to participate (Falk and Heckman, 2009). However, as we do not have any information on the general population (e.g. students in Osnabrueck), we cannot analyse what kind of subjects did not participate.

With respect to the participation of subjects, lab and online experiments have their own advantages and disadvantages. In-person participation in the lab has the advantage that the recruitment process at the university can control the experimental environment and the identity of subjects. This makes their statements on their socio-economic background more reliable. However, physical interactions with the experimenter could undermine the feeling of anonymous participation and enhance an experimenter demand effect (Levitt and List, 2007). Even if the standard experimental code of conduct is followed and all answers in the experiment are given anonymously, subjects in the lab are aware that the experimenter monitors their actions. Therefore the results might be biased due to subjects making socially appropriate choices. Furthermore, the limited availability of participants for lab experiments leads to small sample sizes with reduced statistical power. Even though our samples with 320 students participating at Osnabrueck University and 360 at Hamburg University are large compared to similar studies based on lab experiments, the samples are still small compared to sample sizes achievable by online experiments. We decided to implement the two CPR experiments in the lab due to the group-based design and the lower likelihood of participants dropping out of the experiment once it has started. Running online experiments always includes the risk of high dropout rates, which is especially costly if an individual dropout causes the loss of a group observation.

Not being able to control the accuracy of subjects' identities, the provided socio-economic background information and in which environment they participate is one of the disadvantages of online experiments (Palan and Schitter, 2018). However, this lack of control can also be an advantage because it gives participants a higher level of anonymity. Thus, behavioural responses in online experiments are potentially less biased than in the lab (Levitt and List, 2007).

The ecological complexity incorporated in the two CPR experiments was limited to ensure that the instructions to the game were understandable within a reasonable amount of time. In both CPR experiments (Chapters 2 and 3), participants did not experience any costs connected to the resource extraction, whereas in the real world extraction is costly. Additionally, in the real world, extraction costs often increase if the resource is scarce (Cinner *et al.*, 2011), while resource extraction in the CPR experiments presented in this thesis had a constant rate of return.

In Chapter 2, the exogenous shock was time dependent, and every participant ended on the same post-shock resource level, regardless of the given resource level at the time of the shock. Furthermore, in contrast to natural catastrophes in the real world, the implemented exogenous shock only reduced the resource level and with it, participants' potential future income, but did not affect participants' accumulated wealth. It would be interesting to incorporate these missing complexities one by one in future experiments to further understand their impact on collective action when resource users are simultaneously facing exogenous shocks, resource scarcity and critical thresholds.

Based on our results in Chapter 3, it might be worthwhile to test in the future different levels of imprecision of threshold range knowledge as early warnings. Previous experimental studies found a coordinating effect of threshold range knowledge only for comparatively small ranges (Barrett and Dannenberg, 2014b) and the effect could depend on case-specific settings in the field. Thus, further research is needed to assess how ecological early warning signals could be used to improve collective action and to ensure sustainable resource management even if ecosystems are threatened by ongoing change.

It would also be interesting to vary the level of social uncertainty amongst group members to mirror different situations from the field in the future. With the current design, participants could observe the development of the groups' total extraction and knew how much the other group members extracted in relation to their own extraction. However, such a level of perfect feedback and control is often not given in the field. Given that communication amongst group members enhances cooperation (see Balliet 2010 for a review), it would be interesting to assess how communication affects cooperation and coordination in our setting.

In general, the studies in Chapters 2 and 3 open the chance to extend the experimental design of the (quasi-) continuous-time CPR game to further analyse collective action and resource management in a controlled environment allowing for both the strategic interaction amongst group members and the implementation of ecological complexities. For example, the implementation of external institutions, various sizes of exogenous shocks, varying group sizes or different imprecise early warnings and forecast mechanisms could be assessed in the future. Furthermore, collaborations with ecologists could be a fruitful endeavour in the future. The exchange with ecologists will be essential to identify reasonable environmental parameters that could work as early warning signals, and further research is necessary to identify whether the costly detection of such early warnings could actually be beneficial for sustainable resource management.

Ecologists often neglect the role of human behaviour in socio-ecological systems or base the modelling of it on the simplified assumption of the *homo oeconomicus*, a rational, selfish and payout-maximising individual (Schlüter *et al.*, 2017). Extensions of the CPR game could be used to improve the modelling of human behaviour in ecological models, such as experiments focusing on the individual decision making instead of groups' dynamics to assess how individual decision making varies depending on the ecological dynamics in the socio-ecological system. The experimental outcomes could then be used as parameters for human behaviour in ecological models. An additional idea would be a modification of the CPR experiment to assess how time delays in the knowledge of the resource development impact decision making in the present. For example, time lags in the resource's updating could be designed such that resource users do not know the present state but instead get outdated information on the resource from two to three periods ago.[6]

With regards to the impact of forecast inaccuracies on individuals' adaptation behaviour as presented in Chapter 4, it would be interesting to analyse different types of forecast designs in future work. Given that the communication and understanding of probabilistic forecasts is challenging, especially if actors are of variable education and experience (Budescu *et al.*, 2014), the effect of forecast inaccuracies on adaptation behaviour could change depending on the understanding of the recipients of such forecasts. Not to forget that the stakes in the experiment presented in Chapter 4 were comparatively low and individuals' willingness to adapt might change with increasing risk. With this in mind, testing the design of probabilistic climate forecasts in the field and educating recipients of such forecasts would be interesting in the future.

Finally, it would be interesting for a future study to combine and contrast the ecological and behavioural dynamics from the three academic papers of this thesis. Such a study would assess what effect the experience of forecast inaccuracies has on both collective action and self-management of natural resources simultaneously. This could provide the framework to predict the behavioural consequences of false and/or missed alarms on individuals' behaviour within groups and thus, give policy makers guidance when making policies to respond to climate change.

[6] We thank Matthew Adamson for the fruitful discussions on the use of economic experiments to identify parameters used in future ecological models and how the two strands of research could benefit from each other even more in the future.

References Chapter 1

Aflaki, S. (2013) ‘The effect of environmental uncertainty on the tragedy of the commons’, *Games and Economic Behavior*. Elsevier Inc., 82, pp. 240–253. doi: 10.1016/j.geb.2013.07.011.

Al-Ubaydli, O. and List, J. A. (2015) ‘On the Generalizability of Experimental Results in Economics’, *Handbook of Experimental Economic Methodology*. Edited by G. Frechette and A. Schotter. Oxford University Press, pp. 420–462. doi: 10.1093/acprof:oso/9780195328325.003.0022.

Balliet, D. (2010) ‘Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review’, *Journal of Conflict Resolution*, 54(1), pp. 39–57. doi: 10.1177/0022002709352443.

Barrett, S. (2013) ‘Climate treaties and approaching catastrophes’, *Journal of Environmental Economics and Management*. Elsevier, 66(2), pp. 235–250. doi: 10.1016/j.jeem.2012.12.004.

Barrett, S. (2016) ‘Collective Action to Avoid Catastrophe: When Countries Succeed, When They Fail, and Why’, *Global Policy*, 7(S1), pp. 45–55. doi: 10.1111/1758-5899.12324.

Barrett, S. and Dannenberg, A. (2012) ‘Climate Negotiations under Scientific Uncertainty’, *Proceedings of the National Academy of Sciences*, 109(43), pp. 17372–17376. doi: 10.1073/pnas.1201800109.

Barrett, S. and Dannenberg, A. (2014a) ‘Negotiating to avoid “gradual” versus “dangerous” climate change: An experimental test of two prisoners’ dilemmas’, in Cherry, T. L., Hovi, J., and McEvoy, D. M. (eds) *Toward a New Climate Agreement: Conflict, Resolution and Governance*. Taylor & Francis, pp. 61–75. doi: 10.4324/9780203080009.

Barrett, S. and Dannenberg, A. (2014b) ‘Sensitivity of collective action to uncertainty about climate tipping points’, *Nature Climate Change*, 4, pp. 36–39. doi: 10.1038/nclimate2059.

Berkes, F., Folke, C. and Colding, J. (eds) (1998) *Linking Social and Ecological Systems: Management Practices and Social Mechanisms for Building Resilience*. Cambridge, UK: Cambridge University Press. Available at: https://books.google.de/books?id=CueiTFXCRXQC&pg=PA1&hl=de&source=gbs_toc_r&cad=4#v=onepage&q&f=false.

Biggs, R., Carpenter, S. R. and Brock, W. A. (2009) ‘Turning back from the brink: Detecting an impending regime shift in time to avert it’, *Proceedings of the National Academy of Sciences*, 106(3), pp. 826–831. doi: 10.1073/pnas.0811729106.

Bigoni, M. *et al.* (2015) ‘Time Horizon and Cooperation in Continuous Time’, *Econometrica*, 83(2), pp. 587–616. doi: 10.3982/ECTA11380.

Blanco, E., Lopez, M. C. and Villamayor-Tomas, S. (2015) ‘Exogenous degradation in the commons: Field experimental evidence’, *Ecological Economics*. Elsevier B.V., 120, pp. 430–439. doi: 10.1016/j.ecolecon.2015.03.028.

Bochet, O. *et al.* (2019) ‘Collective risk-taking in the commons’, *Journal of Economic Behavior & Organization*, 163, pp. 277–296.

Bock, O., Baetge, I. and Nicklisch, A. (2014) ‘Hroot: Hamburg Registration and Organization Online Tool’, *European Economic Review*. Elsevier, 71, pp. 117–120. doi: 10.1016/j.euroecorev.2014.07.003.

Boettiger, C., Ross, N. and Hastings, A. (2013) 'Early warning signals: The charted and uncharted territories', *Theoretical Ecology*, 6(3), pp. 255–264. doi: 10.1007/s12080-013-0192-6.

Bohannon, J. (2016) 'Mechanical Turk upends social sciences', *Science*, 352(6291), pp. 1263–1264. doi: 10.1126/science.352.6291.1263.

Brandt, G. *et al.* (2017) 'OGUMI—A new mobile application to conduct common-pool resource experiments in continuous time', *PLoS ONE*, 12(6). doi: 10.1371/journal.pone.0178951.

Brehmer, B. (1992) 'Dynamic decision making: Human control of complex systems', *Acta Psychologica*, 81(3), pp. 211–241. doi: 10.1016/0001-6918(92)90019-A.

Brown, T. and Kroll, S. (2017) 'Avoiding an uncertain catastrophe: climate change mitigation under risk and wealth heterogeneity', *Climatic Change*. Climatic Change, 141(2), pp. 155–166. doi: 10.1007/s10584-016-1889-5.

Bruno Soares, M., Alexander, M. and Dessai, S. (2018) 'Sectoral use of climate information in Europe: A synoptic overview', *Climate Services*, 9, pp. 5–20. doi: 10.1016/j.cliser.2017.06.001.

Budescu, D. V. *et al.* (2014) 'The interpretation of IPCC probabilistic statements around the world', *Nature Climate Change*. doi: 10.1038/nclimate2194.

Budescu, D. V., Rapoport, A. and Suleiman, R. (1990) 'Resource dilemmas with environmental uncertainty and asymmetric players', *European Journal of Social Psychology*, 20(6), pp. 475–487. doi: 10.1002/ejsp.2420200603.

Budescu, D. V., Rapoport, A. and Suleiman, R. (1992) 'Simultaneous vs. sequential requests in resource dilemmas with incomplete information', *Acta Psychologica*, 80(1–3), pp. 297–310. doi: 10.1016/0001-6918(92)90052-F.

Camerer, C. F. *et al.* (2016) 'Evaluating replicability of laboratory experiments in economics', *Science*, 351(6280), pp. 1433–1436. doi: 10.1126/science.aaf0918.

Camerer, C. F. *et al.* (2018) 'Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015', *Nature Human Behaviour*, 2, pp. 637–644. doi: 10.1038/s41562-018-0399-z.

Cardenas, J.-C. and Carpenter, J. (2008) 'Behavioural Development Economics : Lessons from Field Labs in the Developing World', *The Journal of Development Studies*, 44(3), pp. 311–338. doi: 10.1080/00220380701848327.

Cardenas, J., Janssen, M. A. and Bousquet, F. (2013) 'Dynamics of Rules and Resources: Three New Field Experiments on Water, Forests and Fisheries', in *Handbook on Experimental Economics and the Environment*, pp. 1–27. doi: 10.4337/9781781009079.00020.

Cerutti, N. and Schlüter, A. (2019) 'Resource changes: exogenous or endogenous, gradual or abrupt. Experimental evidence', *International Journal of Environmental Studies*. Routledge, pp. 1–15. doi: 10.1080/00207233.2019.1644022.

Chancey, E. T. *et al.* (2015) 'False Alarms vs. Misses: Subjective Trust as a Mediator Between Reliability and Alarm Reaction Measures', in *Proceedings of the Human Factors and Ergonomics Society*. doi: 10.1177/1541931215591141.

Chancey, E. T. *et al.* (2017) 'Trust and the Compliance–Reliance Paradigm: The Effects of Risk, Error Bias, and Reliability on Trust and Dependence', *Human Factors*. SAGE PublicationsSage CA: Los Angeles, CA, 59(3), pp. 333–345. doi: 10.1177/0018720816682648.

Cinner, J. E. *et al.* (2011) 'Responding to change: Using scenarios to understand how socioeconomic factors may influence amplifying or dampening exploitation feedbacks among Tanzanian fishers', *Global Environmental Change*. Elsevier Ltd, 21(1), pp. 7–12. doi: 10.1016/j.gloenvcha.2010.09.001.

Coffman, L. C. and Niederle, M. (2015) 'Pre-analysis plans have limited upside, especially where replications are feasible', *Joiurnal of Economic Perspective*, 29(3), pp. 81–98. doi: 10.1257/jep.29.3.81.

Crépin, A.-S. *et al.* (2012) 'Regime shifts and management', *Ecological Economics*, 84, pp. 15–22. doi: 10.1016/j.ecolecon.2012.09.003.

Dannenberg, A. *et al.* (2015) 'On the Provision of Public Goods with Probabilistic and Ambiguous Thresholds', *Environmental and Resource Economics*, 61(3), pp. 365–383. doi: 10.1007/s10640-014-9796-6.

Dannenberg, A. and Tavoni, A. (2016) 'Collective action in dangerous climate change games', in Botelho, A. (ed.) *The WSPC Reference on Natural Resources and Environmental Policy in the Era of Global Change*. Singapore: World Scientific Publishing, pp. 95–120. doi: 10.1142/9747.

Darbyshire, R. *et al.* (2020) 'Insights into the value of seasonal climate forecasts to agriculture', *Australian Journal of Agricultural and Resource Economics*, 64(4), pp. 1034–1058. doi: 10.1111/1467-8489.12389.

Diekert, F. K. (2017) 'Threatening thresholds? The effect of disastrous regime shifts on the non-cooperative use of environmental goods and services', *Journal of Public Economics*, 147(1), pp. 30–49. doi: 10.1016/j.jpubeco.2017.01.004.

Dow, K. and Cutter, S. L. (1998) 'Crying Wolf: Repeat Responses to Hurricane Evacuation Orders', *Coastal Management*, 26, pp. 237–252.

Dowdy, A. J. *et al.* (2019) 'Future changes in extreme weather and pyroconvection risk factors for Australian wildfires', *Scientific Reports*. Springer US, 9(1), pp. 1–11. doi: 10.1038/s41598-019-46362-x.

Falk, A. and Heckman, J. J. (2009) 'Lab experiments are a major source of knowledge in the social sciences', *Science*, 326(5952), pp. 535–538. doi: 10.1126/science.1168244.

Folke, C. *et al.* (2010) 'Resilience Thinking: Integrating Resilience, Adapatability, and Transformability', *Ecology and Society*, 15(4), p. 20.

Gatiso, T. T., Vollan, B. and Nuppenau, E. (2015) 'Resource scarcity and democratic elections in commons dilemmas : An experiment on forest use in Ethiopia', *Ecological Economics*. Elsevier B.V., 114, pp. 199–207. doi: 10.1016/j.ecolecon.2015.04.005.

Gibson, C. C. (2001) 'Forest resources: institutions for local governance in Guatemala', in Burger, J. et al. (eds) *Protecting the Commons: A Framework For Resource Management In The Americas*. Washington, DC, USA: Island Press, pp. 71–89.

Greiner, B. (2015) 'Subject pool recruitment procedures : organizing experiments with ORSEE', *Journal of the Economic Science Association*. Springer US, 1(1), pp. 114–125. doi: 10.1007/s40881-015-0004-4.

Gustafsson, M., Biel, A. and Gärling, T. (1999) 'Overharvesting of resources of unknown size', *Acta Psychologica*, 103, pp. 47–64. Available at: https://ac.els-cdn.com/S0001691899000244/1-s2.0-S0001691899000244-main.pdf?_tid=7afd0ec9-7901-4eed-8f31-ec5fb0f1f259&acdnat=1520848015_5ce694f2f84a0c70cbcc3581e62d25fe.

Hardin, G. (1968) 'The Tragedy of the Commons', *Science*, 162(June), pp. 1243–1248.

Harrison, G. W. and List, J. A. (2004) 'Field Experiments', *Journal of Economic Literature*, 42(4), pp. 1009–1055. Available at: https://www.jstor.org/stable/3594915.

Henrich, J., Heine, S. J. and Norenzayan, A. (2010) 'The weirdest people in the world?', *Behavioral and Brain Sciences*, 33(2–3), pp. 61–83. doi: 10.1017/S0140525X0999152X.

Hine, D. W. and Gifford, R. (1996) 'Individual Restraint and Group Efficiency in Commons Dilemmas: The Effects of Two Types of Environmental Uncertainty', *Journal of Applied Social Psychology*, 26(11), pp. 993–1009. doi: 10.1111/j.1559-1816.1996.tb01121.x.

IPCC (2014) *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]*. IPCC, Geneva, Switzerland, 151 pp. doi: 10.1017/CBO9781107415324.

IPCC (2019) *Summary for Policymakers*, *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*. Edited by J. M. P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.- O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick,. doi: 10.4337/9781784710644.

Janssen, M. A. (2010) 'Introducing Ecological Dynamics into Common-Pool Resource Experiments', *Ecology and Society*, 15(2). Available at: http://www.ecologyandsociety.org/vol15/iss2/art7/ (Accessed: 10 July 2017).

Janssen, M. A. *et al.* (2010) 'Lab Experiments for the Study of Social-Ecological Systems', *Science*, 328(April), pp. 613–617.

Janssen, M. A., Lindahl, T. and Murphy, J. J. (2015) 'Advancing the understanding of behavior in social-ecological systems: Results from lab and field experiments', *Ecology and Society*, 20(4). doi: 10.5751/ES-08097-200434.

Kimbrough, E. O. and Vostroknutov, A. (2015) 'The social and ecological determinants of common pool resource sustainability', *Journal of Environmental Economics and Management*, 72, pp. 38–53. doi: 10.1016/j.jeem.2015.04.004.

Knudson, C. and Guido, Z. (2019) 'The missing middle of climate services: layering multiway, two-way, and one-way modes of communicating seasonal climate forecasts', *Climatic Change*, 157(1), pp. 171–187. doi: 10.1007/s10584-019-02540-4.

Lade, S. J. *et al.* (2013) 'Regime shifts in a social-ecological system', *Theoretical Ecology*, 6(3), pp. 359–372. doi: 10.1007/s12080-013-0187-3.

LeClerc, J. E. and Joslyn, S. (2015) 'The cry wolf effect and weather-related decision making', *Risk Analysis*, 35(3), pp. 385–395. doi: 10.1111/risa.12336.

Levitt, S. D. and List, J. A. (2007) 'What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?', *Journal of Economic Perspectives*, 21(2), pp. 153–174.

Lim, J. R., Fisher Liu, B. and Egnoto, M. (2019) 'Cry Wolf Effect ? Evaluating the Impact of False Alarms on Public Responses to Tornado Alerts in the Southeastern United States', *Weather, Climate, and Society*, 11, pp. 549–563. doi: 10.1175/WCAS-D-18-0080.1.

Lindahl, T., Crépin, A.-S. and Schill, C. (2016) 'Potential Disasters can Turn the Tragedy into Success', *Environmental and Resource Economics*. Springer Netherlands, 65(3), pp. 657–676. doi: 10.1007/s10640-016-0043-1.

Losee, J. E. and Joslyn, S. (2018) ‘The need to trust: How features of the forecasted weather influence forecast trust’, *International Journal of Disaster Risk Reduction*. Elsevier Ltd, 30, pp. 95–104. doi: 10.1016/j.ijdrr.2018.02.032.

Maas, A. *et al.* (2017) ‘Dilemmas, coordination and defection: How uncertain tipping points induce common pool resource destruction’, *Games and Economic Behavior*. Elsevier Inc., 104, pp. 760–774. doi: 10.1016/j.geb.2017.06.009.

Manzey, D., Gérard, N. and Wiczorek, R. (2014) ‘Decision-making and response strategies in interaction with alarms : the impact of alarm reliability , availability of alarm validity information and workload’, *Ergonomics*, 57(12), pp. 1833–1855. Available at: https://doi.org/10.1080/00140139.2014.957732.

McAllister, R. R. J. *et al.* (2011) ‘Economic Behavior in the Face of Resource Variability and Uncertainty’, *Ecology and Society*, 16(3). doi: 10.5751/ES-04075-160306.

Milinski, M. *et al.* (2008) ‘The Collective-Risk Social Dilemma and the Prevention of Simulated Dangerous Climate Change’, *Proceedings of the National Academy of Sciences of the United States of America*, 105(7), pp. 2291–2294.

Millenium Ecosystems Assessment (2005) *Millenium Ecosystem Assessment. Ecosystems and Human Well-being: Synthesis*, *Island, Washington DC*. doi: 10.1196/annals.1439.003.

Moffatt, P. G. (2021) ‘Experimetrics : A Survey’, *Foundations and Trends(R) in Econometrics*, 11(1–2), pp. 1–152. doi: 10.1561/0800000035.

National Institute of Water and Atmospheric Research (NIWA) (2016) *Improving seasonal climate forecasts*, *2016-02-08*. Available at: https://niwa.co.nz/news/improving-seasonal-climate-forecasts (Accessed: 11 March 2020).

Nie, Z., Yang, X. and Tu, Q. (2020) ‘Resource scarcity and cooperation: Evidence from a gravity irrigation system in China’, *World Development*. Elsevier Ltd, 135. doi: 10.1016/j.worlddev.2020.105035.

NSW Rural Fire Service (2020) *NSW RURAL FIRE SERVICE Prepare your home*. Available at: https://www.rfs.nsw.gov.au/plan-and-prepare/prepare-your-property (Accessed: 16 September 2020).

Nyborg, K. *et al.* (2016) ‘Social norms as solutions’, *Science*, 354(6308), pp. 42–43. doi: 10.1126/science.aaf8317.

Olken, B. A. (2015) ‘Promises and perils of pre-analysis plan’, *Journal of Economic Perspectives*, 29(3), pp. 61–80. doi: 10.1257/jep.29.3.61.

Oprea, R., Charness, G. and Friedman, D. (2014) ‘Continuous time and communication in a public-goods experiment’, *Journal of Economic Behavior & Organization*, 108, pp. 212–223. doi: 10.1016/j.jebo.2014.09.012.

Osés-Eraso, N., Udina, F. and Viladrich-Grau, M. (2008) ‘Environmental versus Human-Induced Scarcity in the Commons: Do They Trigger the Same Response?’, *Environmental and Resource Economics*, 40, pp. 529–550. doi: 10.1007/s10640-007-9168-6.

Osés-Eraso, N. and Viladrich-Grau, M. (2007) ‘Appropriation and concern for resource scarcity in the commons: An experimental study’, *Ecological Economics*, 63(2–3), pp. 435–445. doi: 10.1016/j.ecolecon.2006.11.016.

Ostrom, E. (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*. New York, NY: Cambridge University Press.

Ostrom, E. (2006) 'The value-added of laboratory experiments for the study of institutions and common-pool resources', *Journal of Economic Behavior and Organization*, 61(2), pp. 149–163. doi: 10.1016/j.jebo.2005.02.008.

Ostrom, E. (2009) 'A General Framework for Analyzing Sustainability of Social-Ecological Systems', *Science*, 325(5939), pp. 419–422. doi: 10.1126/science.1172133.

Ostrom, E., Gardner, R. and Walker, J. (1994) *Rules, Games and Common-pool Resources*, *Journal of Chemical Information and Modeling*. Ann Arbor, Michigan: University of Michigan Press.

Ostrom, E., Walker, J. and Gardner, R. (1992) 'Covenants With and Without a Sword: Self-Governance is Possible', *The American Political Science Review*, 86(2), pp. 404–417. Available at: http://links.jstor.org/sici?sici=0003-0554%28199206%2986%3A2%3C404%3ACWAWAS%3E2.0.CO%3B2-V (Accessed: 17 October 2017).

Palan, S. and Schitter, C. (2018) 'Prolific.ac—A subject pool for online experiments', *Journal of Behavioral and Experimental Finance*. Elsevier B.V., 17, pp. 22–27. doi: 10.1016/j.jbef.2017.12.004.

Palmer, J. C. and Strickland, J. (2016) *A beginner's guide to crowdsourcing*, *Psychological Science Agenda*. Available at: http://www.apa.org/science/about/psa/2016/06/changing-minds.

Peer, E. *et al.* (2017) 'Beyond the Turk: Alternative platforms for crowdsourcing behavioral research', *Journal of Experimental Social Psychology*. Elsevier Inc., 70, pp. 153–163. doi: 10.1016/j.jesp.2017.01.006.

Perman, R. *et al.* (2011) *Natural Resource and Environmental Economics*. 4th edn. Pearson Education.

Pettit, J. *et al.* (2014) 'Software for continuous game experiments', *Experimental Economics*, 17, pp. 631–648. doi: 10.1007/s10683-013-9387-3.

Pfaff, A. *et al.* (2015) 'Framed field experiment on resource scarcity & extraction: Path-dependent generosity within sequential water appropriation', *Ecological Economics*. doi: 10.1016/j.ecolecon.2015.06.002.

del Pilar Moreno-Sánchez, R. and Maldonado, J. H. (2010) 'Evaluating the role of co-management in improving governance of marine protected areas: An experimental approach in the Colombian Caribbean', *Ecological Economics*. Elsevier B.V., 69(12), pp. 2557–2567. doi: 10.1016/j.ecolecon.2010.07.032.

Polasky, S., Zeeuw, A. De and Wagener, F. (2011) 'Optimal management with potential regime shifts', *Journal of Environmental Economics and Management*. Elsevier, 62(2), pp. 229–240. doi: 10.1016/j.jeem.2010.09.004.

Poteete, A. R., Janssen, M. A. and Ostrom, E. (2010) *Working Together: Collective Action, the Commons, and Multiple Methods in Practice*, *Princeton University Press*. Princeton, NJ. doi: 10.2307/j.ctt3fgxrd.

Prediger, S., Vollan, B. and Herrmann, B. (2014) 'Resource scarcity and antisocial behavior', *Journal of Public Economics*. Elsevier B.V., 119, pp. 1–9. doi: 10.1016/j.jpubeco.2014.07.007.

Prolific (2021) *Prolific.co*. Oxford, UK. Available at: https://www.prolific.co/ (Accessed: 14 May 2021).

Rapoport, A. *et al.* (1992) *Social dilemmas with uniformly distributed resources.* Internatio. Edited by W. B. G. Liebrand, D. M. Messick, and H. A. M. Wilke. Elmsford, NY, US: Pergamon Press. Available at: https://psycnet.apa.org/record/1993-98342-001.

Readfearn, G. (2020) 'Explainer : what are the underlying causes of Australia ' s shocking bushfire season ? Why has this bushfire season been so devastating ? What has happened to Australia ' s fire weather ?', *The Guardian*, 12 January, pp. 1–3. Available at: https://www.theguardian.com/environment/2020/jan/13/explainer-what-are-the-underlying-causes-of-australias-shocking-bushfire-season.

Ripberger, J. T. *et al.* (2015) 'False Alarms and Missed Events: The Impact and Origins of Perceived Inaccuracy in Tornado Warning Systems', *Risk Analysis*, 35(1), pp. 44–56. doi: 10.1111/risa.12262.

Rocha, J. C. *et al.* (2020) 'Cooperation in the face of thresholds, risk, and uncertainty: Experimental evidence in fisher communities from Colombia', *PLoS ONE*, 15(12 December), pp. 1–14. doi: 10.1371/journal.pone.0242363.

Rutte, C. G., Wilke, H. A. M. and Messick, D. M. (1987) 'Scarcity or Abundance Caused by People or the Environment as Determinants of Behavior in the Resource Dilemma', *JOURNAL OF EXPERIMENTAL SOCIAL PSYCHOLOGY*, 23, pp. 208–216. Available at: https://ac.els-cdn.com/0022103187900321/1-s2.0-0022103187900321-main.pdf?_tid=91c3c340-b341-11e7-b07e-00000aacb35e&acdnat=1508248166_e5159cd90778692408ac1195af77e11a (Accessed: 17 October 2017).

Scheffer, M. *et al.* (2001) 'Catastrophic shifts in ecosystems', *Nature*, 413(6856), pp. 591–596. doi: 10.1038/35098000.

Scheffer, M. (2009) 'Critical Transitions in Nature and Society', *Princeton University Press*, p. 384. doi: 10.5860/CHOICE.47-1380.

Scheffer, M. *et al.* (2012) 'Anticipating Critical Transitions', *Science*, 338(6105), pp. 344–348. doi: 10.1126/science.1225244.

Schill, C., Lindahl, T. and Crépin, A.-S. (2015) 'Collective action and the risk of ecosystem regime shifts: Insights from a laboratory experiment', *Ecology and Society*, 20(1). doi: 10.5751/ES-07318-200148.

Schill, C. and Rocha, J. C. (2019) *Uncertainty Can Help Protect Local Commons in the Face of Climate Change*, *SSRN Electronic Journal*. doi: 10.2139/ssrn.3468677.

Schlüter, M. *et al.* (2017) 'A framework for mapping and comparing behavioural theories in models of social-ecological systems', *Ecological Economics*. Elsevier B.V., 131, pp. 21–35. doi: 10.1016/j.ecolecon.2016.08.008.

Schlüter, M., Tavoni, A. and Levin, S. A. (2016) 'Robustness of norm-driven cooperation in the commons', *Proceedings of the Royal Society B: Biological Sciences*, 283. doi: 10.1098/rspb.2015.2431.

Schultz, D. M. *et al.* (2010) 'Decision making by Austin, Texas, residents in hypothetical tornado scenarios', *Weather, Climate, and Society*, 2(3), pp. 249–254. doi: 10.1175/2010WCAS1067.1.

Simmons, K. M. and Sutter, D. (2009) 'False alarms, tornado warnings, and tornado casualties', *Weather, Climate, and Society*, 1(1), pp. 38–53. doi: 10.1175/2009WCAS1005.1.

Smith, V. L. (1982) 'Microeconomic Systems as an Experimental Science', *American Economic Review*, 72, pp. 923–955.

Sunstein, C. R. (2010) 'Irreversibility', *Law, Probability and Risk*, 9(3–4), pp. 227–245. doi: 10.1093/lpr/mgq010.

Tavoni, A. and Levin, S. (2014) 'Managing the climate commons at the nexus of ecology, behaviour and economics', *Nature Climate Change*, 4(12), pp. 1057–1063. doi: 10.1038/nclimate2375.

Taylor, A. L., Dessai, S. and Bruine De Bruin, W. (2015) 'Communicating uncertainty in seasonal and interannual climate forecasts in Europe', *Philosophical Transactions of the Royal Society A*, 373(20140454). doi: 10.1098/rsta.2014.0454.

Taylor, A. L., Kox, T. and Johnston, D. (2018) 'Communicating high impact weather: Improving warnings and decision making processes', *International Journal of Disaster Risk Reduction*, 30, pp. 1–4. doi: 10.1016/j.ijdrr.2018.04.002.

Trainor, J. E. *et al.* (2015) 'Tornadoes , Social Science , and the False Alarm Effect', *Weather, Climate, and Society*, 7, pp. 333–352. doi: 10.1175/WCAS-D-14-00052.1.

Wagener, F. and de Zeeuw, A. (2021) 'Stable partial cooperation in managing systems with tipping points', *Journal of Environmental Economics and Management*. Elsevier Inc., 109(June), p. 102499. doi: 10.1016/j.jeem.2021.102499.

Walker, J. M. and Gardner, R. (1992) 'Probabilistic Destruction of Common-pool Resources: Experimental Evidence', *The Economic Journal*, 102(414), pp. 1149–1161. doi: http://dx.doi.org/10.2307/2234382.

Webber, S. (2019) 'Putting climate services in contexts: advancing multi-disciplinary understandings: introduction to the special issue', *Climatic Change*. Climatic Change, 157(1), pp. 1–8. doi: 10.1007/s10584-019-02600-9.

Wharton Credibility Lab (2017) *AsPredicted*. Available at: https://aspredicted.org (Accessed: 20 October 2021).

Whitmer, D. E. *et al.* (2017) 'Reconsidering the " Conventional Wisdom " of the Relationship between False Alarms and Risk Beliefs', in *Proceedings of the Human Factors and Ergonomics Society*, pp. 1419–1423. doi: 10.1177/1541931213601838.

Wiczorek, R. and Meyer, J. (2016) 'Asymmetric effects of false positive and false negative indications on the verification of alerts in different risk conditions', *Proceedings of the Human Factors and Ergonomics Society*, pp. 289–292. doi: 10.1177/1541931213601066.

Zommers, Z. (2012) *Climate Early Warning System Feasibility Report: Early Warning Systems and Hazard Prediction*. Available at: http://www.unep.org/science/chief-scientist/Portals/76/docs/EWS_Report_Final_Version1.pdf.

Chapter 2 The interaction of shock experience and threshold knowledge in natural resource management

Katharina Hembach-Stunden[a], Aneeque Javaid[b], Stefanie Engel[a]

[a] School of Business Administration and Economics and Institute for Environmental Systems Research (IUSF), Osnabrueck University, Germany.

[b] Mercator Research Institute on Global Commons and Climate Change (MCC), Germany.

Abstract: Climate change increases the frequency of extreme weather events which increases the risk of exogenous shocks to resource systems. Simultaneously, cooperation and coordination amongst resource users is needed to manage natural resources sustainably and prevent overexploitation beyond critical thresholds causing resources to collapse. Previous experimental studies examining resource users' extraction behaviour separately analysed the impact of either scarcity due to exogenous shocks or the endogenous threat of crossing a critical threshold. Our present study addresses the interaction of both exogenous and endogenous dynamics. We developed a novel (quasi-) continuous-time common-pool resource experiment to analyse how an exogenous shock affects cooperation within groups of resource users facing a critical threshold. We hypothesise that experiencing exogenous shocks, which diminish a resource, potentially undermines the coordinating effect of threshold knowledge that has been demonstrated in previous studies, and reduces cooperation in terms of extraction benefits obtained. We disentangle whether the shock effect is due to the *disruption effect* caused by the shock experience, or merely combined with exposure to resource scarcity (*resource size effect*). Further, we test for a *spillover effect*, if a prior shock experience affects future behaviour. We find that experiencing a shock in a setting with prior threshold knowledge significantly decreases cooperation, but has no significant impact on coordination. While the *resource size effect* partly explains the negative shock effect, we find that the additional *disruption effect* is positive such that subjects who experience the shock cooperate significantly more than subjects who are exposed to initial resource scarcity. In contrast to previous studies, we do not find statistically significant evidence of a *spillover effect* of shock experience in the past on future coordination and cooperation. These results further our understanding of the impact of ecological dynamics on natural resource management.

Keywords: collective action, common-pool resources, tipping points, regime shift, laboratory experiment

2.1 Introduction

The sustainable management of natural resources is threatened by the consequences of climate change. Current evidence indicates that climate change has increased the frequency of extreme weather events, such as droughts or floods, and thereby the risk of exogenous ecological shocks to ecosystems (IPCC, 2014). Such exogenous shocks often have a direct impact on local resources by destroying parts of them, causing resource scarcity (Scheffer et al., 2001). Thus, exogenous shocks due to extreme weather events increase the pressure on resource users that rely on resource extraction and the exploitation of ecosystems for their well-being (Nyborg et al., 2016).

In addition to exogenous shocks, resource users' overexploitation imposes an endogenous threat to local resources by driving gradual change in ecosystem's underlying conditions (Polasky et al., 2011; Scheffer et al., 2001). Depending on an ecosystem's dynamics, the ecosystem may respond smoothly to gradual change until it reaches a *critical threshold* (or *tipping point*). Crossing this critical threshold leads to a sudden switch towards an unfavourable alternative state (*regime shift*), in extreme cases causing a collapse of the ecosystem and thus, the resource (Biggs et al., 2009; Scheffer et al., 2001).[7] The collapse of a resource due to extreme weather events and/or overexploitation of resources is a potential driver of migration of resource users within and across borders (IPCC, 2019). This migration of resource users might cause additional negative spillover effects of exogenous shock experiences to stable resource systems elsewhere.

Multiple field studies have found that resource users' behaviour and their institutional arrangements depend on the ecological dynamics of the resource users' environment (Cardenas et al., 2013; McAllister et al., 2011; Pfaff et al., 2015; Prediger et al., 2014). More

[7] Our study focuses on the impact of negative regime shifts, where crossing a critical threshold and the shift to an alternative state implies a strong reduction in the resource level.

specifically, the level of cooperation amongst group members depends on, amongst other factors, the given resource level and whether that resource level is given by exogenous or endogenous dynamics (Ostrom, 2009; Rutte et al., 1987). However, previous experimental studies on resource users' extraction behaviour analysed the impact of either scarcity due to exogenous shocks (Blanco, Lopez and Villamayor-Tomas, 2015; Cerutti and Schlüter, 2019) or the endogenous threat of crossing a critical threshold separately (Schill, Lindahl and Crépin, 2015; Lindahl, Crépin and Schill, 2016). Thus, they neglect the potential interaction between exogenous and endogenous resource dynamics.

To address this gap in the literature and to extend the understanding of the interaction between human behaviour and ecological dynamics in natural resource systems, we assessed how an unexpected exogenous shock that causes resource scarcity affects subjects' extraction strategies if the resource's critical threshold, at which the resource collapses, is common knowledge. Based on related literature, we hypothesise that the experience of an exogenous shock increases subjects' extraction and decreases *cooperation* (Blanco et al., 2015; Cerutti and Schlüter, 2019; Cinner et al., 2011; Pfaff et al., 2015) even with prior knowledge of a threshold. We further hypothesise that subjects' increased extraction due to the experience of an exogenous shock (ibid.) increases the probability of failed *coordination* and therefore, increases the number of endogenously driven resource collapses.

To test these hypotheses, we designed a novel, (quasi-) continuous-time common-pool resource (CPR) experiment for the laboratory where the resource is developing continuously over time. In each session, subjects were divided in groups of four managing a joint resource. The three experimental treatments all incorporate the endogenous threat of crossing a critical threshold through overexploitation that, once reached, the resource immediately and irreversibly collapses. We are interested in two types of group outcomes: (1) the collapse of the resource as an indicator of failed *coordination* and (2) the group's total resource extraction in relation to the socially optimal extraction as a measure of *cooperation* amongst resource users.

We analyse whether the impact of the shock on both coordination and cooperation can be explained by the mere increase in resource scarcity induced by the shock (which we call the *resource size effect*) or whether there is an additional effect from the experience of an unexpected, sudden drop of the resource level (which we call the *disruption effect*). To distinguish these two effects, our three experimental treatments vary with respect to the initial resource level and shock implementation in the first round.

We also test whether the shock experience has a *spillover effect* on future coordination and cooperation. This is motivated by the idea that resource users may have to migrate to other intact resource systems elsewhere to sustain their livelihood after a shock. To achieve this, subjects are matched in new groups with strangers for a second round that mirrors the control treatment for all groups. We analyse whether the shock experience affects outcomes in the new location by comparing second-round outcomes between groups that experienced a shock in the first round and those who did not.

This research contributes to the body of work on the interaction of human behaviour and ecological dynamics in natural resource management in several ways. Firstly, to our knowledge this is currently the only study to analyse the impact of an exogenous shock in the presence of a known critical threshold. Secondly, we distinguish whether the impact of the exogenous shock on groups' extraction strategies is due to the immediate experience of the shock or due to the resource scarcity itself, and assess a spillover effect of shock experience. And thirdly, we contribute to the growing literature incorporating ecological complexities in a controlled manner into experimental designs, specifically CPR experiments that allow for the continuous-time nature of resource users' extraction choices (Brandt et al., 2017; Cerutti and Schlüter, 2019; Janssen, 2010).

2.2 Experimental setting

The basis of our study is a novel (quasi-) continuous-time CPR game that incorporates a known critical threshold. The experiment is based on a continuous-time design because it has multiple advantages compared to round-based CPR designs. Firstly, time continuity allows for a dynamic resource development and asynchronous, strategic interaction amongst group members (Pettit et al., 2014). Furthermore, time continuity offers the chance to approach long-term interaction amongst subjects in a relatively short time period because it speeds up the process of adjustment dynamics (ibid). Due to technical restrictions, the resource development and subjects' extraction choices are calculated every second, which limits the continuity of the game's development. Thus, it is a quasi-continuous-time game (Bigoni et al., 2015).

Our CPR game is based on, and an advancement of the experimental set-up presented in Brandt et al. (2017). We programmed our CPR game in the experimental software SoPHIE (Hendriks, 2012) and incorporated more complex resource dynamics, such as the critical threshold and the exogenous shock.[8] The main experiment consists of two CPR rounds before which subjects were randomly matched in groups of four (hereinafter referred to as CPR1 and CPR2). We implemented a perfect strangers' matching such that subjects only ever interacted once with each other to reduce social learning effects and avoid strategic interactions throughout all parts of the experiment.

At the beginning of the experimental session, subjects received the instructions of the experiment and had to answer a series of control questions before and after playing two test rounds to secure their understanding of the CPR game. The resource dynamics during the two test rounds were equal to the dynamics in the CPR group rounds, besides the fact that subjects played the test rounds by themselves while the computer simulated the other group members. The two test rounds were not payout-relevant, but allowed subjects to get familiarised with the dynamics of the renewable resource.

---

[8] Overall, the experiment consists of two payout-relevant parts: (1) the CPR game and (2) a game to elicit subjects' social preferences (SVO slider measure, Murphy et al., 2011). Here, we focus on the description of the CPR game, since the outcome of the social preference game is not part of the main analysis reported in this paper. Please see Appendix 2A for further details.

During the CPR game, subjects individually extracted resource units from their group's shared resource. Subjects could observe the resource development over time (in seconds) by looking at a real-time graph on screen and made their extraction decisions simultaneously (Fig. 2.1). Subjects could increase or decrease their individual extraction level at any point by adjusting their extraction slider's position between 0 and 10 (integer numbers). Once chosen, the extraction level was executed each second until subjects changed it again. Subjects received perfect feedback about the development of their individual extraction level, their *individual total extraction* as well as their *group's total extraction*. Thereby, subjects could constantly compare their own earnings to the earnings of the other group members.

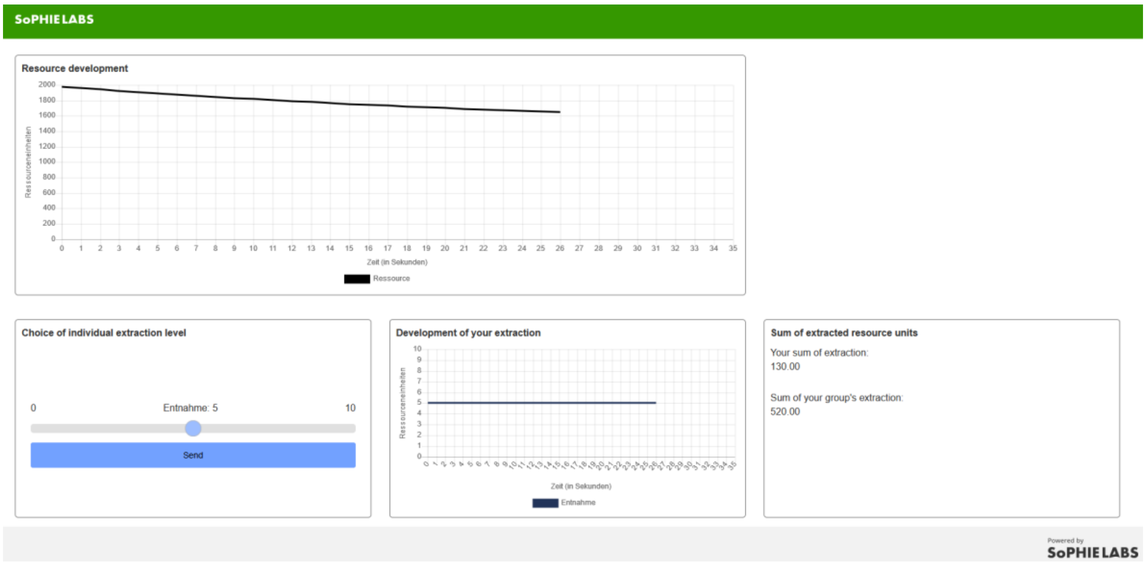


Fig. 2.1 Screenshot of the CPR game (English translation of German original): The graph at the top presents the resource development over time (in seconds). Subjects use a slider to choose their extraction level between 0 and 10 (bottom row, left). Subjects can observe the development of their extraction level over time (bottom row, centre) and their individual total extraction as well as their group's total extraction (bottom row, right).

Subjects knew that their individual total extraction from one of the two CPR group rounds was randomly chosen for the payout of the CPR game in the end. The points were converted with an exchange rate of 100 points = 0.40 Euro. Please see Appendix 2A for details of the experimental design.

2.2.1 User-resource model

The resource development in our CPR game is based on a simple logistic growth model as presented in Perman *et al.* (2011) and also used in Brandt et al. (2017). Further details of the model and the chosen parameters are provided in Appendix 2B.

Equation (2.1) below describes the dynamics of the resource R_t that changes over time t (measured in seconds). The logistic growth term describes the natural growth of the resource, with the resource growth rate $g = 0.04$ and the maximum resource level at the carrying capacity $R_{max} = 2{,}000$ resource units. We implement the threat of a critical threshold by integrating an irreversible and persistent collapse of the resource once subjects' extraction drives the current resource level R_t to the certain threshold $R_{min} = 500$ resource units ($R_{t+1} = 0\ if\ R_t \leq R_{min}$). The resource R_t then stays at zero resource units infinitely since the natural growth of the resource also collapses once the threshold is reached ($gR_t\left(1 - \frac{R_t}{R_{max}}\right) = 0\ if\ R_t \leq R_{min}$). Thus:

$$R_{t+1} = \begin{cases} R_t + gR_t\left(1 - \dfrac{R_t}{R_{max}}\right) - \displaystyle\sum_{i=1}^{n} E_{it} & if\ R_t > R_{min} \\ 0 & if\ R_t \leq R_{min} \end{cases} \tag{2.1}$$

where $\sum_{i=1}^{n} E_{it}$ denotes the joint extraction of a group with $n = 4$ resource users per second t.

The total return B for an individual subject i in the CPR round is given by the sum of resource units that i extracts over time t, with t_{end} denoting the last second of the round:

$$B_i = \sum_{t=0}^{t_{end}} E_{it} \tag{2.2}$$

Subjects' are informed that the expected return for the period after a collapse is zero because $E_{it} = 0\ if\ R_t \leq R_{min}$. In our design, groups that cause a collapse of the resource have to wait until the end of the round without collecting any more resource units for their payout. That way, subjects' incentives are not distorted by the idea that an early depletion of the resource would allow them to leave the experimental session earlier.

The resource's natural growth is highest with a regrowth of 20 resource units per second at the maximum sustainable yield (MSY) $R_{MSY} = 1{,}000$ resource units. If the group's joint extraction $\sum_{i=0}^{n} E_i$ equals the resource's natural growth at MSY with $E_{MSY} = 20\ if\ R_t = R_{MSY}$, the resource can be extracted indefinitely at the steady state of R_{MSY}. This is the socially optimal extraction level for the group since there is no extraction cost in this game. Please see Appendix 2B for details. Optimal extraction paths are discussed in detail below (Section 2.2.2).

Subjects know that the resource development is calculated every second, and that the level of resource regrowth depends on their group's extraction choices. All subjects in all treatments know of the critical threshold at $R_{min} = 500$ and that the resource and regrowth collapse if their extraction drives the resource to R_{min}.

Instead of showing subjects the exact growth function in the instructions, we present the regrowth levels corresponding to resource levels between 2,000 and 500 resource units in steps of 100 in a table (see Table 2A.1 in Appendix 2A). Further, we include a graph of the concave growth function and a graph that represented a collapsed resource to visualise the the resource's dynamic when reaching the critical threshold and the drop in the regrowth rate (see Fig. A.2 and A.3 in Appendix 2A). This set-up of the instructions is inspired by the instructions of Schill et al. (2015).

The end time of each round is determined randomly and unknown to subjects to avoid the depletion of the resource due to an endgame effect (Cerutti and Schlüter, 2017; Janssen, 2010). We implement a random continuation rule (Dal Bó and Fréchette, 2018), which is equivalent to an infinite game horizon and allows incentivising the extraction of the resource at R_{MSY} in each of the three treatments. Beyond the certain round length, every 10 seconds there is a chance of 10% that the round ends. Thus, the round continues for another block of 10 seconds with a probability of 90%.

2.2.2 Treatment design

In the first of the CPR rounds (CPR1), we implemented three treatments in a between-subject design, which differed in four aspects: (1) the initial resource level, (2) a pause after 25 seconds, (3) the resource dynamics at the time of the pause and (4) the minimum duration of the round that is known to subjects with certainty (Table 2.1). Treatments were assigned by session.

Table 2.1. Overview of the treatment design.

four differences between treatments in CPR1:	**T1 Control**	**T2 Exogenous shock**	**T3 Low resource**
1. Initial resource level (in resource units)	$R_{max} = 2{,}000$ (= carrying capacity)	$R_{max} = 2{,}000$ (= carrying capacity)	$R_{shock} = 800$ (= post-shock level)
2. Time of pause (in seconds)	25	25	no pause (shorter round)
3. Resource dynamic at the time of the pause	no change	drop to $R_{shock} = 800$	no change (no pause)
4. Minimum duration of round (in seconds)	240	240	215
the following dynamics are the same in all three treatments:			
End of round (random continuation rule)	beyond the certain round length, every 10 seconds there is a chance of 10% that the round ends		
Carrying capacity (in resource units)	$R_{max} = 2{,}000$		
Maximum sustainable yield (in resource units)	$R_{MSY} = 1{,}000$		
Known threshold (in resource units)	$R_{min} = 500$		
no differences between treatments in the second CPR round (CPR2):	CPR2 is based on the design of CPR1 in T1 Control in all three treatments		

Control treatment T1

The control treatment T1 is our baseline. As in Lindahl et al. (2016) and Schill et al. (2015), the starting level of the resource in the control treatment (T1) equals the carrying capacity of the resource R_{max}. The minimum duration of the round in T1 is 240 seconds. There is a planned pause of the game at $t_{shock} = 25$ seconds, which is unforeseen by subjects. This corresponds to the timing of the exogenous shock in the shock treatment (T2) in CPR1. We implement the pause in T1 to keep comparability between the two treatments. During the pause in T1, subjects receive the information in a pop-up window on screen indicating that the game is paused. Subjects are informed that the CPR game continues at the pre-pause level of the resource once all group members made their choice of an extraction level.

Exogenous shock treatment T2

We evaluate how the experience of a sharp decline in resource availability due to an exogenous shock affects extraction strategies in a setting with an exogenous shock (T2) compared to the control treatment (T1). The only difference in the set-up between treatment T2 and T1 is the introduction of an exogenous shock at the time of the pause $t_{shock} = 25$ in CPR1 (Cerutti and Schlüter, 2019; Kimbrough and Wilson, 2013). Inspired by the observation that stochastic events, like extreme weather events often diminish parts of resources (Scheffer et al., 2001), the resource drops regardless of its current level R to the lower, post-shock resource level $R_{shock} = 800$ resource units. While the game is paused, subjects receive the information that this exogenous reduction of the resource is a one-time event that is not going to happen again and that none of the other resource dynamics have changed. Subjects are informed that the CPR game will continue at the low resource level of $R_{shock} = 800$ resource units once all group members have chosen a new extraction level.

The resource level R_{shock} is chosen such that it satisfies the following two conditions. First, the resource drops below the socially optimal resource level, which corresponds to the MSY in our case ($R_{shock} < R_{MSY} = 1000$). Second, R_{shock} is chosen to be sufficiently higher than the threshold in order to make sure that even in groups where all group members were extracting the resource at the maximum rate, subjects still have sufficient time after the occurrence of the shock to reconsider their extraction choices and prevent a collapse of the resource by lowering their extraction level ($R_{shock} > R_{min} = 500$).[9] We implement the exogenous shock with

---

[9] Our modelling of the resource development under maximum extraction show that it takes approximately 12 seconds to drive the resource from the resource level after the shock $R_{shock} = 800$ to the threshold $R_{min} = 500$. Please see Appendix 2B for details of the parameterisation of the user-resource model.

$t_{shock} = 25$ seconds relatively early in the game to minimise the between-group variation of resource levels prior to the shock.

Low initial resource treatment T3

The set-up in T3 is equivalent to T1, except for two changes. First, the starting level of the resource in T3 is lower than in T1 and corresponds to the post-shock level of the resource in T2. Hence, the resource development in T3 starts at $R_{shock} = 800$. Second, the minimum duration in T3 corresponds to the time given after the pause in treatments T1 and T2. Hence, subjects in T3 have only 215 seconds as minimum duration. There is neither a pause nor an exogenous shock implemented in T3. Thus, with T3 we aim to distinguish the effect of the mere reduction in resource stock from the impacts of a shock (the *resource size effect*).

Second round in all three treatments

The set-up of the second CPR round (CPR2) is based on the idea that exogenous shocks and resulting resource scarcity potentially motivate migration to more abundant resources elsewhere. Such migration would lead to new group formations at the side of the abundant resource, where many of the migrated resource users would have had similar previous experiences. Therefore, in each session we rematched all subjects implementing a stranger matching rule before the start of CPR2. The design of CPR2 is equal to that of the control treatment in CPR1 (Table 2.1). At the start of CPR2, all resources are set at the level of the carrying capacity regardless of the resource level at the end of CPR1. Our aim in implementing CPR2 is to test for a *spillover effect* of the prior shock experience in CPR1 to subsequent cooperation and coordination in CPR2. With this, we compare the outcomes in CPR2 between the groups that experienced a shock in CPR1 (T2) and those that did not (T1).

Optimal extraction strategies

Due to the implemented random continuation rule resulting in an infinite time horizon, it is socially optimal for groups to drive the resource to the level of MSY $R_{MSY} = 1{,}000$ as quickly as possible. Thereafter, it is socially optimal to keep the resource at R_{MSY} since the regrowth of 20 resource units per second is highest and a group extraction of $E_{MSY} = 20$ maintains the resource infinitely, maximising the group's extraction outcome over time. Thus, under the assumption of equal sharing, an individual extraction level of $E_i = 5$ is socially optimal.

Due to the differences between the three treatments, the socially optimal extraction paths that drive the resource to the level of MSY differ between treatments (Fig. 2.2). In T1 it is socially optimal for the group to extract the resource at full capacity, $\sum_{i=1}^{4} E_i = 40$, until it reaches R_{MSY} after about 40 seconds. It is also socially optimal in T2 to extract the resource at full capacity until the exogenous shock at 25 seconds. After the shock, groups in T2 should stop all extraction for about 11 seconds to allow the regrowth of the resource from R_{shock} up to R_{MSY}. The socially optimal group extraction path of T3 is equal to the one of T2 after the shock.

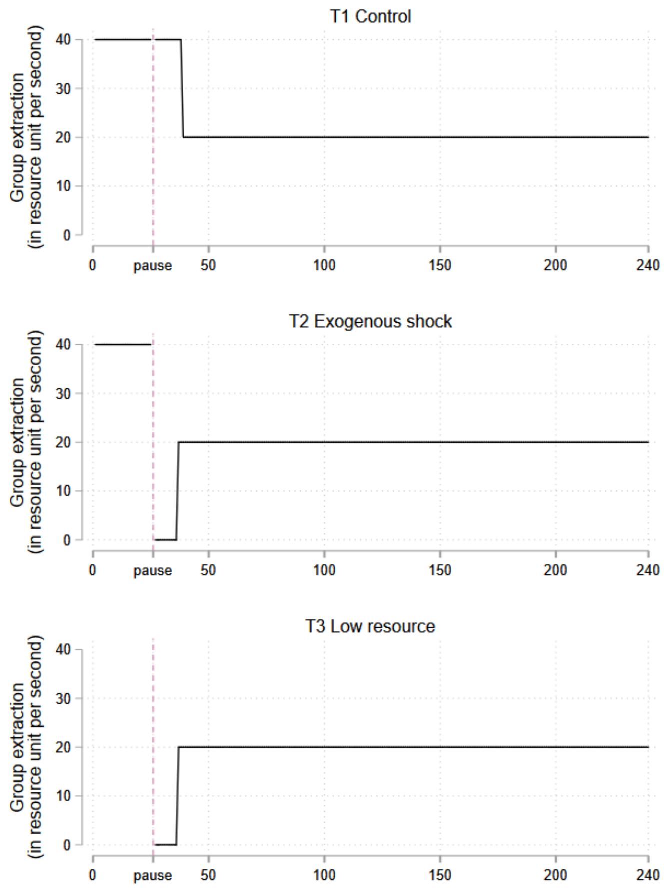


Fig. 2.2 Graphical presentation of the socially optimal extraction paths over time (in seconds). Due to the structural differences between the three treatments in CPR1, the optimal extraction paths differ in the beginning.

In all three treatments, selfish and myopic individuals have an incentive to free ride and extract continuously at the maximum of $E_i = 10$ to maximise their *individual total extraction*. However, due to the critical threshold and under the assumption of an infinite time horizon, it is in the best interest of every group member to prevent the collapse of the resource and coordinate on a resource level above the threshold to secure future extraction.

2.2.3 Experimental procedure

We recruited 320 student subjects using ORSEE (Greiner, 2015) to conduct 17 sessions at the LaER laboratory of Osnabrueck University in May and June 2019. In total, 96 subjects in 24 groups participated in T1, 136 subjects in 34 groups in T2 and 88 subjects in 22 groups in T3.

The average age of subjects is 23 years in T1 (SD = 3), 24 years in T2 (SD = 4) and 25 years in T3 (SD = 7). Unusually old students that participated in T2 (two subjects at age 44 and 46) and T3 (three subjects at age 44, 53 and 65) lead to a significant difference in the age distribution between treatments (Mann–Whitney–Wilcoxon tests: T1 *vs.* T2, p = 0.015 and T1 *vs.* T3, p = 0.016). We ran robustness checks to control for the imperfect randomisation with respect to subjects' age, which are in line with our main results (see Appendix 2E).

Apart from this, we find no significant differences between the three treatments with respect to subjects' gender, student status, field of study, monthly income and previous experience with experiments. There is no evidence for statistically significant differences in subjects' understanding of the experimental instructions, subjects' elicited social preference, their stated risk preference, or their ecological worldviews.[10]

[10] We elicited subjects' social preference by using the social value orientation method as presented in Murphy et al. (2011). Subjects' risk preference was measured by implementing the risk question as presented in Dohmen et al. (2011) and their ecological worldviews were measured by using the New Environmental Paradigm (NEP) scale based on Dunlap et al. (2000), Hawcroft and Milfont (2010) and Schleyer-Lindenmann et al. (2018). Please see Appendix 2D for details.

Upon arrival, the subjects were randomly seated in our laboratory and the login codes for the experimental software SoPHIE (Hendriks 2012) were randomly distributed by the experimenter. Following standard procedures of laboratory experiments, no communication between subjects and no use of technical devices like mobile phones was allowed during the entire experiment. Throughout the experiment, subjects remained anonymous and they received their payout individually in cash at the end. The average payout including a three Euro attendance fee was 13.50 Euro and each session lasted about 90 minutes. Please see Appendix 2A for details of the procedures.

2.3 Formulating hypotheses

Our study focuses on the analysis of a shock experience in the presence of a critical threshold. We analyse two main group outcomes: (1) the collapse of the resource as measurement for failed *coordination* and (2) the group's total extraction in relation to the socially optimal extraction as measure of *cooperation* among group members.

Firstly, we focus on the failed coordination, measured by groups' average probability of causing a collapse of the resource in the different treatments. In particular, we compare the average probability to cause a collapse of the resource for groups with exogenous shock (T2) to groups in the control treatment without shock (T1).

Blanco et al. (2015) find that extraction significantly increases after an exogenous reduction of the resource. Hence, we expect that individuals will increase their extraction as a response to the exogenous shock and that they fail to cooperate by not stopping all extraction of the resource to allow for sufficient regrowth after the shock experience. Due to the proximity of the resource to the critical threshold, subjects might perceive the collapse as inevitable. Since they anticipate the increase in their group's members' extraction and the future loss of the resource, they increase their current extraction (Reed and Heras, 1992). Consequently, we expect that groups' average probability of causing a collapse of the resource is greater in T2 than in T1.

Secondly, we analyse the level of cooperation amongst group members. We define a cooperative group as one that follows the socially optimal extraction path. We analyse the level of cooperation by using the *normalised group extraction (NGE)* as a proxy for cooperation (Cerutti and Schlüter, 2019). The NGE is calculated as follows:

$$NGE = \frac{total\ group\ extraction}{socially\ optimal\ group\ extraction}$$

where *total group extraction* is the total sum of extracted resource that groups extracted until the end of the round's minimum duration. The *socially optimal group extraction* is the maximum sum of extracted resource units that a group can extract if following the socially optimal extraction path of the given treatment. Please see Appendix 2C for details of the calculations.

We are interested in the comparison of the average NGE of treatment groups with exogenous shock experience (T2) as compared to groups without it (T1). Higher values of the NGE represent higher cooperation and more socially optimal group extraction. Previous studies outline that an exogenous reduction in the available resource level causes an increase in subjects' extraction choices (Blanco et al., 2015; Cinner et al., 2011) and that the experience of exogenous reductions lowers the relevance of subjects' decisions and their incentive to cooperate (Cerutti and Schlüter, 2019). Therefore, we expect that average NGE in treatment groups (T2) is lower than its counterpart in control groups (T1).

Thus, we formulate the first hypothesis on the *shock effect*:

Hypothesis 1.1: With the prior knowledge of a threshold, experiencing an exogenous shock results in both lower *coordination* causing a higher probability of the resource collapsing and lower *cooperation* amongst the resource users as a group.

Our design allows us to distinguish between two effects as drivers of a decrease in both coordination and cooperation in response to the experience of an exogenous shock. First, the experience of resource scarcity could be driving the effect of exogenous shock on coordination and cooperation. Exogenous shocks cause resource scarcity, which limits the size of the resource available to subjects. The mere difference in resource levels that subjects are exposed to could drive the changed outcome after a shock (*resource size effect)*. Earlier literature on the link between group cooperation and resource size, especially resource scarcity is inconclusive (Nie et al., 2020 and citations therein). On the one hand, Osés-Eraso and Viladrich-Grau (2007) and Osés-Eraso et al. (2008) find that resource scarcity induces caution amongst resource users and therefore lowers subjects' extraction efforts. The results of these papers are in line with Gibson (2001) who stated that the experience of resource scarcity is necessary to motivate collective action amongst resource users, and Nie et al. (2020) who find higher levels of cooperation amongst resource users if they are facing water scarcity in an irrigation system. Contradicting these results, others have found that the experience of scarcity decreases subjects' level of cooperation and increases their individual extraction efforts (Cinner et al., 2011; Blanco et al., 2015; Gatiso et al., 2015; Cerutti and Schlüter, 2019). A potential explanation is that resource scarcity is a potential driver of competition amongst resource users that leads to a faster rate of resource exhaustion (Grossman and Mendoza, 2003; Prediger et al., 2014).

Second, the experience of exogenous shocks as independent events themselves could drive the effect on both coordination and cooperation. Exogenous shocks possibly cause the disruption of a group's coordination efforts, making it harder for subjects to prevent a collapse of the resource. Additionally, we expect that cooperation deteriorates in groups that experience a disruption due to an exogenous shock and subjects are less likely to agree on the socially optimal extraction path. We refer to these additional impacts of a shock beyond the mere exposure to resource scarcity as the *disruption effect*. We expect that there is a disruption effect that leads groups that have experienced a low resource level due to an exogenous shock (T2) to coordinate and cooperate less in CPR1 than groups who experience no shock, but a low resource level to start with (T3). Thus, we formulate:

Hypothesis 1.2: The shock induces a decrease in both *coordination* and *cooperation* due to a *disruption effect*. This *disruption effect* is independent of what can be explained by changing resource availability (the *resource size effect*).

Next, we assess whether a shock experience (and consequential resource scarcity) causes a *spillover effect* on future resource management in a scenario where all individuals are rematched in groups with strangers and the initial resource rebounds to an abundant level (CPR2). Previous studies' evidence of spillover effects of resource scarcity on future cooperation and collective action is inconclusive. Some studies find evidence for a *negative spillover effect*, where scarcity in the past motivates spiteful behaviour and erodes cooperation (Pfaff et al., 2015; Prediger et al., 2014). Furthermore, the experience of an exogenous shock in a linear resource appropriation game without the implementation of a critical threshold leads to higher extraction when facing an abundant resource at a later stage in comparison to individuals that did not experience exogenous change to the resource (Blanco et al., 2015). In contrast, Nie et al. (2020) find that the experience of resource scarcity enhances cooperation (*positive spillover effect*). Since we find more evidence for a negative spillover effect in the literature, we formulate as third hypothesis:

Hypothesis 1.3: There is a negative *spillover effect* of shock experience on both future *coordination* and *cooperation*.

We test for the spillover effect by comparing the level of coordination and cooperation in CPR2 between T1 and T2. To our knowledge, we are the first to analyse the spillover effect with the endogenous threat to cross a critical threshold.

2.4 Results

Our data analysis focuses on the *coordination* and *cooperation* of the groups, based on the resource level and total extraction at 240 seconds (215 in CPR1 of T3), the minimum duration of CPR1 known to all subjects with certainty. Our study differs from previous studies because all treatments of our experiment include the critical threshold. The statistical analysis focuses on nonparametric tests and is conducted in STATA 15 (StataCorp. 2017).

2.4.1 Shock effect on coordination and cooperation

At the start of CPR1, we do not find significant differences between T1 and T2. Subjects in both treatments chose high levels of extraction, which result in a sharp decline of the resource between the start and the pause at 25 seconds (Fig. 2.3a).

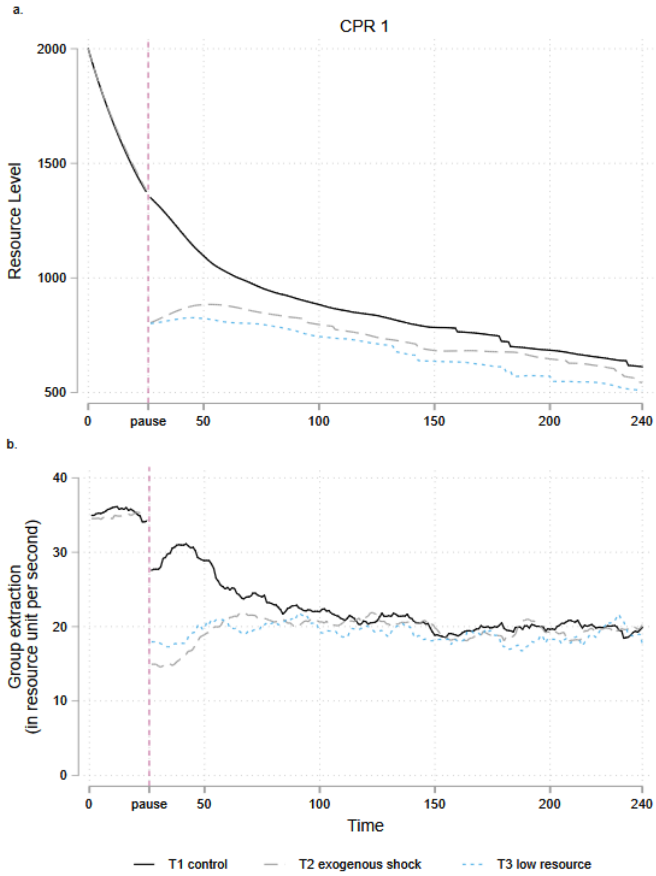


Fig. 2.3. Graphical presentation of (a) the development of the average resource level per treatment over time in CPR1 and (b) the corresponding average extraction rate of groups per treatment. At the start of CPR1, the resource development starts at the maximum carrying capacity of 2,000 resource units, except for T3 in CPR1 where it starts at the post-shock level of 800 resource units. Groups' extraction ranges between 0 and 40 resource units per second. The vertical dashed line at "pause" marks 25 seconds, which is the time of pause in T1 and T2 and corresponds to the start of T3 in CPR1. The development of the resource and the extraction choices is shown for the minimum duration of 240 seconds (215 seconds in T3).

We find no significant differences in resource levels at the time of the pause, neither between groups within treatments T1 (M=1,378, SD=64, Kruskal-Wallis test (KW): Chi square=23, p=0.461, df=23) and T2 (M=1,385, SD=67, KW: Chi square=32.95, p=0.47, df=33) nor between T1 and T2 (pairwise Mann–Whitney–Wilcoxon test (MWW): p=0.693). Therefore, we find no evidence that groups' behaviour in T1 and T2 differed significantly prior to the pause. Any observed differences in both coordination and cooperation should thus be driven by the difference in experience in the pause and afterwards.

After the pause in CPR1, we observe a significant difference in the average decrease of groups' extraction levels (in resource units per second (RUS)) between T1 (-7 RUS, SD=6) and T2 (-20 RUS, SD=5; MWW: p = 0.000)[11], which shows a strong treatment effect with respect to the immediate response in groups' extraction to the shock experience in T2. Groups' average extraction level in T2 (15 RUS, SD=4) is significantly lower than in T1 (28 RUS, SD=5, MWW: p = 0.000) (Fig. 2.3b).

As intended by the experimental design, the majority of subjects in T2 stated that they interpreted the exogenous shock as a natural catastrophe, like a bushfire or a drought (46%), whereas the majority of subjects in T1 interpreted the pause as a nudge to reflect on their own extraction choices (stated by 30% in T1 but only 15% in T2) which might explain the observed decrease in extraction levels in T1.[12] Overall, the majority of groups in all treatments did not follow the socially optimal extraction path and overexploited the resource, obtaining an average resource level below the MSY of 1,000 resource units at the end (Fig. 2.3a, Table 2.2).

Fig. 2.4 presents the percentage of groups per treatment that cause a collapse of the resource in CPR1: 17% in T1 and 21% in T2. These relatively low percentages of collapses are in line with the finding in the literature that certain threshold knowledge has a strong coordinating effect (Barrett and Dannenberg, 2014; Brown and Kroll, 2017; Dannenberg et al., 2015; Lindahl et al., 2016; Milinski et al., 2008; Schill et al., 2015). The endogenous threat to drive the resource below its critical threshold and cause a collapse changes the collective action problem fundamentally as compared to a situation without threshold (Barrett, 2013).

---

[11] The average decrease is computed by subtracting groups' average extraction level at 25 seconds (pre-pause) from groups' average extraction level at 27 seconds (post-pause) per treatment.

[12] Other common answers to the feedback question on the pause were "no interpretation" (20% in T1 and 16% in T2) and "software glitch" (13% in T1 and 5% in T2). 16% in T1 also reported an interpretation in the context of a natural catastrophe.

Table 2.2. Overview of the treatment effects in CPR1.

	T1 Control			T2 Exogenous shock			T3 Low resource				
	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Fisher's exact tests[a]	Mann-Whitney-Wilcoxon tests[a]
Fraction of groups that caused a collapse of the resource	0.1667 (0.3807)			0.2059 (0.4104)			0.2273 (0.4289)			1.000 1.000 0.718	
Average resource level at the end[b]	612 (306)	0	969	543 (306)	0	1,012	509 (308)	0	994		0.235 0.463 0.107
Normalised group extraction (NGE)[c]	1.011 (0.096)	0.74	1.08	0.976 (0.164)	0.39	1.07	0.927 (0.157)	0.52	1.04		0.069* 0.007** 0.0001***

Note: 96 subjects in 24 groups participated in T1, 136 subjects in 34 groups in T2 and 88 subjects in 22 groups in T3. The means include groups that caused a collapse of the resource prior to the end of the minimum duration. Standard deviations (SD) are presented in parentheses. Min refers to the lowest observed value and max to the highest observed value.
[a] The **p-values** of the pairwise two-sided Fisher's exact and Mann-Whitney-Wilcoxon tests tests are in the following order: T1 vs. T2, T2 vs. T3 and T3 vs. T1.
[b] **At the end** refers to the end of the minimum duration: 240 seconds in T1 and T1, 215 seconds in T3.
[c] To achieve comparability between treatments, the NGE is calculated using groups' total extraction after the pause in T1 and T2.
* p < 0.1; ** p < 0.05; *** p < 0.01

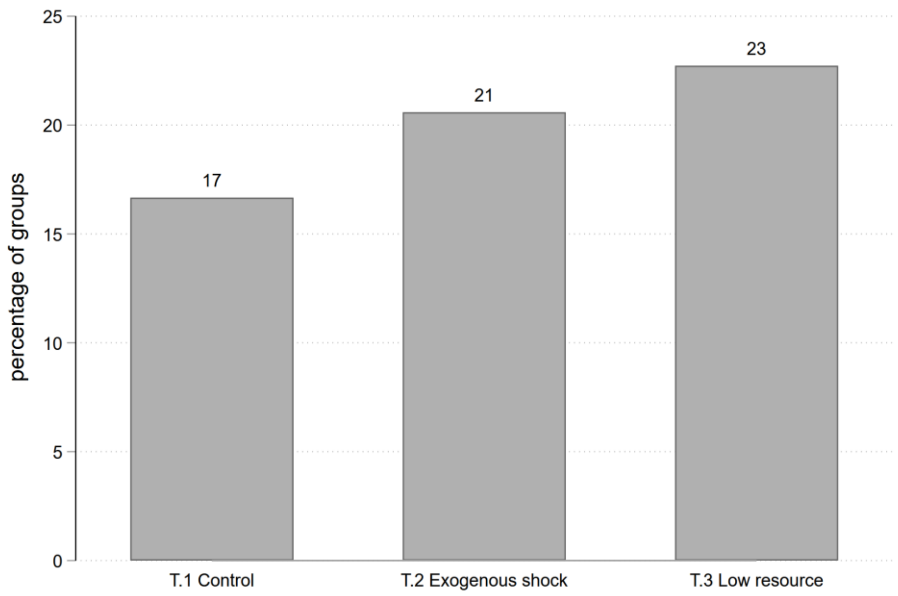


Fig. 2.4 Graphical presentation of the percentage of groups that caused a collapse per treatment in CPR1: 17% (SD=38) in T1, 21% (SD=41) in T2 and 23% (SD=43) in T3.

We then ran pairwise two-sided Fisher's exact (FET) and Mann-Whitney-Wilcoxon tests (MWW) to systematically test for a shock effect on both coordination and cooperation in CPR1 (Table 2.2). As to coordination, the probability to observe a collapse is 24% higher in T2 than in T1. Yet, this difference is not statistically significant (FET: p=1.00). As to cooperation, the analysis of the *normalised group extraction (NGE)* shows that groups in T1 have a significantly higher NGE and thus, a higher level of cooperation than groups in T2 (MWW: p=0.07, Table 2.2). Thus, we find evidence that the shock experience in CPR1 significantly decreases cooperation in T2 in comparison to the control groups (T1).

In summary, with regard to our first hypothesis:

Result 1.1: Experiencing an exogenous shock leads to lower *cooperation* amongst group members, while the negative impact on *coordination* is statistically insignificant.

It should be noted that even though the value of the NGE theoretically ranges between 0 and 1, we observed that some groups reach values higher than 1 (Table 2.2). We argue that some subjects potentially do not believe that the round will continue long after the minimum duration that is known by subjects with certainty. Therefore, these subjects have an incentive to drive the resource down to the critical threshold towards the perceived end of the minimum duration at 240 seconds. Thus, groups overexploited the resource and gained a higher sum of extraction than is socially optimal under the assumption of an infinite time horizon. However, we do not find evidence that such a potential endgame effect is driving our result of a negative shock effect on cooperation (please see robustness checks in Appendix 2C).

2.4.2 Analysis of disruption and resource size effect

Fig. 2.3b shows that the groups' level of extraction is briefly significantly lower in T2 after the pause (14.94 RUS, SD=4.42) than at the start of T3 (17.95 RUS, SD=4.98, MWW: p = 0.015). This indicates that groups' extraction behaviour differs whether resource scarcity occurs after an exogenous shock (T2) or as the initially given state (T3).

Regarding coordination, we find that the probability of collapsing is higher in T3 (23%) than in either T1 (17%) or T2 (21%). However, these differences are not statistically significant (FET T3 vs. T1: p=0.718; T3 vs. T2: p=1.00; Table 2.2), so we do not find evidence for a resource size effect on coordination. The experience of initial resource scarcity (T3) in comparison to initial resource abundance (T1) does not lead to a statistically significant increase in collapsing resources. Furthermore, we do not find evidence that the experience of a disruption (T2) increases the probability of the collapse of the resource in CPR1 in addition to the resource size effect (T3). We conclude that neither the exogenous shock, nor the initial resource scarcity have an effect on coordination.

When analysing cooperation, we find evidence for a negative resource size effect. The average NGE is 9% lower in T3 than in T1, which is statistically significant (MWW: p=0.0001, Table 2.2). Experiencing initial resource scarcity (T3) leads to higher rates of resource extraction and thus, a lower level of cooperation in comparison to groups that only experience resource abundance without any exogenous shock (T1).

With respect to the disruption effect, we find that the average NGE is 5% higher in T2 than in T3, which is statistically significant (MWW: p=0.007, Table 2.2), so we find evidence for a disruption effect. However, the direction of this disruption effect is contrary to what we expected. Subjects' prior experience of resource abundance with the shock thereafter in T2 appears to motivate an additional *increase* in cooperation in comparison to T3 where the initial resource availability was already low. Our results add to the body of evidence indicating that (initial) resource scarcity has a negative immediate impact on cooperation and sustainable resource management (Gatiso et al., 2015; Grossman and Mendoza, 2003). However, this negative impact is partly offset in the case where the resource scarcity stems from an exogenous shock. One possible explanation may be that experiencing the shock makes subjects more cautious.

Thus, regarding our second hypothesis, we find:

Result 1.2: Experiencing the shock induces an increase in *cooperation* (a positive *disruption effect*), but no significant change in *coordination*, independent from the *resource size effect*.

2.4.3 Spillover effect of shock experience

We next examined the *spillover effect* of shock experience in CPR1 on behaviour in CPR2. To do this, we compared the levels of both coordination and cooperation in groups of subjects that all had prior shock experience (T2) to the control groups without shock experience (T1).

At the beginning of CPR2, we observed slightly more cautious resource extraction levels in T2 than in T1 and after the pause the average extraction levels of groups in T2 (-8 RUS, SD 5) is significantly lower than it is in T1 (-3 RUS, SD 5, MWW: p=0.001, Fig. 2.5b).

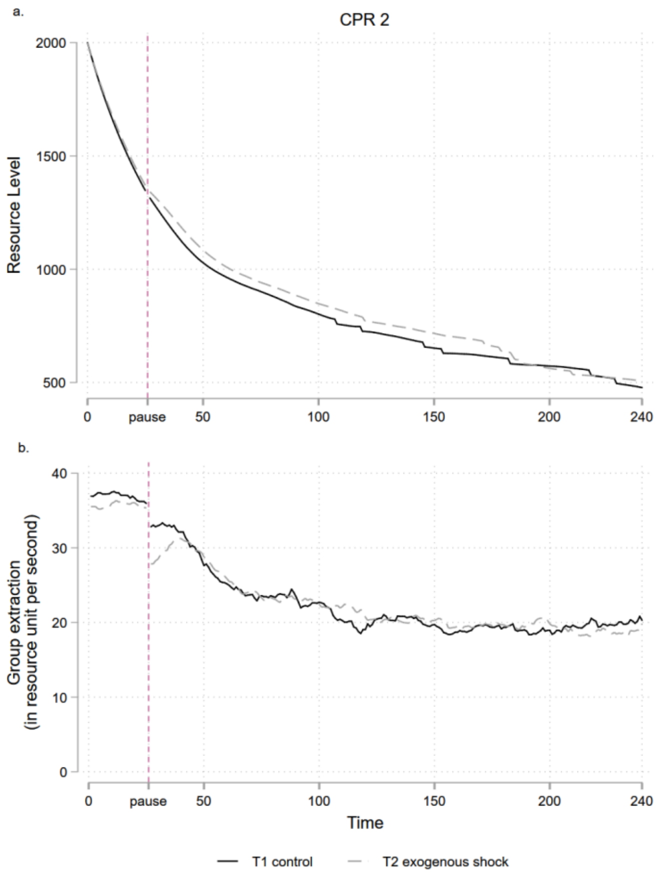


Fig. 2.5. Graphical presentation of (a) the development of the average resource level in T1 and T2 over time in CPR2 and (b) the corresponding average extraction rate of groups per treatment. At the start of CPR2, the resource development starts at the maximum carrying capacity of 2,000 resource units. Groups' extraction ranges between 0 and 40 resource units per second. The vertical dashed line at "pause" marks 25 seconds, which is the time of pause. The development of the resource and the extraction choices is shown for the minimum duration of 240 seconds.

Furthermore, we observe that the shock experience in T2 stabilises both coordination (percentage of groups that cause a collapse of the resource is 21% in CPR1 and CPR2) and cooperation (NGE of 0.98 in CPR1 *vs*. 0.99 in CPR2) in subsequent rounds with resource abundance. Thus, we find some weak indications that support the finding that resource scarcity due to an exogenous shock induces higher levels of caution amongst resource users when they face resource abundance in subsequent rounds (Gibson, 2001; Nie et al., 2020; Osés-Eraso et al., 2008; Osés-Eraso and Viladrich-Grau, 2007).

In CPR2, groups in T2 have on average, a lower probability to cause a collapse of the resource than groups in T1 (21% *vs*. 29%) and a higher NGE measure (0.99 *vs*. 0.95). Hence, in CPR2, groups in T2 are more likely to coordinate and cooperate than groups in T1. However, these differences between treatments T1 and T2 are not statistically significant (Table 2.3). In contrast to previous studies (Blanco et al., 2015; Pfaff et al., 2015), we do not find evidence of a negative spillover effect of shock experience where subjects that have experienced an exogenous shock and resource scarcity extract more once they face resource abundance again (T2) compared to subjects that experienced repeated resource abundance (T1).

Table 2.3. Overview of the treatment effects of T1 and T2 in CPR2.

	T1 Control			**T2 Exogenous shock**				
	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Fisher's exact test (p-value)	Mann-Whitney-Wilcoxon test (p-value)
Fraction of groups that caused a collapse of the resource	0.2917 (0.4643)			0.2059 (0.4104)			0.539	
Average resource level at the end[a]	478 (330)	0	939	506 (276)	0	855		0.956
Normalised group extraction (NGE)[b]	0.951 (0.168)	0.52	1.08	0.990 (0.112)	0.59	1.08		0.613

Note: 96 subjects in 24 groups participated in T1, 136 subjects in 34 groups in T2 and 88 subjects in 22 groups in T3. The means include groups that caused a collapse of the resource prior to the end of the minimum duration. Standard deviations (SD) are presented in parentheses. Min refers to the lowest observed value and max to the highest observed value.
[a] **At the end** refers to the end of the minimum duration: 240 seconds in T1 and T1, 215 seconds in T3.
[b] To achieve comparability between treatments, the NGE is calculated using groups' total extraction after the pause in T1 and T2.
* p < 0.1; ** p < 0.05; *** p < 0.01

With regard to our third hypothesis, we thus summarise:

Result 1.3: There is no statistically significant evidence of a *spillover effect* of shock experience on both *coordination* and *cooperation.*

2.5 Discussion and conclusion

Climate change increases the frequency of extreme weather events (IPCC, 2014) and thereby, increases the risk of exogenous shocks to resource systems. At the same time, cooperation and coordination amongst resource users is needed to manage natural resources sustainably while preventing overexploitation that potentially drives resource systems beyond critical thresholds causing a collapse of the resources. We designed a novel (quasi-) continuous-time common-pool resource experiment for the laboratory to analyse the potential impact that the interaction of exogenous ecological and endogenous extraction dynamics has on *coordination* amongst resource users (avoiding collapse) and their *cooperation* (group's total resource extraction in relation to the socially optimal extraction). This interaction of exogenous and endogenous dynamics in resource systems had so far been neglected (Blanco et al., 2015; Cerutti and Schlüter, 2019; Lindahl et al., 2016).

Even if subjects experience an exogenous shock, our results confirm the coordinating effect from threshold knowledge (e.g. Barrett, 2013; Dannenberg et al., 2015; Lindahl et al., 2016). While groups that experience an exogenous shock (T2) appear to be somewhat more likely to cause a collapse of the resource due to overexploitation, the difference to the control treatment (T1) is not statistically significant. Nonetheless, the average level of cooperation is significantly lower in groups that experience an exogenous shock (T2) as compared to the control treatment without exogenous shock (T1). Thus, the known critical threshold seems to balance out the potential negative effect of shock experience on coordination, but not on cooperation.

Besides the negative effect of the exogenous shock on cooperation due to the exposure to resource scarcity (the *resource size effect*), we find evidence that the disruption caused by shock experience has an additional effect on cooperation (which we term *disruption effect*), but not on coordination. Contrary to our initial expectation, this disruption effect on cooperation is positive, i.e. it increases cooperation. Specifically, we observe a lower average level of cooperation in groups that experience initial resource scarcity (T3) than in groups that experience scarcity due to the exogenous shock (T2). Thus, the shock experience seems to make subjects actually more cautious when facing resource scarcity in comparison to subjects that experience scarcity as initial resource state. Thereby, the shock experience partly counteracts the negative effect of the scarcity experience on cooperation which had been found in previous studies (Gatiso et al., 2015; Grossman and Mendoza, 2003; Prediger et al., 2014).

Unlike Blanco et al. (2015) and Pfaff et al. (2015), we do not find evidence for a negative *spillover effect* of shock experience in the past on both future coordination and cooperation in our setting where resource users are aware of a critical threshold. Instead we observe weak indications that the shock experience prevents a deterioration of coordination and cooperation in subsequent rounds with resource abundance. Thus, our findings are partly in line with previous studies who outline that the experience of scarcity has a positive effect on cooperation (Gibson, 2001; Nie et al., 2020; Osés-Eraso et al., 2008; Osés-Eraso and Viladrich-Grau, 2007).

Our experiment builds upon previous studies that analyse human behaviour in socio-ecological systems by introducing more complex resource dynamics to CPR experiments in the laboratory (amongst others Cerutti and Schlüter, 2019; Janssen, 2010; Lindahl et al., 2016; Schill et al., 2015) and contributes to the understanding of human responses to environmental change. It has been shown that institutional arrangements not only depend on human interactions but also on the given ecological dynamics of the resource (Janssen et al., 2015; McAllister et al., 2011) and thus, advanced understanding of behavioural responses to ecological change can potentially optimise management strategies of natural resources. We see our analysis of coordination and cooperation amongst group members and thus, an advanced understanding of self-governance when exposed to exogenous shocks and critical thresholds, as a first step. Given the increase of damaging exogenous shocks to natural resources due to climate change (IPCC, 2014), further research is needed to analyse how the implementation of different institutions affects human behaviour and the management of natural resources in the described context of ecological dynamics.

While we aimed for a high level of ecological complexity in our experiment, we simplified the dynamics to secure subjects' understanding of the novel continuous-time design of the CPR game. The exogenous shock implemented in our experiment has a fixed size for all groups and only reduces the size of the resource, but not individuals' accumulated income. Yet, natural catastrophes as exogenous shocks often destroy the prospect of future income as well as individual's accumulated wealth. Moreover, subjects in our experiment did not have any costs connected to their resource extraction, whereas in the real world the cost of extraction often increases if the size of the resource decreases (Cinner et al., 2011). In a real-world scenario, the loss of accumulated wealth provides an additional incentive to earn income and therefore, potentially motivates increased extraction efforts after the experience of an exogenous shock, which could amplify the negative effect of shock experience on both coordination and cooperation. On the other hand, the loss of accumulated wealth due to an exogenous shock might restrict resource users' extraction activities since they cannot cover the costs of increased extraction. In the future, it would be interesting to assess in how far the impact of shock experience depends on the size of the shock and its consequences on individuals' accumulated income.

Furthermore, subjects in our experiment faced a constant rate of return to the level of extraction effort such that the amount of extracted resource units with a chosen extraction level did not depend on the size of the resource. In reality however, with a constant level of extraction effort the generated income often varies depending on the size of the resource, and resource users who experience an exogenous shock are likely to increase their extraction efforts to compensate for the decrease in generated income per level of extraction effort (Cinner et al., 2011). However, by increasing their level of extraction effort to keep the generated income steady after a shock, resource users' extraction is no longer sustainable due to the resource's damaged state after the shock. These increased extraction efforts prevent the recovery of the resource stock after the shock, leading to overexploitation of the resource and potentially amplifying negative trends in the resource system after the shock (ibid). Therefore, we assume that the negative impact of resource scarcity induced by exogenous shocks on natural resource management is more severe in the real world than it is in our experiment. In future work, it would be interesting to analyse if the experience of an exogenous shock that also diminishes accumulated wealth amplifies the effect of shock experience in comparison to the resource size effect in a scenario where individuals are aware of the resource's critical threshold. Further, it would be interesting to assess in how far a flexible rate of return that depends on the resource level would affect behavioural response to the shock experience.

Overall, our results emphasise the impact of ecological dynamics on human behavioural responses. We see this as an important contribution to the exchange between policy makers, ecologists and resource users who all have an interest in identifying sustainable management strategies. Precise knowledge of local resource dynamics and their critical thresholds potentially prevent a collapse of the resources even in the light of an increase in exogenous shocks due to climate change. While subjects that participated in our laboratory experiment received perfect feedback about their groups' total extraction and had perfect knowledge about the ecosystem dynamics with respect to the location of the critical threshold, resource users in the real world face higher levels of ecological and social uncertainty. Thus, it would be worthwhile to increase the complexity of ecological dynamics and social uncertainty in future experiments to further the understanding of the interaction of exogenous and endogenous dynamics in natural resource management.

References Chapter 2

Barrett, S., 2013. Climate treaties and approaching catastrophes. J. Environ. Econ. Manage. 66, 235–250. https://doi.org/10.1016/j.jeem.2012.12.004

Barrett, S., Dannenberg, A., 2014. Sensitivity of collective action to uncertainty about climate tipping points. Nat. Clim. Chang. 4, 36–39. https://doi.org/10.1038/nclimate2059

Biggs, R., Carpenter, S.R., Brock, W.A., 2009. Turning back from the brink: Detecting an impending regime shift in time to avert it. Proc. Natl. Acad. Sci. 106, 826–831. https://doi.org/10.1073/pnas.0811729106

Bigoni, M., Casari, M., Skrzypacz, A., Spagnolo, G., 2015. Time Horizon and Cooperation in Continuous Time. Econometrica 83, 587–616. https://doi.org/10.3982/ECTA11380

Blanco, E., Lopez, M.C., Villamayor-Tomas, S., 2015. Exogenous degradation in the commons: Field experimental evidence. Ecol. Econ. 120, 430–439. https://doi.org/10.1016/j.ecolecon.2015.03.028

Brandt, G., Kulesz, M.M., Nissen, D., Merico, A., 2017. OGUMI—A new mobile application to conduct common-pool resource experiments in continuous time. PLoS One 12. https://doi.org/10.1371/journal.pone.0178951

Brown, T., Kroll, S., 2017. Avoiding an uncertain catastrophe: climate change mitigation under risk and wealth heterogeneity. Clim. Change 141, 155–166. https://doi.org/10.1007/s10584-016-1889-5

Cardenas, J., Janssen, M.A., Bousquet, F., 2013. Dynamics of Rules and Resources: Three New Field Experiments on Water, Forests and Fisheries, in: Handbook on Experimental Economics and the Environment. pp. 1–27. https://doi.org/10.4337/9781781009079.00020

Cerutti, N., Schlüter, A., 2019. Resource changes: exogenous or endogenous, gradual or abrupt. Experimental evidence. Int. J. Environ. Stud., 1–15. https://doi.org/10.1080/00207233.2019.1644022

Cerutti, N., Schlüter, A., 2017. Resource Changes: Exogenous or Endogenous , Gradual or Abrupt. Experimental Evidence.

Cinner, J.E., Folke, C., Daw, T., Hicks, C.C., 2011. Responding to change: Using scenarios to understand how socioeconomic factors may influence amplifying or dampening exploitation feedbacks among Tanzanian fishers. Glob. Environ. Chang. 21, 7–12. https://doi.org/10.1016/j.gloenvcha.2010.09.001

Dal Bó, P., Fréchette, G.R., 2018. On the Determinants of Cooperation in Infinitely Repeated Games : A Survey. J. Econ. Lit. 56, 60–114.

Dannenberg, A., Löschel, A., Paolacci, G., Reif, C., Tavoni, A., 2015. On the Provision of Public Goods with Probabilistic and Ambiguous Thresholds. Environ. Resour. Econ. 61, 365–383. https://doi.org/10.1007/s10640-014-9796-6

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. J. Eur. Econ. Assoc. 9, 522–550. https://doi.org/10.1111/j.1542-4774.2011.01015.x

Dunlap, R.E., Van Liere, K.D., Mertig, A.G., Jones, R.E., 2000. New Trends in Measuring Environmental Attitudes: Measuring Endorsement of the New Ecological Paradigm: A Revised NEP Scale. J. Soc. Issues 56, 425–442. https://doi.org/10.1111/0022-4537.00176

Gatiso, T.T., Vollan, B., Nuppenau, E., 2015. Resource scarcity and democratic elections in commons dilemmas : An experiment on forest use in Ethiopia. Ecol. Econ. 114, 199–207. https://doi.org/10.1016/j.ecolecon.2015.04.005

Gibson, C.C., 2001. Forest resources: institutions for local governance in Guatemala, in: Burger, J., Ostrom, E., Norgaard, R.B., Policansky, D., Goldstein, B.D. (Eds.), Protecting the Commons: A Framework For Resource Management In The Americas. Island Press, Washington, DC, USA, pp. 71–89.

Greiner, B., 2015. Subject pool recruitment procedures : organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1, 114–125. https://doi.org/10.1007/s40881-015-0004-4

Grossman, H.I., Mendoza, J., 2003. Scarcity and appropriative competition. Eur. J. Polit. Econ. 19, 747–758. https://doi.org/10.1016/S0176-2680(03)00033-8

Hawcroft, L.J., Milfont, T.L., 2010. The use (and abuse) of the new environmental paradigm scale over the last 30 years: A meta-analysis. J. Environ. Psychol. 30, 143–158. https://doi.org/10.1016/j.jenvp.2009.10.003

Hendriks, A., 2012. SoPHIE-Software platform for human interaction experiments, Working Paper. Osnabrueck: Working Paper, Osnabrueck University.

IPCC, 2019: Summary for Policymakers. In: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems [P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.- O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, (eds.)]. In press. https://doi.org/10.4337/9781784710644

IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.. https://doi.org/10.1017/CBO9781107415324

Janssen, M.A., 2010. Introducing Ecological Dynamics into Common-Pool Resource Experiments. Ecol. Soc. 15.

Janssen, M.A., Lindahl, T., Murphy, J.J., 2015. Advancing the understanding of behavior in social-ecological systems: Results from lab and field experiments. Ecol. Soc. 20. https://doi.org/10.5751/ES-08097-200434

Kimbrough, E.O., Wilson, B.J., 2013. Insiders, outsiders, and the adaptability of informal rules to ecological shocks. Ecol. Econ. 90, 29–40. https://doi.org/10.1016/j.ecolecon.2013.02.008

Lindahl, T., Crépin, A.-S., Schill, C., 2016. Potential Disasters can Turn the Tragedy into Success. Environ. Resour. Econ. 65, 657–676. https://doi.org/10.1007/s10640-016-0043-1

McAllister, R.R.J., Tisdell, J.G., Reeson, A.F., Gordon, I.J., 2011. Economic Behavior in the Face of Resource Variability and Uncertainty. Ecol. Soc. 16. https://doi.org/10.5751/ES-04075-160306

Milinski, M., Sommerfeld, R.D., Krambeck, H.-J., Reed, F.A., Marotzke, J., 2008. The Collective-Risk Social Dilemma and the Prevention of Simulated Dangerous Climate Change. Proc. Natl. Acad. Sci. U. S. A. 105, 2291–2294.

Murphy, R.O., Ackermann, K.A., Handgraaf, M., 2011. Measuring Social Value Orientation. Judgm. Decis. Mak. 6, 771–781. https://doi.org/10.2139/ssrn.1804189

Nie, Z., Yang, X., Tu, Q., 2020. Resource scarcity and cooperation: Evidence from a gravity irrigation system in China. World Dev. 135. https://doi.org/10.1016/j.worlddev.2020.105035

Nyborg, K., Anderies, J.M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger, W.N., Arrow, K., Barrett, S., Carpenter, S.R., Chapin, F.S., Crépin, A.-S., Daily, G.C., Ehrlich, P.R., 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., de Zeeuw, A., 2016. Social norms as solutions. Science (80-.). 354, 42–43. https://doi.org/10.1126/science.aaf8317

Osés-Eraso, N., Udina, F., Viladrich-Grau, M., 2008. Environmental versus Human-Induced Scarcity in the Commons: Do They Trigger the Same Response? Environ. Resour. Econ. 40, 529–550. https://doi.org/10.1007/s10640-007-9168-6

Osés-Eraso, N., Viladrich-Grau, M., 2007. Appropriation and concern for resource scarcity in the commons: An experimental study. Ecol. Econ. 63, 435–445. https://doi.org/10.1016/j.ecolecon.2006.11.016

Ostrom, E., 2009. A General Framework for Analyzing Sustainability of Social-Ecological Systems. Science. 325, 419–422. https://doi.org/10.1126/science.1172133

Perman, R., Ma, Y., Common, M., Maddison, D., McGilvray, J., 2011. Natural Resource and Environmental Economics, 4th ed. Pearson Education.

Pettit, J., Friedman, D., Kephart, C., Oprea, R., 2014. Software for continuous game experiments. Exp. Econ. 17, 631–648. https://doi.org/10.1007/s10683-013-9387-3

Pfaff, A., Vélez, M.A., Ramos, P.A., Molina, A., 2015. Framed field experiment on resource scarcity & extraction: Path-dependent generosity within sequential water appropriation. Ecol. Econ. https://doi.org/10.1016/j.ecolecon.2015.06.002

Polasky, S., Zeeuw, A. De, Wagener, F., 2011. Optimal management with potential regime shifts. J. Environ. Econ. Manage. 62, 229–240. https://doi.org/10.1016/j.jeem.2010.09.004

Prediger, S., Vollan, B., Herrmann, B., 2014. Resource scarcity and antisocial behavior. J. Public Econ. 119, 1–9. https://doi.org/10.1016/j.jpubeco.2014.07.007

Reed, W., Heras, H., 1992. The conversation and exploitation of vulnerable resources. Bull. Math.Biol. 54, 185-207. https://doi.org/10.1007/BF02464829

Rutte, C.G., Wilke, H.A.M., Messick, D.M., 1987. Scarcity or Abundance Caused by People or the Environment as Determinants of Behavior in the Resource Dilemma. J. Exp. Soc. Psychol. 23, 208–216.

Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B., 2001. Catastrophic shifts in ecosystems. Nature 413, 591–596. https://doi.org/10.1038/35098000

Schill, C., Lindahl, T., Crépin, A.-S., 2015. Collective action and the risk of ecosystem regime shifts: Insights from a laboratory experiment. Ecol. Soc. 20. https://doi.org/10.5751/ES-07318-200148

Schleyer-Lindenmann, A., Ittner, H., Dauvier, B., Piolat, M., 2018. Die NEP-Skala – hinter den (deutschen) Kulissen des Umweltbewusstseins. Diagnostica 64, 156–167. https://doi.org/10.1026/0012-1924/a000202

StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.

Appendices Chapter 2

Appendix 2A: Details of the experimental design and procedure

Fig. 2.A.1 depicts the overall set-up of the experiment, which consisted of two payout-relevant parts, a (quasi-) continuous-time common-pool resource (CPR) game and a measure to elicit subjects' Social Value Orientation (SVO) (Murphy et al., 2011). The focus of our study is the CPR game with its two individual test rounds and control questions followed by two CPR group rounds. In addition, we implemented a post-experimental questionnaire that consist of four parts: (1) non-incentivised risk preferences (Dohmen et al., 2011); (2) the New Environmental Paradigm (NEP) scale to control for subjects' ecological worldviews (Dunlap et al., 2000); (3) feedback questions on the experimental design, and (4) questions on subjects' socio-economic background.

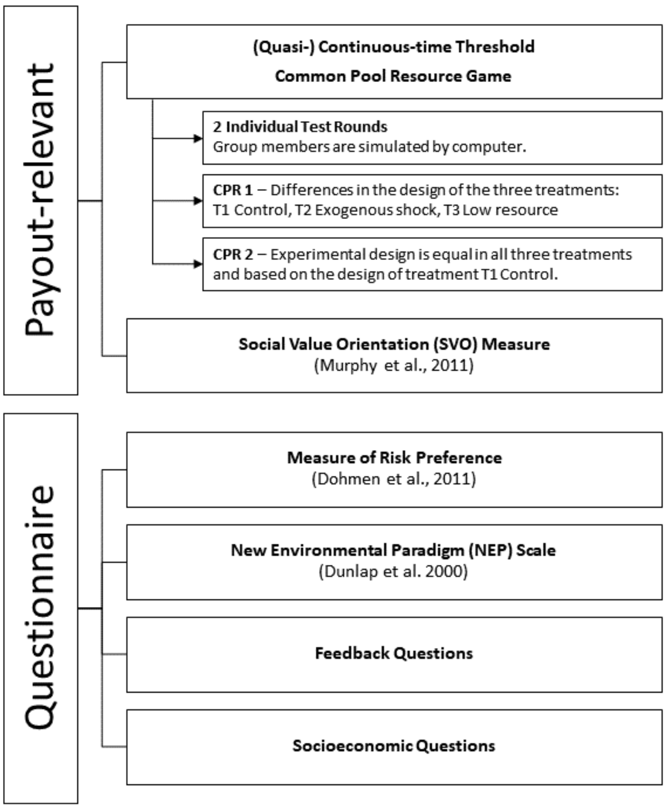


Fig. 2A.1 Graphical presentation of the two parts of the experiment. The (quasi-) continuous-time threshold common-pool resource game and the social value orientation were payout-relevant. Participants' answers to the questionnaire were not payout-relevant.

Before the start of each experimental session, subjects who registered for the session via the organisation software ORSEE (Greiner, 2015) freely chose their seats in the experimental laboratory. Once everyone was seated, the experimenter welcomed the subjects and clearly announced that no communication amongst subjects was allowed during the session. All electronic devices had to be switched off and stored away. To participate in the experimental session, subjects randomly picked a login code for SoPHIE (Hendriks, 2012). This way we secured subjects' random assignment to groups.

Subjects received a first part of the instructions for the CPR game as a printout before moving on to the rest of the instructions presented on screen. The experimenter read the printed instructions aloud, the on-screen instructions were only read by subjects in silence. We asked subjects to answer a series of control questions before playing the two individual test rounds and again before playing the two group rounds of the CPR game to secure their understanding of the game. In addition, the experimenter encouraged subjects to ask any questions throughout the experiment.

Table 2A.1. Overview of the resource's growth rates.

level of the resource (in resource units)	**growth of the resource** (in resource units per second)
2000	0
1900	3.8
1800	7.2
1700	10.2
1600	12.8
1500	15
1400	16.8
1300	18.2
1200	19.2
1100	19.8
1000	20
900	19.8
800	19.2
700	18.2
600	16.8
500	0
0	0

Details of the instructions

Subjects did not know the exact resource growth function but instead we presented them with an overview of different growth rates of the resource at different resource levels as presented in Table 2A.1. Please also see a translation of the original instructions in Appendix 2F.

Furthermore, subjects were presented with a graph that represented the growth of the resource for all the levels of the resource between the values presented in Table 2A.1 (Fig. 2A.2). Subjects knew that the development of the resource and also the growth of the resource dropped down to zero resource units per second as soon as the threshold of 500 resource units was reached. Also, we showed an exemplary depiction of a collapsing resource to subjects (Fig. 2A.3) to give them a better understanding of the dynamics once the group's resource would reach the threshold.

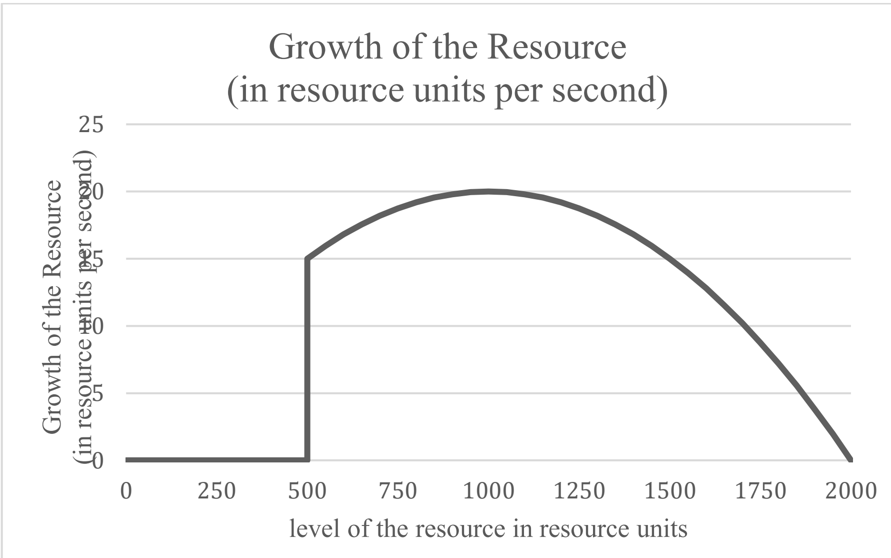


Fig. 2A.2. Graphical presentation of the growth of the resource as presented in the instructions to all three treatments.

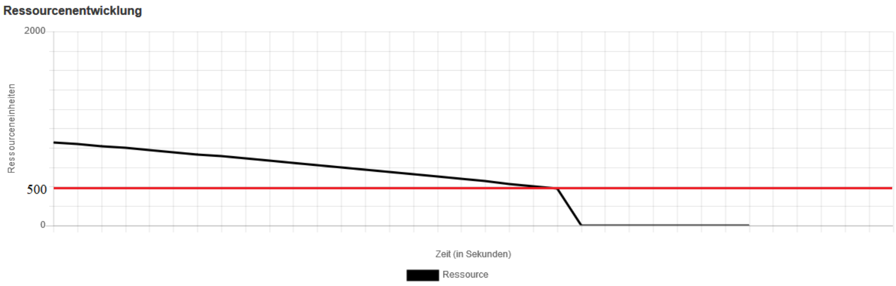


Fig. 2A.3. Demonstration of the resource development (Ressourcenentwicklung) once the resource reaches the critical threshold. To make it easier for subjects to understand the threshold mechanism we presented them with this graphical illustration in the instructions. The x-axis shows the time in seconds (Zeit in Sekunden) and the y-axis shows the resource units (Ressourceneinheiten)

Individual test rounds

Subjects could familiarise themselves individually with the resource dynamics and the resource's response to varying extraction levels in two non-payout relevant individual test rounds. To rule out any social interaction, the computer simulated the three group members during the individual test rounds such that a subject's individual choice was mirrored and multiplied by four as the group's choice. The computer simulation and shorter minimum round length were the only differences in the set-up between individual test rounds and group rounds. Both individual test rounds started at 2,000 resource units in T1 and T2 and 800 resource units in T3 and lasted for at least 80 seconds with certainty. The exact end of each test round was again determined by the random continuation rule that was also implemented in the CPR group rounds.

Use of the slider

Subjects could adjust their chosen extraction level any time by choosing any integer number between zero and ten with adjusting their slider's position. Once a subject clicked on the "Send"-button, the chosen extraction level was executed each second until subjects actively increased or decreased it. Any change to the slider needed to be confirmed by a click on the "Send"-button.

Elicitation of the social value orientation

Following the CPR game, we elicited subjects' social value orientation (SVO) as a control variable for subjects' concern towards others by using a computerised version of the SVO slider measure presented in Murphy et al. (2011). Subjects had to make the 15 distributive decisions of the SVO slider measure where they allocated points between themselves and another anonymous person. Depending on how many points subjects allocated to themselves and the other person in the six primary items of the SVO, we could calculate a reliable unidimensional index of subjects' general social preferences (SVO index) to differentiate between altruistic, prosocial, individualistic and competitive types (ibid).

Post-experimental questionnaire

The post-experimental questionnaire contained the risk question, where each subject rated itself between "0 – not all willing to take risk" and "10 – very willing to take risk" (Dohmen et al., 2011) and the New Environmental Paradigm (NEP) scale to compute subjects' ecological world view as a proxy for their environmental preferences (Dunlap et al., 2000; Hawcroft and Milfont, 2010).

The NEP scale consisted of 15 items that covered the five hypothesised facets of an ecological worldview. These five facets included the perceived reality of limits to growth, antianthropocentrism, the fragility of nature's balance, rejection of exemptionalism and the possibility of an ecocrisis. Subjects stated their agreement/disagreement to each of the 15 given statements on a 5-point-Likert scale between strongly agree, mildly agree, unsure, mildly disagree and strongly disagree (Dunlap et al., 2000).

Additionally, we elicited subjects' socio-economic background characteristics to test for structural differences between subjects. These characteristics included subjects' age, gender, student status, their highest degree, their monthly income and whether or not they have experience with participation in economic experiments.

Payout determination

Subjects' payout consisted of the payment for the CPR game and the SVO part. Firstly, subjects' individual total extraction from one of the two CPR group rounds was randomly chosen and converted into points such that one resource unit equalled one point. The points were converted with an exchange rate of 100 points = 0.40 Euro. Secondly, one of subjects' 15 distributive decisions in the SVO part was randomly chosen as payout-relevant. Each subject received the points that they allocated to themselves in their chosen decision, as well as the points that another subject in the experiment assigns to the other person in his/her randomly chosen decision. The exchange rate from points to Euro for the SVO part was 50 points = 1.50 Euro.

Subjects only received the instructions for the SVO part after finishing the CPR game. This was done to prevent them from averaging their payout across the two payout-relevant parts of the experiment. In addition, we did not provide direct feedback on subjects' payout in Euro from the first part of the experiment until the very end of the experiment. However, subjects could form an expectation on their payment for the CPR game since they knew the exchange rate from points to Euro, they were told that one of the two CPR rounds would be randomly chosen as payout-relevant and they could observe their total extraction during both CPR rounds. To reduce social learning effects and avoid strategic interactions throughout all parts of the experiment, we implemented a perfect strangers' matching such that subjects only ever interact once with each other.

Perfect strangers' matching rule

We told subjects that they only interact once with each other and ran an algorithm to secure the perfect strangers' matching. The fixed algorithm was necessary to prevent unreasonably long waiting time during the experiment. Even with the fixed matching algorithm, random matching was secured by subjects' random choice of login codes to the experimental software SoPHIE (Hendriks, 2012) before the start of the experiment. Subjects were matched three times: Initially, subjects were matched in groups of four for CPR1, then rematched in groups of four for CPR2, and rematched once again for the SVO part. To reduce the impact of indirect reciprocity, the four subjects A, B, C and D per group in the SVO part were matched such that A allocated points to B, B to C, C to D and D to A. Subjects knew that the individual that they allocated points to was a different individual than the one that allocated points to them.

Please see the English translation of the original German instructions including the printed and the on-screen part for further details (Appendix 2F).

Appendix 2B: Details of the implemented user-resource model

The logistic growth term in equation (2B.1) below describes natural growth $G(R_t)$ of the resource R_t that changes over time t, with the resource growth rate $g = 0.04$ and the carrying capacity $R_{max} = 2{,}000$ resource units as long as the resource R is above the threshold R_{min}.

$$R_{t+1} - R_t = \begin{cases} gR_t\left(1 - \frac{R_t}{R_{max}}\right) - \sum_{i=1}^{n} E_{it} & if\ R_t > R_{min} \\ 0 & if\ R_t \leq R_{min} \end{cases} \tag{2B.1}$$

where the resource users' extraction per second $\sum_{i=1}^{n} E_{it}$ is the cumulative extraction of $n = 4$ resource users in one group. Subjects' extraction is cost-free in our experiment and the amount of extracted resource units is constant per level of extraction. For example, an extraction level of 1 always results in 1 extracted resource unit independent of the current resource level.

We implement the threat of a critical threshold by integrating an irreversible and persistent collapse of the resource once subjects' resource extraction drives the resource level R_t to the certain threshold $R_{min} = 500$ resource units: $R_{t+1} = 0\ if\ R_t \leq R_{min}$. The resource R_t stays at the level of 0 resource units infinitely since the growth rate also collapses ($R_{t+1} - R_t = 0$) once the threshold is reached.

In what follows, we outline the calculation of the resource level of maximum sustainable yield (MSY) and the corresponding extraction level. The dynamics of the resource R_t change over time t depending on the natural growth of the resource $G(R_t)$ and the change of the resource through users' extraction per second $\sum_{i=0}^{n} E_{it}$.

The resource's natural growth is based on a logistic growth function (Perman et al., 2011), thus

$$R_{t+1} - R_t = G(R_t) - \sum_{i=0}^{n} E_{it} \tag{2B.2}$$

$$\Leftrightarrow\ R_{t+1} - R_t = gR_t\left(1 - \frac{R_t}{R_{max}}\right) - \sum_{i=0}^{n} E_{it}$$

with the resource level R_t, the instrinsic resource growth rate g and the carrying capacity R_{max}.

Given no harvest by any user, $\sum_{i=0}^{n} E_{it} = 0$, the resource's regrowth is at its maximum when the resource level is at the level of maximum sustainable yield (MSY):

$$\max_{R} \frac{dR}{dt} = gR\left(1 - \frac{R}{R_{max}}\right) \tag{2B.3}$$

$$FOC: g - \frac{2gR}{R_{max}} = 0$$

$$\Leftrightarrow g = \frac{2gR}{R_{max}}$$

$$\Leftrightarrow R = \frac{R_{max}}{2}$$

$$\Rightarrow R_{MSY} = \frac{R_{max}}{2} = 1{,}000$$

To reach a steady state at the resource level of maximum sustainable yield $R_{MSY} = \frac{R_{max}}{2}$ where the natural growth is at its maximum, the resource users' group's extraction $\sum_{i=0}^{n} E_i$ needs to be equal to the amount of the resource's natural growth. Thus,

$$\sum_{i=0}^{n} E_i = g\frac{R_{max}}{2}\left(1 - \frac{\frac{R_{max}}{2}}{R_{max}}\right) \tag{2B.4}$$

$$\Leftrightarrow \sum_{i=0}^{n} E_i = g\frac{R_{max}}{2}\frac{1}{2}$$

$$\Leftrightarrow \sum_{i=0}^{n} E_i = g\frac{R_{max}}{4}$$

$$\Leftrightarrow E_{MSY} = g\frac{R_{max}}{4}$$

$$\Leftrightarrow E_{MSY} = 0.04\frac{2{,}000}{4} = 20$$

Therefore, the optimal joint extraction at $R_{MSY} = 1{,}000$ resource units is $E_{MSY} = 20$.

While the experiment is running, the resource is updated every second such that the resource in the next second R_{t+1} is based on the resource level in the previous second R_t, its growth and the cumulative users' extraction $\sum_{i=0}^{n} E_i$ in the previous second t.

Thus:

$$R_{t+1} = R_t + gR_t\left(1 - \frac{R_t}{R_{max}}\right) - \sum_{i=0}^{n} E_i$$

Parameterisation of the user-resource model

The parameterisation of the resource development in the CPR experiment is based on the following constraints:

I. At the time of the occurrence of the exogenous shock, the current resource level R_t should be higher than the resource level after the implementation of the shock R_{shock}, as well as it needs to be higher than the threshold R_{min}:

$R_t > R_{shock}$ and $R_{shock} > R_{min}$

This constraint needs to hold in the extreme case that each group member choses 10 as extraction level and thus, the group extracts at the maximum rate of 40 resource units per second from the start of the round.

II. Under the assumption of an infinite time horizon, extracting at the highest rate until the resource reaches the resource level of MSY R_{MSY}, followed by the choice of the maximum sustainable extraction level of all group members, i.e. the choice of E_{MSY} at the resource level R_{MSY}, should maximise the joint group outcome.

   a. Joint group maximisation does not mean that all group members need to have the same choice of extraction. As long as the cumulative extraction $\sum_{i=0}^{n} E_i$ is at the level of the MSY extraction $E_{MSY} = g\,\frac{R_{max}}{4}$, some group members can have a lower extraction rate while others might have a higher extraction rate.

III. Choice of the extraction rate that maintains the MSY, once it is reached, avoids reaching the threshold and the collapse of the resource. Under that cumulative extraction choice, the resource should be sustained indefinitely.

a. Implementation of a random continuation rule to set the incentive to choose $E_{MSY} = g\frac{R_{max}}{4}$ once the resource is at R_{MSY} (Dal Bó and Fréchette, 2018).

b. In T2 Exogenous shock, subjects need to adjust their extraction strategy given the sudden reduction of the resource level after the shock. In this instance, group extraction needs to be lower than $E_{MSY} = g\frac{R_{max}}{4}$ (at least for some time) to prevent the collapse of the resource.

Further testing of the experimental design and parameterisation with the help of participants at the “Ocean and Coastal Governance for Sustainability” workshop in Bremen in December 2018 as well as the results and feedback from two pilot sessions run in the LaER laboratory at Osnabrueck University in early 2019 lead to the following parameter choices:

- Carrying capacity of the resource in all three treatments: $R_{max} = 2000$ resource units
    - Corresponds to starting level of the resource in treatment T1 and T2
- Resource level after the occurrence of the exogenous shock in T2: $R_{shock} = 800$
    - Corresponds to starting level of the resource in treatment T3
- Point of time at which the experiment pauses in T1 and the exogenous shock occurs in T2: $t_{shock} = 25$ seconds
- Growth factor of the resource in all three treatments: $g = 0.04$
- Threshold value in all three treatments at $R_{min} = 500$ resource units

Under the assumption of an infinite time horizon these parameterisation leads to:

- $R_{MSY} = \frac{R_{max}}{2} = 1{,}000$ resource units
- $E_{MSY} = g\frac{R_{max}}{4} = 20$ resource units as group extraction per second

Appendix 2C: Normalised group extraction

We use the normalised group extraction (NGE) as a proxy for group cooperation due to the structural differences between treatments (Cerutti and Schlüter, 2019). We calculate the NGE as such:

$$NGE = \frac{total\ group\ extraction}{socially\ optimal\ group\ extraction\ (SGE)}$$

Based on the parameters of our experiment and structural differences between the three treatments, we calculate the socially optimal group extraction (SGE) per treatment.

First, we present the calculation of the NGE using the SGE that is based on the assumption of an infinite game horizon which we created by implementing a random continuation rule (Dal Bó and Fréchette, 2018). Due to the structural differences between the three treatments in CPR1, we have different SGEs between the treatments in CPR1. To create comparability between the three treatments, we calculate the NGE based on the post-pause time (26 to 240 seconds equals a total of 215 seconds) in T1 and T2 in CPR1 and in all three treatments in CPR2. The calculated SGE_{T1} for T1 equals the SGE in CPR2 since the design of CPR2 in all three treatments is based on T1 in CPR1:

$$SGE_{T1} = 40\ units\ \cdot 13\ seconds + 20\ units\ \cdot 202\ seconds = 4{,}560\ units$$

Furthermore, we calculate as $SGE_{T2/T3}$ for T2 and T3 in CPR1:

$$SGE_{T2/T3} = 0\ units\ \cdot 11\ seconds + 20\ units\ \cdot 204\ seconds = 4{,}080\ units$$

Robustness checks based on assumption of endgame effect

Theoretically, the NGE should range between values of zero and one. However, we observe that some groups reach values higher than one. We argue that $NGE > 1$ occurs because subjects do not belief that the game will continue much longer after the certain end of the round and thus, groups cause an endgame effect by overexploiting the resource. Overexploitation based on an endgame effect would result in the resource being driven close to the critical threshold towards the certain end of the round instead of being kept at the level of MSY (1,000 units). The MSY level would be socially optimal under the assumption of an infinite time horizon.

We calculate the $SGE_{endgame}$ under the assumption of an endgame effect that would motivate maximum extraction of 40 units per second starting approximately 25 seconds before the certain end of the round in CPR1 and CPR2 for all three treatments. We calculate as $SGE_{endgame}$ in T1 and thus, for all treatments in CPR2:

$$SGE_{endgame\ T1} = 40\ units\ \cdot 13\ seconds + 20\ units\ \cdot 177\ seconds + 40\ units \cdot 25\ seconds$$
$$= 5{,}060\ units$$

The $SGE_{endgame\ T2,T3}$ for T2 and T3 in CPR1 is:

$$SGE_{endgame\ T2,T3} = 0\ units\ \cdot 11\ seconds + 20\ units\ \cdot 179\ seconds + 40\ units\ \cdot 25\ seconds$$
$$= 4{,}580\ units$$

Table 2C.1 presents the results for the NGE analysis based on the assumption of an endgame effect, which are in line with our main findings. Thus, we conclude that a potential endgame effect does not drive our main results. Furthermore, we test whether groups that fail to coordinate and cause a collapse of the resource prior to the certain end drive our main results by excluding them from the analysis (defined as “excluding collapsed groups” in Table 2C.1). Again, we find no evidence that groups that collapsed drive our main results on cooperation.

Table 2C.1. Robustness checks of cooperation results based on the normalised sum of group extraction.

	T1 Control				T2 Exogenous shock				T3 Low resource				Mann-Whitney-Wilcoxon test (p-value)[a]
	N	Mean (SD)	Min	Max	N	Mean (SD)	Min	Max	N	Mean (SD)	Min	Max	
CPR1 excluding collapsed groups[b]	20	1.05 (0.02)	1.00	1.08	27	1.03 (0.02)	1.00	1.07	17	1.00	0.94	1.04	0.030** 0.004*** 0.000***
CPR1 endgame effect[c]	24	0.91 (0.09)	0.67	0.97	34	0.87 (0.15)	0.34	0.96	22	0.83 (0.14)	0.47	0.93	0.002*** 0.007*** 0.000***
CPR2 excluding collapsed groups	17	1.04 (0.4)	0.96	1.08	27	1.04 (0.03)	0.97	1.08	20	1.04 (0.03)	0.98	1.09	0.838 0.659 0.891
CPR2 endgame effect	24	0.86 (0.15)	0.47	0.97	34	0.89 (0.10)	0.53	0.97	22	0.90 (0.15)	0.43	0.98	0.613 0.343 0.267

Note: N denotes number of group observations. Standard deviations (SD) are presented in parentheses. Min refers to the lowest observed value and max to the highest observed value.
[a]The **p-values** of the pairwise Mann-Whitney-Wilcoxon tests are in the following order: T1 vs. T2; T2 vs. T3 and T3 vs. T1.
[b]**Excluding collapsed groups** indicates that all groups that caused a collapse of the resource prior to the end of the round were excluded from this robustness check. Thus, the N is smaller.
[c]**Endgame effect** indicates that the normalised sum of group extraction is calculated based on the assumption that groups expected the round to end immediately at the end of the certain round length. An endgame effect would have caused them to drive the resource to the threshold to collapse right at 240 seconds (215 seconds in T3).
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Appendix 2D: Details of sample's socio-economic characteristics

Details of the sample's socio-economic characteristics and the controls of the understanding of the game and elicited preferences are presented in Table 2D.1. The vast majority of subjects is below the age of 35 years (Fig. 2D.1). However, we have two subjects of the age of 44 and 46 in treatment T2 and three subjects of the age 44, 53 and 65 in T3 who drive the significant difference in the average age across treatments. We find a significant difference between treatments T1 and T2 (MWW: p = 0.015) and T3 (p = 0.016).

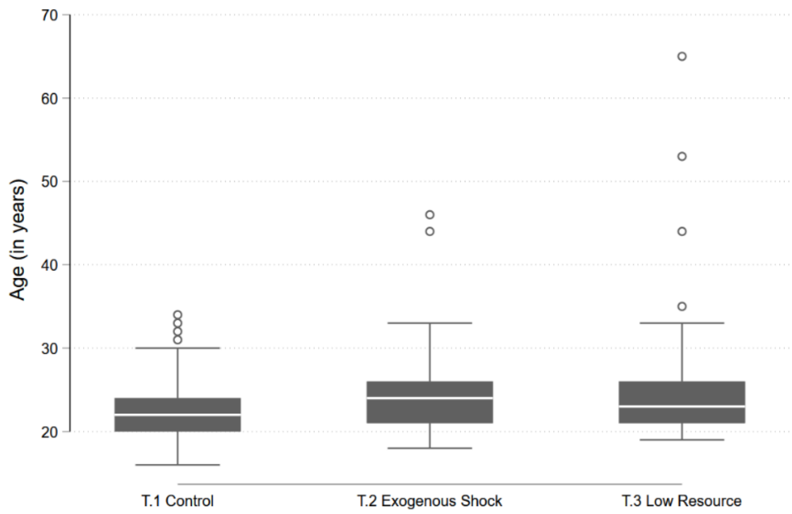


Fig. 2D.1 Boxplot of age distribution per treatment. The outlier values are the unusually old subjects (older than 30 years) that especially participated in T2 and T3.

Risk preferences

As part of the post-experimental questionnaire, we elicited subjects risk preferences with the non-incentivised risk question (Dohmen et al., 2011). The risk question asks subjects how willing they are in general to take risk and subjects can state any value between 0 (risk averse) to 10 (risk loving). We do not find evidence for a significant difference in subjects' risk preferences between treatments (Table 2D.1).

Social preferences

We implemented the Social Value Orientation (SVO) based on Murphy et al. (2011). Based on the first six primary items of the 15 distributive decisions of the SVO slider measure that subjects made, we calculate the SVO index for each subject. The higher the value of the SVO index the more prosocial is a subject. Altruists have an SVO index higher than 57.15, prosociality is defined between 22.45° and 57.15. Individualists are subjects between - 12.04 and 22.45, while subjects are classified as competitive if their SVO index is below - 12.04. We find the majority of our subjects classifies as prosocial or individualistic and there is no evidence for statistically significant differences between treatments (Table 2D.1).

Environmental preferences

We used the New Environmental Paradigm (NEP) scale to elicit subjects' ecological worldviews (Dunlap et al., 2000). The scale consists of 15 items to cover 5 different aspects of the ecological worldview: (1) the reality of limit of growth, (2) antianthropocentrism, (3) the fragility of nature's balance, (4) rejection of exemptionalism and (5) the possibility of an ecocrisis (ibid). We used the German translation of the items that has been tested and is provided by Schleyer-Lindenmann et al. (2018). We find no evidence for statistically significant differences in subjects' ecological worldview between treatments (Table 2D.1).

Understanding of the experiment

Subjects' understanding of the experimental instructions was measured by their answers to the eleven control questions that they answered before the start of the two CPR group rounds. The control question score denotes the mean number of tries that subjects needed to answer all questions correctly. A score of one means that the subject answered all eleven questions correctly at the first try. Subjects were not able to continue with the game before giving the correct answer to each question. We find no evidence for significant differences in subjects' understanding of the CPR game between treatments (Table 2D.1)

All control questions and the post-experimental questionnaire including all feedback questions are presented in the full set of the instructions (English translation of German original) in Appendix 2F.

Table 2D.1. Balance table of socio-economic characteristics

	T1 Control			T2 Exogenous shock			T3 Low resource				
	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Fisher's exact tests[a]	MWW tests[a]
Age (years)	22.802 (3.320)	16	34	23.838 (3.918)	18	46	24.909 (6.770)	19	65		0.015** 0.702 0.016**
Female (fraction)	0.563			0.515			0.591				0.473 0.264 0.698
Number of previous experiments	3.198 (3.217)	0	15	3.691 (2.993)	0	15	3.648 (2.905)	0	15		0.113 0.943 0.123
Income in Euro per month (fractions)										0.194 0.955 0.160	
0 – 300	0.146			0.132			0.125				
301 - 600	0.438			0.324			0.341				
601 - 900	0.313			0.338			0.296				
more than 900	0.063			0.147			0.171				
not specified	0.042			0.059			0.068				
Student status (fraction)	0.979			0.963			0.943			0.703 0.520 0.262	
Field of study (fractions)	N=94			N=131			N=83			0.615 0.930 0.701	
Humanities	0.160			0.130			0.145				
Social sciences	0.075			0.069			0.084				
Economics	0.319			0.267			0.289				
Natural sciences	0.223			0.244			0.253				
Engineering	0.178			0.115			0.072				
Other	0.096			0.176			0.157				
Control question score[b]	1.091 (0.144)	1	2.091	1.144 (0.396)	1	5.273	1.092 (0.159)	1	2.091		0.388 0.253 0.786
Risk measure (0 – risk averse; 10 – risk loving)	5.490 (2.201)	0	10	5.331 (2.215)	0	10	5.330 (2.268)	0	10		0.524 0.986 0.594
Social value orientation index[c]	17.802 (13.267)	-7.815	61.390	18.022 (13.281)	-7.815	45.892	17.590 (13.419)	-9.943	45		0.509 0.698 0.914

Table 2D.1. Continued. Balance table of socio-economic characteristics

	T1 Control			T2 Exogenous shock			T3 Low resource				
	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Fisher's exact tests[a]	MWW tests[a]
New ecological paradigm scale[d]	3.852 (0.448)	2.467	4.8	3.825 (0.484)	2.2	4.867	3.889 (0.467)	1.867	4.733		0.651 0.337 0.604
Proportion of sessions run by female experimenter[e]	60.00			57.14			60.00			1.000	

Note: We run 5 sessions with 96 subjects in 24 groups in T1, 7 sessions with 136 subjects in 34 groups in T2 and 5 sessions including 88 subjects in 22 groups in T3. N denotes the number of individuals that answer the question. Standard deviations are presented in brackets.

[a] The p-values of the pairwise Fisher's exact tests (FET) and the pairwise Mann–Whitney–Wilcoxon tests (MWW) are ordered as such: T1 vs. T2; T2 vs. T3 and T3 vs. T1.

[b] Subjects answered eleven control questions before playing the two CPR rounds. The control question score is the mean number of tries that subjects needed to answer the questions correctly. A score of one means that the subject answered all eleven questions correctly at the first try. [c]The Social Value Orientation index is a measure of subjects' social preferences. Higher values of the index indicate pro-sociality whereas low values indicate competitive preferences. [d]The New Ecological Paradigm (NEP) Scale presents subjects' agreement with NEP statements on a scale from 1 (strong disagreement) to 5 (strong agreement). Higher means indicate that subjects have a more pro-ecological worldview. [e]Two different experimenters, one female and one male, conducted the experimental sessions.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Appendix 2E: Robustness checks of main results

In Table 2E.1 and 2E.2, we present regression results of Probit and Tobit models as robustness checks of our main results on both *coordination* and *cooperation*.

As a robustness check of our coordination analysis, we first run Probit regressions with the probability of causing a collapse of the resource as dependent variable and group's socio-economic characteristics including groups' mean age as control variables (Table 2E.1). As before, we find no statistically significant treatment effects of T2 and T3 in comparison to T1 on the probability to cause a collapse (Table 2E.1, column 1 and 2).

If we control for groups' socio-economics characteristics (Table 2E.1, column 2), we find that a higher mean group age makes the groups more likely to collapse ($p<0.05$). Additionally, a higher SVO index and NEP level significantly decrease ($p<0.1$ and $p<0.05$ respectively) the likelihood to cause a collapse. However, the treatment effects are still insignificant and in line with our main results.

Second, we run Tobit regressions with the normalised sum of group extraction (NGE) as dependent variable (Table 2E.1). Models 3 and 5 include all groups while models 4 and 6 only include groups that did not cause a collapse in CPR1. We find a decreasing, insignificant treatment effect for T2 in model 3 ($p=0.142$). However, T2 decreases the NGE in CPR1 significantly in the sub-sample that did not cause a collapse of the resource and coordinated successfully (Table 2E.1, column 4). The treatment effect of T3 is significantly negative for both samples; including and excluding the groups that failed to coordinate successfully (Table 2E.1, column 3 and 4). However, both treatment effects are statistically insignificant if we include control variables in model 5 (sample includes all groups), but statistically significant in the sub-sample excluding groups that failed to coordinate (Table 2E.1, column 5 and 6). Overall, our finding that shock experience and initial resource scarcity cause a decrease in *cooperation* in CPR1 is robust.

Analysing the *spillover effect* of previous shock experience in CPR2, we find no evidence for a significant treatment effect of T2 on NGE in CPR2 (Table 2E.2), neither with regard to *coordination* (column 1 and 2) nor *cooperation* (column 3 to 6). Again, this finding is in line with our main results and robust to the inclusion of groups' socio-economics characteristics as control variables.

Table 2E.1. Probit and Tobit models as robustness checks on coordination and cooperation results of analysis for CPR1

	Outcome variable Probit models (1) and (2): Probability to cause a collapse of the resource (1=collapse, 0 otherwise); **Outcome variable Tobit models (3) and (4):** Groups' normalised sum of group extraction (NGE)					
Variables	**(1)** Probit	**(2)** Probit	**(3)** Tobit	**(4)** Tobit	**(5)** Tobit	**(6)** Tobit
T.2 Exogenous Shock	0.15 [-0.50 - 0.79] (0.66)	-0.07 [-0.69 - 0.54] (0.81)	-0.04 [-0.08 - 0.01] (0.14)	-0.02** [-0.03 - - 0.00] (0.03)	-0.00 [-0.05 - 0.04] (0.91)	-0.01** [-0.03 - - 0.00] (0.04)
T.3 Low Resource	0.22 [-0.32 - 0.76] (0.43)	0.00 [-0.84 - 0.85] (0.99)	-0.08*** [-0.13 - -0.04] (0.00)	-0.04*** [-0.06 - - 0.02] (0.00)	-0.04 [-0.10 - 0.03] (0.23)	-0.04*** [-0.06 - - 0.02] (0.00)
Group age[a]		0.14*** [0.04 - 0.24] (0.01)			-0.02*** [-0.04 - - 0.01] (0.00)	-0.00 [-0.00 - 0.00] (0.98)
Group female fraction[b]		-1.24 [-3.09 - 0.61] (0.19)			0.11 [-0.04 - 0.27] (0.14)	0.01 [-0.02 - 0.03] (0.65)
Group risk preference[c]		-0.05 [-0.44 - 0.34] (0.80)			0.01 [-0.04 - 0.05] (0.73)	0.00 [-0.01 - 0.01] (0.87)
Group CQ score[d]		1.06 [-0.89 - 3.01] (0.29)			0.01 [-0.14 - 0.17] (0.86)	-0.01 [-0.08 - 0.07] (0.86)
Group SVO index[e]		-0.04** [-0.08 - -0.00] (0.05)			0.00 [-0.00 - 0.01] (0.79)	0.00 [-0.00 - 0.00] (0.28)
Group NEP value[f]		-1.22* [-2.64 - 0.20] (0.09)			0.08 [-0.02 - 0.18] (0.13)	-0.01 [-0.03 - 0.01] (0.46)

Table 2E.1. Continued. Probit and Tobit models as robustness checks on coordination and cooperation results of analysis for CPR1.

Variables	**(1)** Probit	**(2)** Probit	**(3)** Tobit	**(4)** Tobit	**(5)** Tobit	**(6)** Tobit
Female experimenter (1=yes, 0=no)		-0.56* [-1.14 - 0.01] (0.05)			0.02 [-0.03 - 0.07] (0.34)	-0.00 [-0.01 - 0.01] (0.83)
Constant	-0.97***	1.20	1.01***	1.05***	1.11***	1.07***
	[-1.29 - -0.65]	[-4.68 - 7.09]	[0.99 - 1.03]	[1.03 - 1.06]	[0.36 - 1.86]	[0.95 - 1.20]
	(0.00)	(0.69)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	80	80	80	64	80	64
left-censored			0	0	0	0
Prob > chi2	0.710	0.000				
Prob > F			0.001	0.000	0.000	0.000
Pseudo R2	0.00	0.24	-0.05	-0.09	-0.33	-0.09

Note: In the Probit models (column 1 and 2) the dependent variable is the probability to cause a collapse of the resource in CPR1. In the Tobit models (column 3 to 6) the dependent variable is groups' normalised sum of group extraction (NGE) in CPR1. Tobit models 3 and 5 include all groups including the groups that caused a collapse of the resource prior to the known end of CPR1, while models 4 and 6 only include the groups that coordinated successfully and did not cause a collapse. All models are based on group outcomes and all control variables (besides "female experimenter") are the group means of individual outcomes.
[a]**Group age** is the mean age of all four group members. [b]**Group female fraction** is the gender ratio per group. [c]**Group risk preference** is the mean value of group members' individual risk preference that was stated between 0 = risk averse and 10 = risk liking. [d]**Group CQ score** is the mean value of group members' individual control question score. The higher the score, the more tries group member needed on average to answer all eleven control questions correctly. [e]**Group SVO index** is the mean of group members' individual Social Value Orientation (SVO) index. The SVO index is a measure of subjects' social preferences. Higher values of the index indicate pro-sociality whereas low values indicate competitive preferences. [f]**Group NEP value** is the mean value of group members' individual New Ecological Paradigm (NEP) Scale values. The NEP scale measures subjects' agreement with NEP statements on a scale from 1 (strong disagreement) to 5 (strong agreement). Higher values indicate that group members' have a more pro-ecological worldview. [g]Two different experimenters, one female and one male, conducted the experimental sessions.
The lower censoring limit for the dependent variable NGE of the Tobit models was zero.
Clustering was done at the session level (17 clusters). Standard errors are displayed in brackets and adjusted for 17 clusters. Confidence intervals (at the 95% confidence level) are presented in square brackets.
* p < 0.1; ** p < 0.05; *** p < 0.01

Table 2E.2 Probit and Tobit models as robustness checks on coordination and cooperation results of analysis for CPR2.

	Outcome variable Probit models (1) and (2): Probability to cause a collapse of the resource (1=collapse, 0 otherwise); **Outcome variable Tobit models (3) and (4):** Groups' normalised sum of group extraction (NGE)					
Variables	(1) Probit	(2) Probit	(3) Tobit	(4) Tobit	(5) Tobit	(6) Tobit
T.2 Exogenous Shock	-0.27 [-0.92 - 0.38] (0.41)	-0.14 [-0.87 - 0.59] (0.70)	0.04 [-0.04 - 0.11] (0.30)	0.00 [-0.01 - 0.02] (0.90)	0.04 [-0.01 - 0.10] (0.14)	0.00 [-0.01 - 0.01] (0.55)
Group age[a]		-0.11 [-0.34 - 0.11] (0.32)			-0.00 [-0.01 - 0.01] (0.69)	0.00 [-0.00 - 0.01] (0.87)
Group female fraction[b]		-0.65 [-2.13 - 0.84] (0.39)			0.13** [0.02 - 0.24] (0.03)	0.03 [-0.02 - 0.08] (0.18)
Group risk preference[c]		0.16 [-0.11 - 0.43] (0.26)			-0.00 [-0.03 - 0.02] (0.76)	0.00 [-0.01 - 0.02] (0.68)
Group CQ score[d]		-1.29 [-3.89 - 1.30] (0.33)			0.08 [-0.03 - 0.20] (0.16)	-0.01 [-0.03 - 0.02] (0.61)
Group SVO index[e]		-0.05 [-0.12 - 0.02] (0.15)			0.01** [0.00 - 0.01] (0.04)	-0.00 [-0.00 - 0.00] (0.76)
Group NEP value[f]		-0.84 [-2.58 - 0.90] (0.34)			0.02 [-0.13 - 0.16] (0.81)	0.00 [-0.05 - 0.05] (0.98)
Female experimenter (1=yes, 0=no)		0.32 [-0.43 - 1.07] (0.41)			-0.05** [-0.10 - - 0.00] (0.03)	0.01 [-0.00 - 0.03] (0.13)
Constant	-0.55** [-1.02 - - 0.08] (0.02)	6.80 [-2.66 - 16.26] (0.16)	0.95*** [0.88 - 1.02] (0.00)	1.04*** [1.03 - 1.05] (0.00)	0.69*** [0.21 - 1.17] (0.01)	0.99*** [0.76 - 1.23] (0.00)
Observations	58	58	58	44	58	44
left-censored			0	0	0	0
Prob > chi2	0.41	0.00				
Prob > F			0.30	0.90	0.31	0.00
Pseudo R2	0.01	0.09	-0.01	-0.00	-0.17	-0.03

Note: In the Probit models (column 1 and 2) the dependent variable is the probability to cause a collapse of the resource in CPR2. In the Tobit models (column 3 to 6) the dependent variable is groups' normalised sum of group extraction (NGE) in CPR2. Tobit models 3 and 5 include all groups including the groups that caused a collapse of the resource prior to the known end of CPR2, while models 4 and 6 only include the groups that coordinated successfully and did not cause a collapse. All models are based on group outcomes and all control variables (besides "female experimenter") are the group means of individual outcomes.
[a]**Group age** is the mean age of all four group members. [b]**Group female fraction** is the gender ratio per group. [c]**Group risk preference** is the mean value of group members' individual risk preference that was stated between 0 = risk averse and 10 = risk liking. [d]**Group CQ score** is the mean value of group members' individual control question score. The higher the score, the more tries group member needed on average to answer all eleven control questions correctly. [e]**Group SVO index** is the mean of group members' individual Social Value Orientation (SVO) index. The SVO index is a measure of subjects' social preferences. Higher values of the index indicate pro-sociality whereas low values indicate competitive preferences. [f]**Group NEP value** is the mean value of group members' individual New Ecological Paradigm (NEP) Scale values. The NEP scale measures subjects' agreement with NEP statements on a scale from 1 (strong disagreement) to 5 (strong agreement). Higher values indicate that group members' have a more pro-ecological worldview. [g]Two different experimenters, one female and one male, conducted the experimental sessions.
The lower censoring limit for the dependent variable NGE of the Tobit models was zero.
Clustering was done at the session level (12 clusters). Standard errors are displayed in brackets and adjusted for 12 clusters. Confidence intervals (at the 95% confidence level) are presented in square brackets.
* p < 0.1; ** p < 0.05; *** p < 0.01

Appendix 2F: Instructions of the experiment

English translation of the original German experimental instructions. German original is available upon request. Additional information and explanations of experimental processes are marked in italics. The first part of the instructions was printed and read aloud by the experimenter. Horizontal lines _____ mark the switch to the next screen/step of the experiment as programmed in the experimental software SoPHIE (Hendriks 2012).

Instructions of the CPR game

Welcome to this experiment

You are now taking part in an economic experiment. Please read the instructions carefully. The decisions you and the other participants make during the experiment, determine your payout at the end of the experiment. Since your earnings depend on your choices, it is important that you read and understand the instructions carefully.

All data that we collect during this experiment is treated with great care and confidentiality.

You make your decisions anonymously and it is impossible to connect your choices to your true identity.

The instructions that you receive from us are for your own private information. In economics experiments like this one, the experimenter is not allowed to deceive participants, hence, all information given to you in the instructions are true. There are no correct or incorrect decisions in this experiment. You are asked to decide based upon your own personal preferences.

You are not allowed to talk to any other participant during the experiment, and/or use communication devices such as your mobile phone. Please turn your communication devices off and put them in your bags. If you have a question, please raise your hand and a member of the support staff will come and answer your question privately. If you fail to comply with these rules or cause undue disruptions, we reserve the right to exclude you from the experiment and all payouts.

Experimental procedure: This experiment consist of two parts: part 1 and part 2, and will take approximately 90 minutes in total. Initially, you will only receive the instructions for part 1 of the experiment. Upon completion of part 1, you will receive the instructions for part 2 of the experiment on screen.

Each part of the experiment is independent from the other. Decisions that you make in part 1 have no impact on any probabilities or payouts in part 2 of the experiment.

Payout: The payout you receive at the end of the experiment is based on the earnings from part 1 and part 2 plus a fixed 3€ attendance fee. You will learn the specific method that determines your payout for part 1 and part 2 of the experiment in the instructions for each part. At the end of the entire experiment, we will inform you of your total payout.

Your earnings are referred to as points during the experiment, not Euro. All your earnings are calculated in points and your total sum of points will be converted into Euro at the end of the experiment, where your 3€ attendance fee will be added. The exchange rate for the points to Euro conversion are presented in the instructions for the individual parts of the experiment.

To receive your cash payout after the experiment, please present your participation code that you got for your login at the beginning of the experiment. The payouts will be handed out individually and anonymously.

Instructions part 1

In part 1 of the experiment, you manage a renewable resource in a group of four people. The resource develops dynamically over time and your task is to decide how many resource units you want to extract from the resource during a given time period. You can change the amount that you extract at any time during the experiment.

The resource units that you extract determine the number of points that you collect for your payout.

One resource unit equals one point (1 resource unit = 1 point).

The more resource units you extract the higher your number of points and thus your payout for the corresponding round.

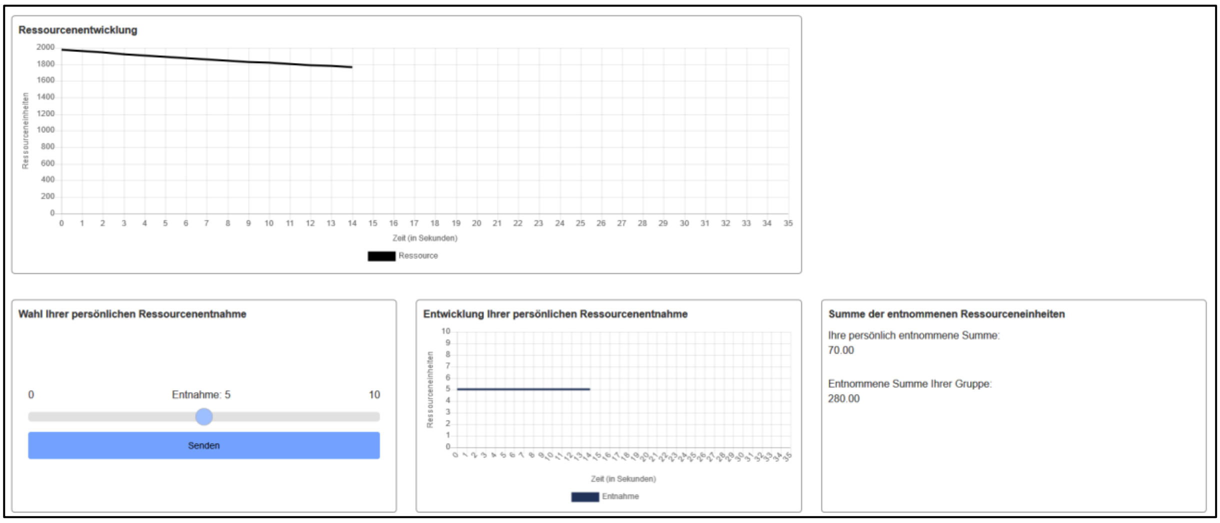
Part 1 consists of four rounds in total. The exact end of each round will be determined randomly and is not known at the start of each round. The first two rounds are test rounds, allowing you to familiarise yourself with the experimental set-up. Your decisions in these **two test rounds** do not have any impact on your payout.

Following the two test rounds, you make decisions in **two potentially relevant payout rounds**. Only one of these two payout-relevant rounds will be randomly selected to determine your payout at the end of the experiment. The points you have earned based on your decisions made in that randomly chosen round are converted into Euro and given to you in cash at the end of the experiment. In this part 1 of the experiment, the exchange rate for points in Euro is **100 points = 0.40 Euro.**

The decisions that you make in one of the two test rounds will not influence the development of the resource or any probabilities in the payout-relevant rounds. Further, the decisions that you make in the first payout-relevant round do not influence the development of the resource or any probabilities in the second payout-relevant round. Each round of the four rounds in the first part is independent of the others.

Resource dynamics and resource extraction procedure

This is a screenshot of the screen as it will be presented to you during the experiment:



The box **"development of the resource" (Ressourcenentwicklung)** in the top left corner of the screen shows the development of the resource of your group. The vertical axis with the caption "Resource units (Ressourceneinheiten)" presents the level of the resource in resource units from 0 to 2,000. The time is displayed in seconds on the horizontal axis. The graph of the resource develops in the frame that you can see from the beginning for the first 30 seconds. From 30 seconds onwards, the numbers that label the seconds on the x-axis will change over time. As mentioned before, the end of each round is determined randomly and not known at the beginning.

You can choose the number of resource units that you want to extract per second via the slider in the box **"choice of your personal resource extraction" (Wahl Ihrer persönlichen Ressourcenentnahme)** (see Figure 2F.1, box on the bottom left). You can choose your personal resource extraction from the numbers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 resource units per second.

For example, if you choose 5 as your personal resource extraction per second and you keep that choice for 14 seconds, you extract 5 resource units each second. Therefore, over 14 seconds it would be $5 * 14 = 70$, which gets you 70 points. You can change your personal extraction per second anytime throughout each round.

Resource dynamics – Growth of the resource

The resource development is dynamic such that the level of the resource in one second depends on the amount of resource units present in the previous second. This means that your extraction of resource units in the one second has an impact on the resource level in the following second.

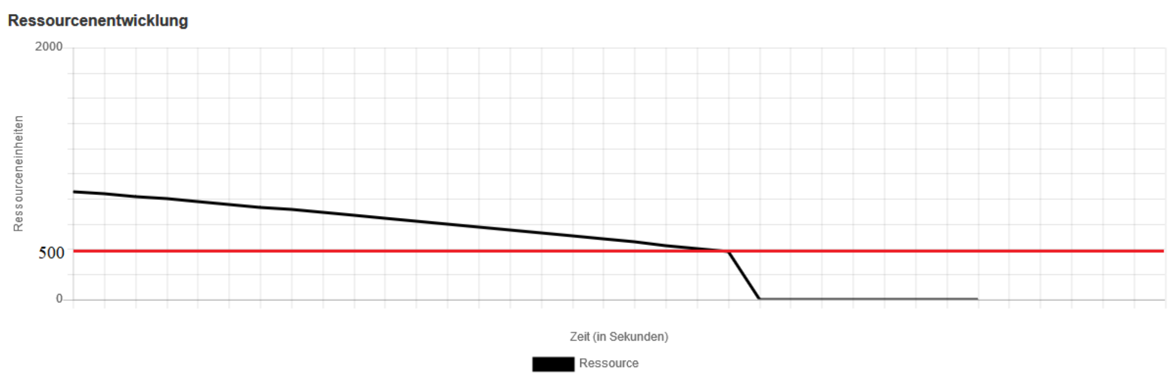
For example, the resource grows by 15 resource units in one second if the level of the resource is at 1,500 resource units. In case that the group extracts 40 resource units in that same second, the level of the resource in the next second would be:

$1.500 + 15 - 40 = 1.475$ resource units

The resource is at its maximum resource level of capacity at a level of 2,000 resource units. The resource cannot grow higher than 2,000 resource units.

Further, there is a threshold in the development of the resource. As soon as the resource level reaches the threshold of 500 resource units, which means the resource level is smaller or equal to 500 resource units, the resource immediately drops down to 0 resource units. Additionally, once the resource is at 0 resource units, it stops growing and the growth rate is changed to 0 resource units per second.

The graph below shows an example of the presentation of the resource development on screen once the critical threshold of 500 resource units is reached (black line). The vertical, red line is only visible in this example to outline the critical threshold at 500 resource units. It will not be visible on screen during the game.



Once the critical threshold of the resource is reached, the resource stays at the level of zero resource units with zero growth until the time of the round is up. This means that the development of the resource comes to an irreversible stop. At this point, changes of your extraction level do not have any impact on the development of the resource.

Please be aware that until the end of the round neither you nor any other person in your group is able to extract additional resource units at this point. It is impossible to collect further points in this round.

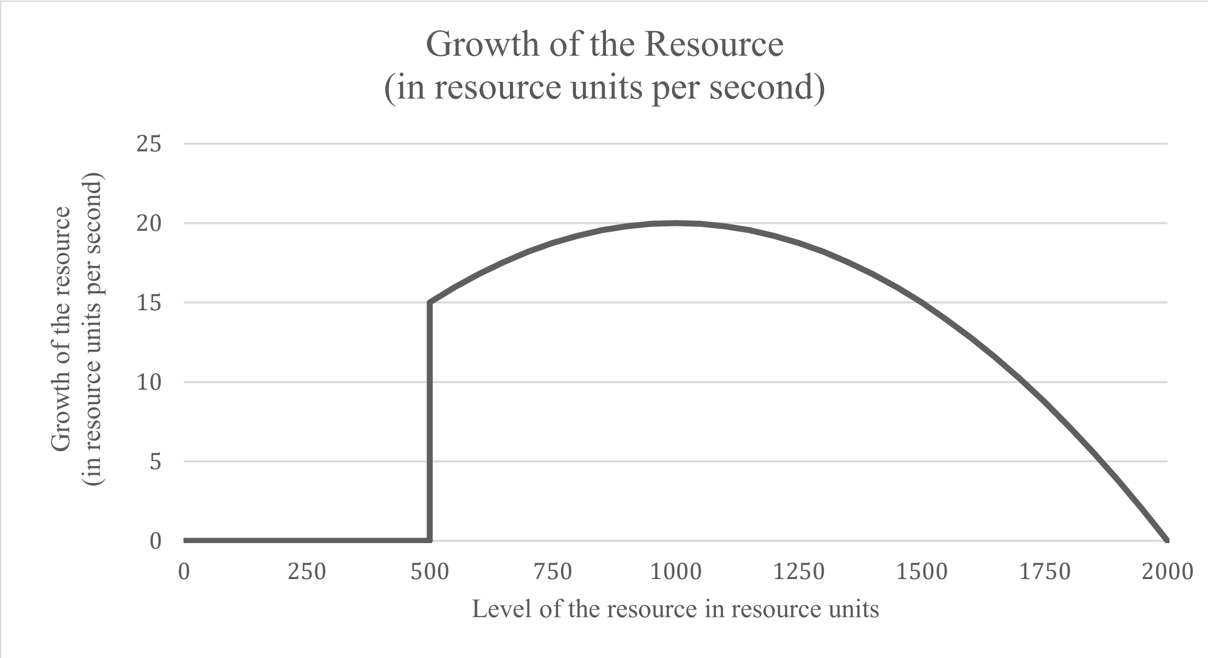
An overview of different growth rates of the resource at different resource levels is given in the following table:

level of the resource (in resource units)	**growth of the resource (in resource units per second)**
2000	0
1900	3.8
1800	7.2
1700	10.2
1600	12.8
1500	15
1400	16.8
1300	18.2
1200	19.2
1100	19.8
1000	20
900	19.8
800	19.2
700	18.2
600	16.8
500	0
0	0

Example 1: The resource does not grow at a level of 2,000 resource units since it is at its maximum at that level. The growth is at 0 resource units per second.

Example 2: At a level of 1,000 resource units, the growth of the resource is at its highest. It grows with 20 resource units per second.

Please be aware that the table above only presents the possible levels of the resource between 2,000 and 500 resource units in steps of 100. The following graph shows the growth of the resource for all the levels of the resource between these values. The development of the resource and also the graph of the growth of the resource drops down to 0 resource units per second as soon as the threshold of 500 resource units is reached.



You will now receive further explanations on the use of the slider to set your personal level of extraction per second and the general set-up of the experiment on screen.

End of the printed instructions on paper and transition to the instructions that are presented on screen in the SoPHIE format. The experimenter prompts subjects to login at the computers and asks if anyone has any questions about the instructions. Once all questions are clarified, the instructions on screen are started.

Instructions – Part 1

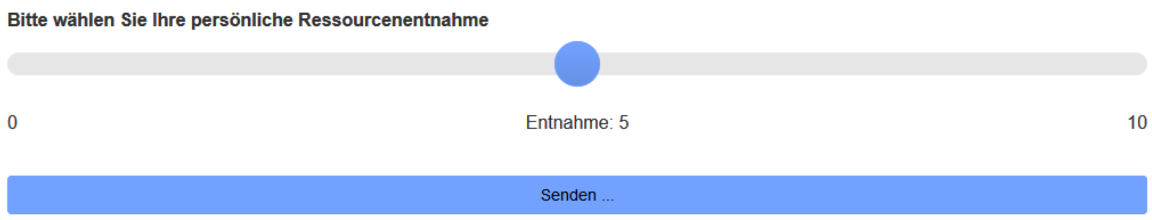
The following instructions explain the use of the slider to choose your personal extraction level. You will also receive further information about the other boxes displayed on the screen during the experiments.

You will play two individual test rounds before the start of the payout-relevant rounds to familiarise yourself with the reaction of the resource development to different extraction levels and the general development of the experiment.

The decisions that you make in the two test rounds of the first part do not have any impact on the development of the resource or any probabilities in the payout-relevant rounds.

Information on “Your personal choice of resource extraction”

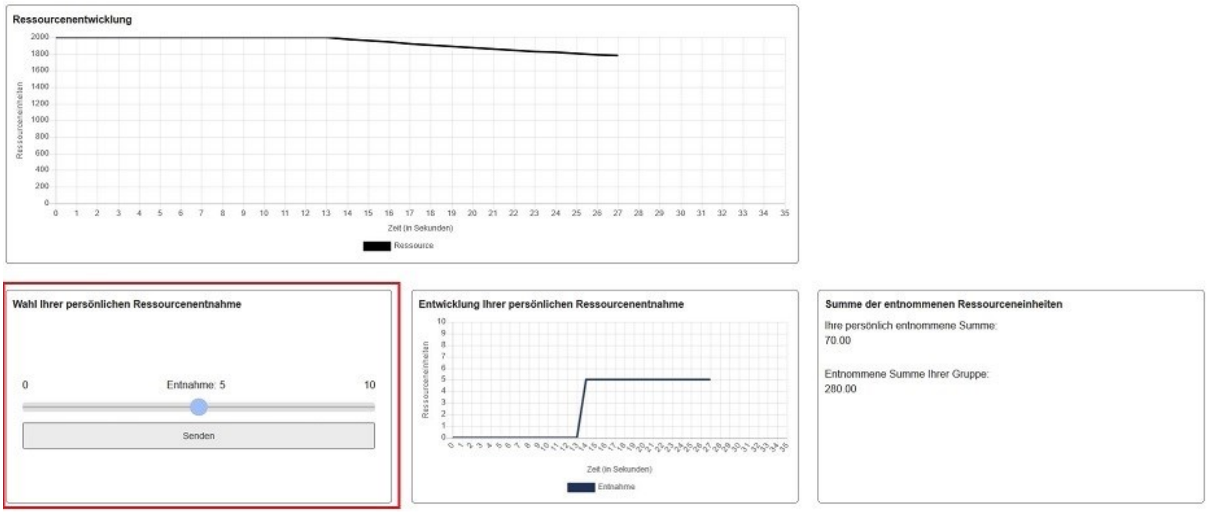
You are asked to choose the starting level of your personal extraction per second prior to the start of each round. The development of the resource starts at a level of 2,000 resource units *(800 resource units in treatment T.3 Low resource)* as soon as every member of your group chose their extraction level. The sum of extraction per second of your group is executed from the beginning.



You choose your starting level of your personal extraction with a slider that looks like the one shown in the above figure. At first, the blue dot on the slider is invisible. You have to click on the grey bar above the “Continue (Senden)” button to make the blue dot appear. You can freely move this blue dot to the left or right to adjust the level of your personal resource extraction between 0 and 10 extracted resource units per second. The text under the grey bar clearly states the exact number of resource units that are extracted when the corresponding spot on the grey bar is chosen via positioning the blue dot on it. The example in the screenshot shows “Extraction (Entnahme): 5”, which means that you extract 5 resource units per second if you choose that position on the slider.

As soon as you have chosen your preferred level of resource extraction, you have to confirm your choice by clicking “Send”. The round of the extraction game will start as soon as all members of your group clicked on send.

Information on “Your personal choice of resource extraction”



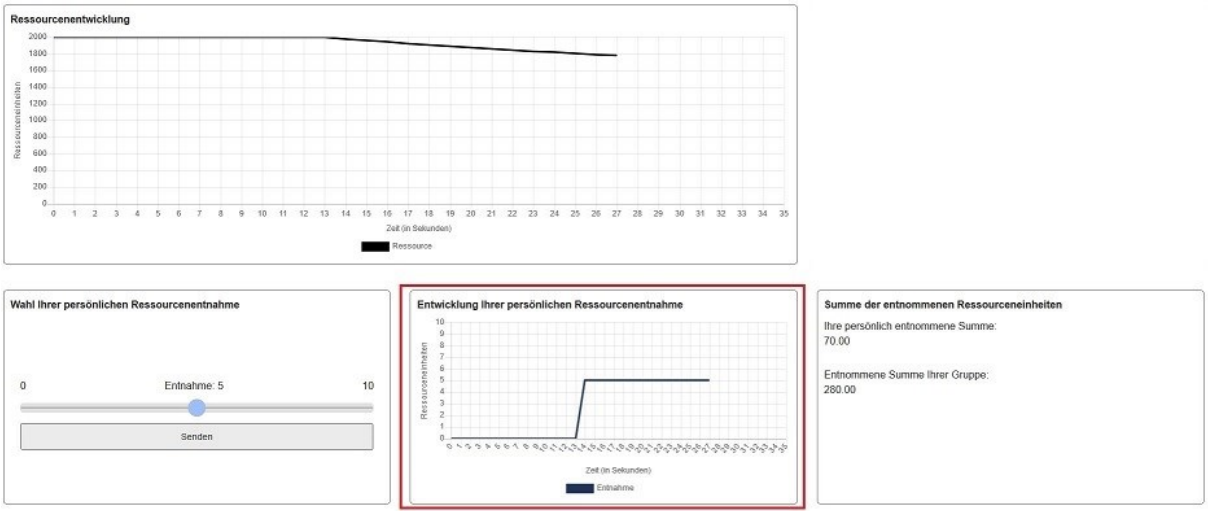
At the start of each round, you will extract the amount of resource units per second that you choose by the slider. Your starting level of extraction is continuously executed until you make an active change.

The box in the bottom corner on the left (see Figure) presents your personally chosen starting level of your extraction on the slider. You can change your chosen level of resource extraction anytime during the round by moving the slider (blue dot) to the left or the right and by confirming your new choice with a click on “send”.

Please be aware that you have to click “send” to confirm your choice. If you do not click “send”, your choice of resource extraction is not implemented.

Your chosen level of resource extraction is executed every second until your change your level and confirm that change by clicking on “send”.

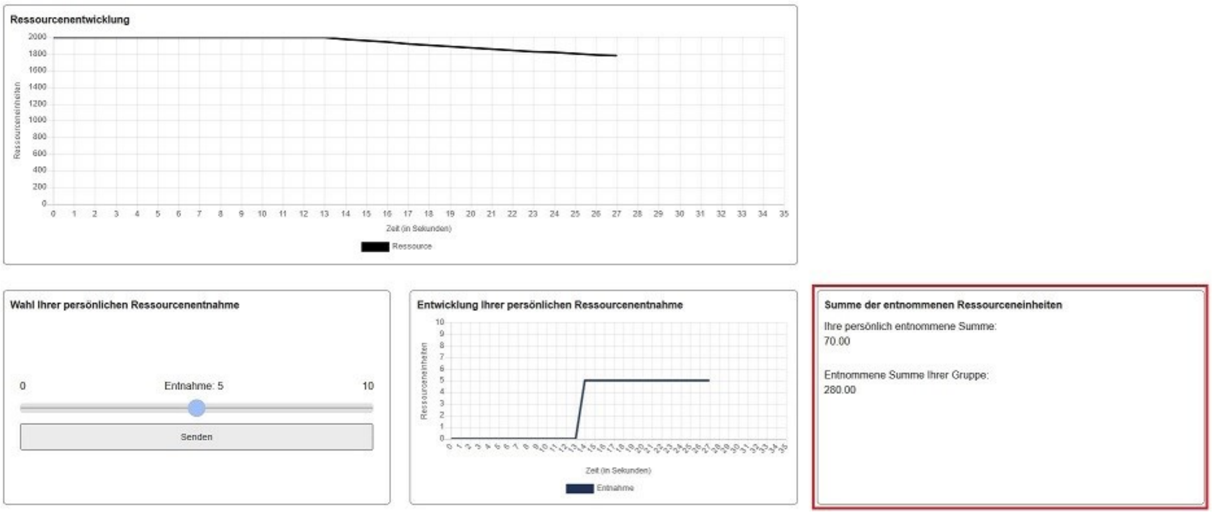
Information on “Development of your personal resource extraction”



The graph “**Development of your personal resource extraction**” (box in the middle of the bottom line) shows you if your extraction made with the slider is implemented. If you move the blue dot to the right to choose a higher resource extraction per second than before, the graph increases. If you move the blue dot to the left to choose a lower resource extraction per second than before, the graph decreases.

Please click on “send” again, in case that you do not observe a change in the graph after you changed your level of resource extraction per second with the slider. You need to confirm your choice by clicking on “send”, otherwise the change is not implemented.

Information on “Sum of extracted resource units”: Your own individual sum

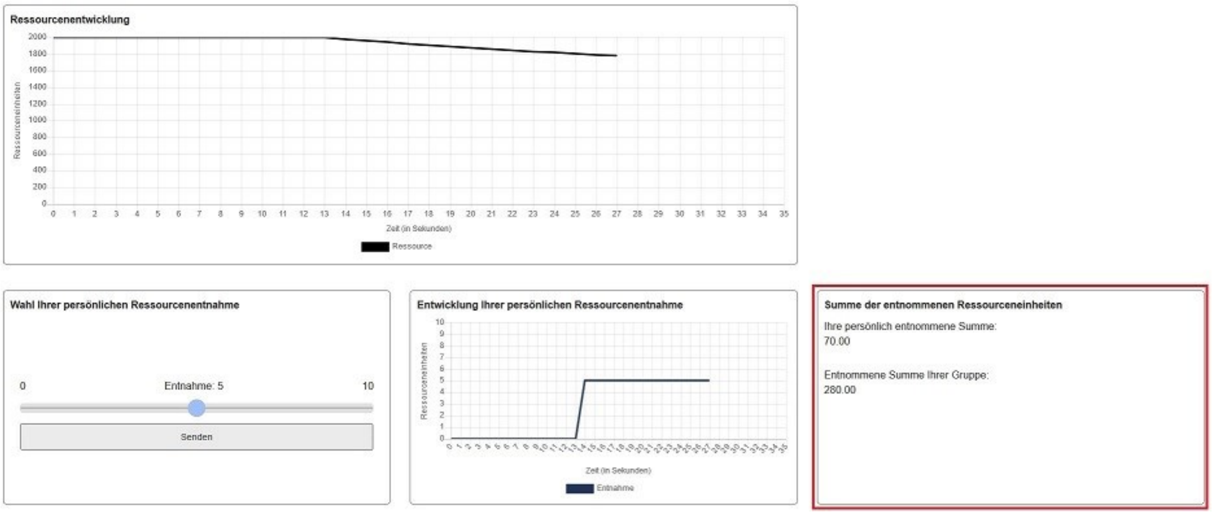


You get an overview about the sum of resource units that you yourself and all persons in your group have extracted in total up to a given point in time by looking at the information shown in the box **“Sum of extracted resource units”** (bottom right corner).

“Your own individual sum” (Ihre persönlich entnommene Summe) shows how many resource units have been extracted by you personally up to the given point in time. For example if you extract 10 resource units in the first and the second second, your individual sum will show 10 + 10 =20 resource units as total sum for yourself.

One of the two payout-relevant rounds will be randomly chosen for your payout at the end of the experiment. That payout is determined by the amounts of resource units stated under **“your individual sum”** and will be converted into Euro.

Information on “Sum of extracted resource units”: Extracted sum of your group



“Extracted sum of your group” shows how many resource units all four persons in your group have extracted in total up to a given point in time. The sum of the group includes the resource units that you extracted for yourself plus the sum of resource units that the other three persons in your group extracted up to the given point in time.

For example: If at 14 seconds **“Extracted sum of your group”** shows 280 resource unit and **“Your own individual sum”** shows 70 resource units, you know that up to that point in time, you extracted 70 resource units for yourself and the other three persons from your group extracted a total of 210 resource units. “Your own individual sum” and “Extracted sum of your group” will be updated at least every 5 seconds.

Instructions test rounds

The following part describes the procedure in the two test rounds in detail.

In the two test rounds, which are not relevant for your payout, the other three persons in your group will be simulated by the computer. The computer multiplies your own choice of resource extraction by 4, such that the other three simulated persons in your group choose the same level of resource extraction per second as you. For example, if you choose the level of resource extraction of 10, the group as a whole will extract:

4*10 resource units per second = 40 resource units per second.

The table below lists all possible extraction levels that can be chosen during the two test rounds. The column "Your choose" lists the choices that you can choose from by setting the extraction slider. The column on the right-hand side presents the level of extraction per second that is simulated by the computer according to your personal choice in the two test rounds.

You choose:	**The computer simulates the resource extraction of all four persons in your group as:**
0	0
1	4
2	8
3	12
4	16
5	20
6	24
7	28
8	32
9	36
10	40

By multiplying your own resource extraction choice by four, you get the chance to experience the development of the resource depending on different levels of resource extraction per second in a group of four persons. However, in the test rounds these other three persons are simulated by the computer such that they mirror your own choice of resource extraction per second.

Length of the two test rounds

Each test round lasts at least 80 seconds with certainty, which means, you have a minimum of 80 seconds to extract resource units and thereby to collect points for yourself.

Once the minimum time of 80 seconds is up each test round will continue for an unknown amount of time. **Hence, the exact end of the round and duration of the round is unknown at the beginning.**

Every 10 seconds it is randomly determined whether the round continues for another 10 seconds or not. There is a 90 percent probability that the round will continue for at least another 10 seconds during which you can collect further points for your payout. The round will end with a probability of 10 percent. Hence, in 9 of 10 cases the round will continue for another 10 seconds. Hence, at any decision point after the certain minimum duration of 80 seconds, the probability that the round will continue is nine times higher than the probability that the round ends.

For example, if you are at 80 seconds, the probability that there will be at least another 10 seconds is 90 percent and if you are at 100 seconds, the probability that there will be another 10 seconds is also 90 percent.

Control Questions - Part 1

Please answer a few questions to clarify your understanding of the instructions before you the start of the two test rounds.

Control Questions - Part 1 *[CQ1]*

Please answer the following question.

At which resource level is the growth of the resource at its maximum?

Answer: Number of resource units:

Correct answer: 1,000 resource units

Control Questions - Part 1 *[CQ2]*

Please choose the correct statement.

If you choose 10 as the level of resource extraction,

a) You extract 10 resource units per second. Your choice continuously implemented each second until you actively change it.

b) You once extract 10 resource units. You have to confirm your choice each second.

Correct answer: a

Control Questions - Part 1 *[CQ3]*

Please answer the following question.

What happens when the threshold of 500 resource units is reached (resource level is equal to or smaller than 500)?

a) The growth of the resource drops to 0, but the resource level remains at 500 resource units.

b) The resource growth drops to 0 and the resource level drops down to 0. No one can extract anymore resource units until the round is over.

Correct answer: b

Control Questions - Part 1 *[CQ4]*

Please choose the correct statement.

If you choose the level of 10 as personal resource extraction in one of the test rounds,

a) the computer simulates the other three persons in your group such that all four persons from one group choose an individual resource extraction of 10 resource units per second. The total resource extraction of the group is 4*10=40 resource units per second.

b) the computer simulates the other three persons in your group such that all four persons choose a combined resource extraction of 10 resource units per second. The total group extraction is then 10 resource units per second.

Correct answer: a

Start of the first test round

Summary:

- Your decisions in the two test rounds have no impact on your payout at the end of the experiment.
- During the two test rounds, the computer multiplies your chosen extraction level by four such that the other three members of your group are simulated.
- The exact length of each test round is unknown at the beginning. You have a minimum of 80 seconds with certainty to extract resource units and collect points.

Please click on “Continue” to choose the starting level of your personal extraction for the first test round. As soon as you make your choice and confirm it by clicking on “send”, the resource development of the first test round will start at the level of 2,000 resource units *(800 resource units in treatment T.3 Low resource accordingly)* under consideration of the chosen extraction level of your group.

Please choose your personal extraction level:

The first individual test round started automatically, once subjects chose their initial resource extraction level.

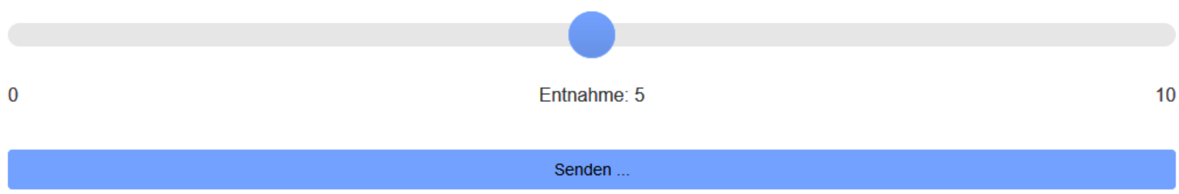
Start of the second Test Round

The first test round has ended and you will now start with the second test round.

Please remember that your decisions in the two test rounds have no impact on your payout at the end of the experiment. The points that you are collecting are not relevant for your payout.

Please click on “Continue” to choose the starting level of your personal extraction for the second test round. As soon as you make your choice and confirm it by clicking on “send”, the resource development of the second test round will start at the level of 2,000 resource units *(800 resource units in treatment T.3 Low resource accordingly)* under consideration of the chosen extraction level of your group.

Please choose your personal extraction level:



The second individual test round started automatically, once the subjects chose their initial resource extraction level.

Instructions payout-relevant rounds

You finished the two test rounds and now proceed to the instructions of the two payout-relevant rounds.

Before the start of the two rounds, that are relevant for your payout, all participants, who take part in the current experiment and are present in this room, are randomly matched in groups of four. Therefore, you will extract resource units from a resource that is shared with three other persons in the two payout-relevant rounds.

The three participants in your group of the first payout-relevant round will be changed for the second payout-relevant round. Hence, you will only share a group with any other person once.

Instructions payout-relevant rounds

None of the decisions in the two payout-relevant rounds are simulated by the computer.

Participants of this experiment make all the decisions on resource extraction. Each person in your group, including yourself, makes their decision regarding their personal level of resource extraction per second independently and anonymously.

Each person in your group can choose any of the given levels of resource extraction per second from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 to 10 resource units per second. Such that the lowest possible resource extraction per second for all four persons in your group together is 0 resource units per second. And the highest possible resource extraction of all four persons together is 40 resource units per second.

The possible levels of resource extraction for a group of four are all the different combinations of individual resource levels between 0 and 10. Hence, all the integer numbers between 0, 1, 2, 3, 4, … up to …, 36, 37, 38, 39, 40 are possible as joint group resource extraction levels per second at any given point in the experiment round. You and the other three individuals in your group can use the slider to change the level of the resource extraction per second anytime during the two payout-relevant rounds.

Duration of the payout-relevant round

The two payout-relevant rounds are longer than the two test rounds.

In each payout-relevant round, you have a minimum of **240 seconds** to extract resource units and thereby to collect points for yourself. After these 240 seconds, the round continues for an unknown period of time which means that every 10 seconds it is randomly determined whether the round continues for another 10 seconds.

Again, after each block of 10 seconds, there is a 90 percent probability that the round will continue for at least another 10 seconds. The round will end with a probability of 10 percent. Hence, in 9 of 10 cases the round will continue for another 10 seconds. At any point after the certain minimum duration of 80 seconds, the probability that the round will continue is nine times higher than the probability that the round ends.

For example, if you are at 240 seconds, the probability that there will be at least another 10 seconds is 90 percent and if you are at 300 seconds, the probability that there will be another 10 seconds is also 90 percent.

Control questions – payout-relevant rounds

Before you start the two rounds that are relevant for your payout, we ask you to answer a few more questions that help you to understand your task in the two payout-relevant rounds.

Control questions – Part 1 *[CQ5]*

Please answer the following question.

Under the assumption that person A and person B are in the same group in the first of the two payout-relevant rounds. How likely is it that person A and person B are also in the same group in the second of the two payout-relevant rounds?

a) Highly likely, person A and person B are definitely playing in the same group in both payout-relevant rounds

b) Possible, person A and person B could be in the same group in both payout-relevant rounds

c) Impossible, if person A and person B were in the same group in the first payout-relevant round, they are definitely not in the same group in the second payout-relevant round.

Correct answer: c

Control questions – Part 1 *[CQ6]*

Please answer the following question.

How high is the sum of the resource extraction of all persons in your group in the payout-relevant rounds, if you personally would choose a resource extraction of 5 resource units per second?

a) I cannot know that since the other three persons in my group can choose their level of resource extraction per second independent from my own personal choice.

b) The sum of the level of resource extraction per second of all four persons in the group is 4*5=20 resource units per second.

Correct answer: a

Control questions – Part 1 *[CQ7]*

Please answer the following question.

How long does each of the two payout-relevant rounds last?

a) The rounds lasts certainly 240 seconds *(215 seconds in T.3 Low resource)* and after that might randomly continue every 10 seconds with a probability of 90 percent.

b) The rounds stop immediately after 240 seconds. *(215 seconds in T.3 Low resource)*

Correct answer: a

Control questions – Part 1 *[CQ8]*

Please answer the following question.

When does the level of the resource drop to 0 resource units?

Please state the level of the resource at which the resource drops down to 0 resource units.

Correct answer: 500

Control questions – Part 1 *[CQ9]*

Please answer the following question.

Which possible level of resource extraction can all four persons in one group reach when their individual levels are summed up?

a) The possible levels of resource extraction per second of the group are all integer numbers between 0, 1, 2, 3, 4, 5, … to …, 38, 39, 40 due to different combinations of the individually chosen levels of resource extraction per second by each person in the group.

b) The possible levels of resource extraction per second of the group are combinations in which all four persons in the group choose the same level of resource extraction per second. Hence, 0, 4, 8, 12, 16, 20, 24, 28, 32, 36 or 40 resource units per second.

Correct answer: a

Control questions – Part 1 *[CQ10]*

Please answer the following question.

How high is the growth of the resource at a level of 500 resource units?

Correct answer: 0

Control questions – Part 1 *[CQ11]*

Please answer the following question.

How many points do you collect per seconds once the level of the resource and the growth of the resource are at zero?

a) As many points, as I chose with positioning the extraction slider.

b) I cannot collect any more points.

Correct answer: b

Instructions payout-relevant rounds - Summary

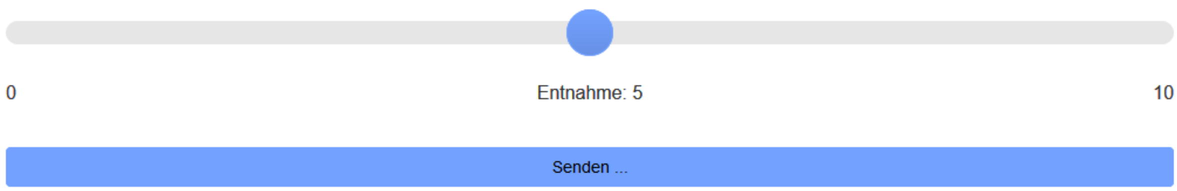
- You choose your personal level of resource extraction per second between 0 and 10 with the slider in the box at the bottom left.
- If the resource level reaches the threshold of 500 resource units, it will drop to the level of 0 resource units and the growth of the resource also drops to 0 resource units per second. Once this happens no more points can be collected.
- You have a minimum of 240 seconds per payout-relevant round to extract resource units and collect points for yourself. After these first 240 seconds each round will continue for an unknown amount of seconds. The exact end of the round is unknown.

- Before the start of each payout-relevant round, all participants will be randomly matched in groups of four persons. You cannot influence the choice of the extraction level of the other persons in your group directly. Every person in your group, including yourself, makes their extraction choices independently.

One of the payout-relevant rounds will be randomly chosen at the end of the experiment. The points that you collected based on your decisions in that randomly chosen round will determine your payout for the first part of the experiment and will be given to you in cash in Euro after the experiment. There are no right or wrong decisions.

Please click on “Continue” to choose the starting level of your personal extraction for the first of the two payout-relevant rounds. As soon as all persons, including yourself, in your group make their choices and confirm them by clicking on “send”, the resource development of the first payout-relevant round will start at the level of 2,000 resource units *(800 resource units in treatment T.3 Low resource accordingly)* under consideration of the chosen extraction level of your group.

Please choose your personal extraction level:



The first payout-relevant round started automatically, once all group members chose their initial resource extraction level.

Second Payout-Relevant Round

The fist payout-relevant round is finished and the second payout-relevant round will follow now. Before the start of the second round all participants of the experiment are again matched in new groups of four individuals.

[Additional in treatment T.3 Low resource:

Please be aware that the minimum round length of the second payout-relevant round is longer than it was in the first. Your now have a minimum of 240 seconds to extract resource units and collect points for yourself. Once these 240 seconds are up, the round continues for an unknown period of time which means that every 10 seconds it is randomly determined whether the round continues for another 10 seconds. There is a 90 percent probability that the round will continue for at least another 10 seconds and the round will end with a probability of 10 percent. The exact end of the round is unknown.*]*

Please click again on “Continue” to choose the starting level of your personal extraction for the first of the two payout-relevant rounds. As soon as all persons, including yourself, in your group make their choices and confirm them by clicking on “send”, the resource development of the first payout-relevant round will start.

The resource development starts at the level of 2,000 resource units *(in all three treatments)* under consideration of the chosen extraction level of your group.

Please choose your personal extraction level:

The second payout-relevant round started automatically, once all group members chose their initial resource extraction level.

The following text was presented in the pop-up window on screen in SoPHIE at the time of the pause (25 second) in CPR1 in treatment T1 Control and in all three treatments in CPR2:

Please read the following text carefully:

A sudden event caused a stop in the resource development. This one-time event occurred independent of the choices made by persons in your group and will not repeat itself during this round of the experiment. The properties of the resource like the growth of the resource per second and the impact that the extraction of resource units have on the resource have not changed.

While you are reading this information, the development of the resource is stopped, and no one can extract any resource units at the moment. The resource development will continue at its current resource level under consideration of your group's extraction choice, as soon as all persons in your groups decide on an extraction level by using the slider below and clicking on "send". You can change your chosen extraction level and the points that you are collecting per second at any time during the round by changing the slider in the box in the bottom left corner.

Please choose the extraction level with which you would like to continue your personal resource extraction at the current level of the resource:

The following text was presented in the pop-up window on screen in SoPHIE at the time of the pause (25 seconds) only in treatment T2 Exogenous Shock in CPR1:

Please read the following text carefully:

A sudden event caused a change in the environmental conditions of the resource and destroyed part of the resource of your group. This one-time event occurred independent of the choices made by persons in your group and will not repeat itself during this round of the experiment. Due to the caused destruction of resource units, the resource of your group drops down to 800 resource units. Other properties of the resource like the growth of the resource per second and the impact that the extraction of resource units have on the resource have not changed and continue to be like they were before the event.

While you are reading this information, the development of the resource is stopped, and no one can extract any resource units at the moment. The resource development will continue at its new resource level of 800 resource units under consideration of your group's extraction choice, as soon as all persons in your groups decide on an extraction level by using the slider below and clicking on “send”. You can change your chosen extraction level and the points that you are collecting per second at any time during the round by changing the slider in the box in the bottom left corner.

Please choose the extraction level with which you would like to continue your personal resource extraction at the new, lower resource level of 800 resource units:

Instructions of the SVO measure

Instructions Part 2

In part 2 of the experiment, you are asked to make 15 decisions about how to allocate a number of points between yourself and another person. In the following, this other person will simply be called “other”. This “other” is a person who is also a participant of the experiment here in the room. You have not interacted with the “other” before. Thus, you have not been in the same group during the first part of the experiment. You did not have any contact with “other” in the first part of the experiment.

The “other” is someone you do not know and will remain mutually anonymous. All of your choices are completely confidential. The “other” will not know who allocated points to them in this part of the experiment. **Please choose your preferred allocation of points for each of the following decisions.**

Your decisions are allocating points to yourself and to the “other”. Thereby you decide on the payout in Euro for yourself and the “other”. At the end of the second part of the experiment one of your 15 decisions will be randomly chosen by the computer and the points that you allocated to yourself in that randomly chosen decisions, will be credited towards your account. In addition, the person that was chosen as your “other” gets the points that you allocated to the “other” credited to their account.

As a general rule, the person who you send points to and the person who sends points to you, are two different persons.

Your payout for this second part of the experiment will be the sum of points that you allocate to yourself and that you receive from another person. The exchange rate for points to Euro for this part of experiment is **50 points = 1.50 Euro**.

The amount of points in the top line state the amount of points that would be allocated to your own account (Sie erhalten) and the points in the bottom line state the amount of points that the other person would get. You choose one of the options per decision by clicking on the dot in the middle.

Sie erhalten	100	96	93	89	85	81	78	74	70
	○	○	○	○	○	○	○	○	○
Jemand erhält	50	56	63	69	75	81	88	94	100

Please click on the allocation that you want to choose in each decisions and confirm your choice by clicking on "continue". Once you clicked on "continue", you cannot change your choice for that specific decision anymore.

There are no right or wrong answers in this part of the experiment; it is all about personal preferences.

Each of the following 15 decisions (Entscheidung 1 to Entscheidung 15) is presented on a separate screen. Subjects choose their allocation by clicking on the corresponding dot and then confirm their choice by clicking on "Continue".

Teil 2

Bitte treffen Sie eine Auswahl für die folgende Entscheidung.

Entscheidung 1

Sie erhalten	85	85	85	85	85	85	85	85	85
	○	○	○	○	○	○	○	○	○
Jemand erhält	85	76	68	59	50	41	33	24	15

Entscheidung 2

Sie erhalten	85	87	89	91	93	94	96	98	100
	○	○	○	○	○	○	○	○	○
Jemand erhält	15	19	24	28	33	37	41	46	50

Entscheidung 3

Sie erhalten	50	54	59	63	68	72	76	81	85
	○	○	○	○	○	○	○	○	○
Jemand erhält	100	98	96	94	93	91	89	87	85

Entscheidung 4

Sie erhalten	50	54	59	63	68	72	76	81	85
	○	○	○	○	○	○	○	○	○
Jemand erhält	100	89	79	68	58	47	36	26	15

Entscheidung 5

Sie erhalten	100	94	88	81	75	69	63	56	50
	○	○	○	○	○	○	○	○	○
Jemand erhält	50	56	63	69	75	81	88	94	100

Entscheidung 6

Sie erhalten	100	98	96	94	93	91	89	87	85
	○	○	○	○	○	○	○	○	○
Jemand erhält	50	54	59	63	68	72	76	81	85

Entscheidung 7

Sie erhalten	100	96	93	89	85	81	78	74	70
	○	○	○	○	○	○	○	○	○
Jemand erhält	50	56	63	69	75	81	88	94	100

Entscheidung 8

Sie erhalten	90	91	93	94	95	96	98	99	100
	○	○	○	○	○	○	○	○	○
Jemand erhält	100	99	98	96	95	94	93	91	90

Entscheidung 9

Sie erhalten	100	94	88	81	75	69	63	56	50
	○	○	○	○	○	○	○	○	○
Jemand erhält	70	74	78	81	85	89	93	96	100

Entscheidung 10

Sie erhalten	100	99	98	96	95	94	93	91	90
	○	○	○	○	○	○	○	○	○
Jemand erhält	70	74	78	81	85	89	93	96	100

Entscheidung 11

Sie erhalten	70	74	78	81	85	89	93	96	100
	○	○	○	○	○	○	○	○	○
Jemand erhält	100	96	93	89	85	81	78	74	70

Entscheidung 12

Sie erhalten	50	56	63	69	75	81	88	94	100
	○	○	○	○	○	○	○	○	○
Jemand erhält	100	99	98	96	95	94	93	91	90

Entscheidung 13

Sie erhalten	50	56	63	69	75	81	88	94	100
	○	○	○	○	○	○	○	○	○
Jemand erhält	100	94	88	81	75	69	63	56	50

Entscheidung 14

Sie erhalten	100	96	93	89	85	81	78	74	70
	○	○	○	○	○	○	○	○	○
Jemand erhält	90	91	93	94	95	96	98	99	100

Entscheidung 15

Sie erhalten	90	91	93	94	95	96	98	99	100
	○	○	○	○	○	○	○	○	○
Jemand erhält	100	94	88	81	75	69	63	56	50

Post-experimental questionnaire

Questionnaire

To finalise the experiment, we kindly ask you to answer a few questions. In the meantime, our team is preparing your payout. Your answers to the following questions have no effect on your payout.

It is mandatory to answer all questions marked with an asterisk *.

All of your answers are anonymised and treated with confidentiality.

[S1]

Please answer the following question.

How do you see yourself: are you generally a person who is prepared to take risks, or do you try to avoid taking risks?

Please tick a box on the scale, where the value 0 means: ‘not at all willing to take risks’ and the value 10 means: ‘very willing to take risks’.

You can use the values in between to grade your assessment.

not at all willing to take risks										very willing to take risks
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
○	○	○	○	○	○	○	○	○	○	○

[NEP 1-15]

(German translation of the NEP scale as published in Schleyer-Lindenmann et al, 2018)

Please answer the following questions.

Listed below are statements about the relationship between humans and the environment. For each one, please indicate to which extent you agree with it.

Please tick a box on the scale from 1 to 5 to indicate to what extent you reject or agree with the following statements. The value 1 means: "strongly reject" and the value 5 means: "strongly approve". You can use the values in between to grade your rejection or approval respectively.

There are no right or wrong answers. Please decide based upon your own personal opinion.

1. We are approaching the maximum population of humans the earth can support.
2. Humans have the right to modify the natural environment to suit their needs.
3. When humans interfere with nature it often has disastrous consequences.
4. Human ingenuity will insure that we do NOT make the earth unlivable.
5. Humans are severely abusing the environment.
6. The earth has plenty of natural resources if we just learn how to develop them.
7. Plants and animals have as much right as humans to exist.
8. The balance of nature is strong enough to cope with impacts of modern industrial nations.
9. Despite our advanced state, humans are still subject to the laws of nature.
10. The so-called "ecological crisis" facing humankind has been greatly exaggerated.
11. The earth is like a spaceship with very limited room and resources.
12. Humans were meant to rule over the rest of nature.

13. The balance of nature is very delicate and easily upset.

14. Humans will eventually learn enough about how nature works to be able to control it.

15. If things continue on their present course, we will soon experience a major ecological catastrophe.

[S2]

What were your expectations regarding the length of the **payout-relevant** rounds?

Did you expect that after the first 240 seconds the rounds

a) would end immediately.

b) would end after less than 30 seconds.

c) would end after another 30 seconds.

d) would end after 30 to 60 more seconds.

e) would end after 60 to 90 seconds.

f) would end after more than 90 seconds.

g) You had no expectation regarding the end of the rounds.

[S3]

Assume that you knew exactly when the **payout-relevant rounds** would have ended. Would you have chosen a different level of personal resource extraction?

y: Yes, I would have increased my personal resource extraction before the end of the payout-relevant rounds.

n: No, I would not have increased my personal resource extraction before end of the payout-relevant rounds.

m: Neither, I would have behaved differently than described under “yes” and “no”.

(only if S3 was answered with y)

[S3a]

To which level would you have increased your resource extraction per second?

(choice of integer numbers between 0 and 10)

[S3b]

How many seconds before the end of the payout-relevant rounds would you have increased your resource extraction per second? *(open text field)*

[S4]

Please answer the following question.

Do you think it is possible to understand the experiment without the two test rounds?

- Yes
- No

[S4a]

To what extent did the two test rounds help you to understand the development of the resource and your task during the first part of the experiment? *(open text field)*

[S5]

Please tell us how you perceived the differences in the experimental procedure of between the two test rounds and the two payout-relevant rounds.

a) I think that the differences were explained clearly in the instructions.

b) I did not understand the differences between the test rounds and the payout-relevant rounds.

c) It was not clear to me that there were differences between the test rounds and the two payout-relevant rounds.

[S6]

In how far did the test-rounds and the payout-relevant rounds differ? To what extent was the procedure comparable?

Please click on all statements that you think apply (multiple answers possible):

a) The payout-relevant rounds lasted at least 240 seconds *(215 in T.3 Low Resource)*.

b) The payout-relevant rounds lasted maximum 240 seconds *(215 in T.3 Low Resource)*.

c) The probability with which the round could end every 10 seconds was known in the payout relevant rounds.

d) The test rounds and payout-relevant rounds ended randomly. The exact end was unknown at the beginning.

e) During the test rounds, the computer simulated the other three people in my group.

f) In the payout-relevant rounds, the computer simulated the other three people in my group.

g) In the payout-relevant rounds, the three other people in my group were real people who also participated in the experiment.

[S7]

Were you able to directly influence the level of resource extraction of the other people in your group during the payout-relevant rounds?

a) Yes, during the payout-relevant rounds, the other people always automatically selected the same level of resource extraction as I did.

b) No, in the payout-relevant rounds, the other people in my group freely chose their level of resource extraction independent of my personal choice.

[S8]

We would like to know how you perceived the explanation of the interrupted resource development in the payout-relevant rounds. By interruption, we mean the sudden event, which happened after approx. 25 seconds, the appearance of the explanation pop-up window and the renewed request to choose an extraction level.

Did you understand the interruption in resource development as part of the experiment?

- Yes
- No

[S8a]

Please briefly explain how you interpreted the sudden event that caused the interruption of the resource development: *(open text field)*

Question was only present in Treatment T.2 Exogenous shock: **[S8shock]**

Please briefly describe how you perceived the interruption after 25 seconds and the information about the sudden loss of resource units in the first payment-relevant round.

Did the information about the sudden loss motivate you to decrease or increase your personal resource extraction?

- Decrease, I chose a lower extraction level.
- Increase, I chose a higher extraction level.
- Neither, I have chosen the same extraction level as before.

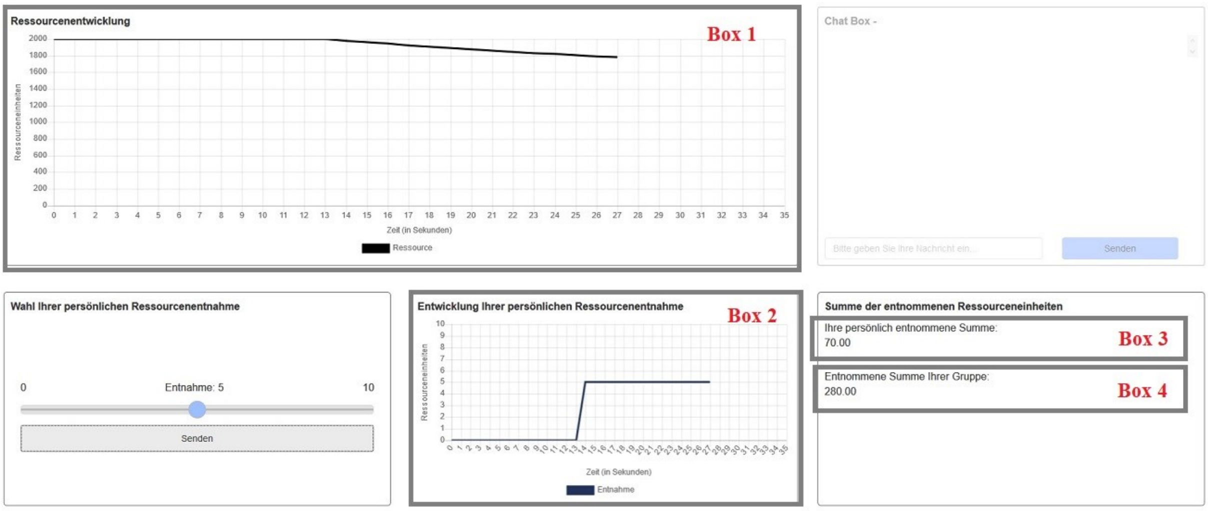
Question was only present in Treatment T.2 Exogenous shock: **[S8shocka]**

Please explain briefly, why you reacted like that. *(open text field)*

[S9]

What information did you focus on during the payout-relevant rounds?

Did you focus on the graph that showed the development of the resource (Box 1), on the graph that showed the development of your personal extraction (Box 2), on the display of your individual sum of extracted resource units (Box 3) or on the display of your groups' sum of the resource extraction (Box 4)?



Please indicate which information you mostly focused on during the payout-relevant rounds. You can choose multiple answers:

I focused on the

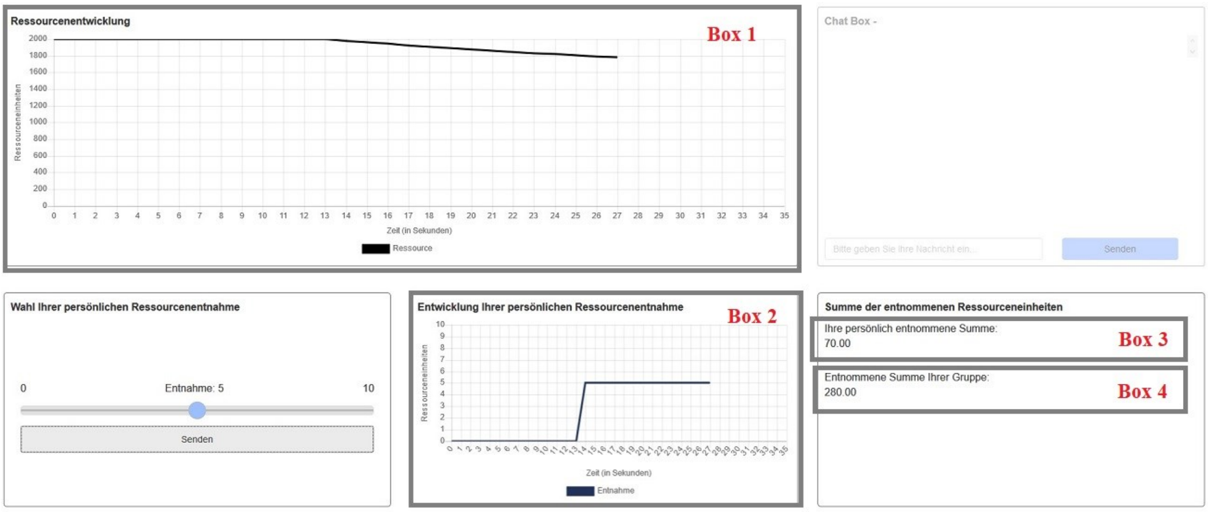
- graph that showed the development of the resource over time (Box 1).
- graph that shows the development of my personal resource extraction over time (Box 2).
- display of my personally extracted sum of extracted resource units (Box 3).
- display of my group's sum of the resource extraction (Box 4).

[S9a]

Please describe briefly, why you focused on the selected information. (open text field)

[S10]

What information did you not pay attention to during the payout-relevant rounds?



Please indicate which information you mostly focused on during the payout-relevant rounds. You can choose multiple answers:

I did not pay attention to the

- graph that showed the development of the resource over time (Box 1).
- graph that shows the development of my personal resource extraction over time (Box 2).
- display of my personally extracted sum of resource units (Box 3).
- display of the sum of the resource units extracted from my group (Box 4).

[S10a]

Please briefly describe why you did not pay attention to the selected information. *(open text field)*

[S11 a - e]

To which extend do you agree with the following statements?

Please tick a box on the scale from 1 to 5. The value 1 means: “I strongly disagree” and the value 5 means: “I strongly agree”. You can use the values in between to grade your approval.

- The instructions were written understandably.
- I fully understood the development of the resource.
- It was important for me to prevent the resource from reaching the threshold.
- It was easy to adjust my personal resource extraction with the slider during a round.
- It was important for me to allocate the resource fairly among all people in a group.

[S12]

Were there moments during the experiment when you felt stressed?

- Yes
- No

[S12a]

Please briefly explain when and why you felt stressed. *(open text field)*

[S13]

Do you have any <u>further</u> comments on the experiment?

You can also explain here why you made certain decisions during the experiment. *(open text field)*

[CV1]

Which gender are you?

- Female
- Male

[CV2]

How old are you?

- (open number field) age in years

[CV3]

Have you participated in one or more economic or psychological experiments before this experiment?

- Yes
- No

(if yes to CV3) **[CV3a]**

How many times have you participated in economic or psychological experiments? (open number field)

[CV4]

Are you a student?

a) Student at Osnabrueck University

b) Student at Osnabrueck University of Applied Sciences

c) Student at another university/university of applied sciences

d) Not a student

[CV4a]

Which field of study are you studying (predominantly)?

a) Humanities

b) Social sciences (not economics)

c) Economics

d) Natural sciences

e) Engineering

f) Other

[CV4b]

What is the name of your study program? *(open text field)*

[CV4c]

What type of degree is it?

- Bachelor
- Master
- PhD

[CV5]

What is your monthly income (including subsidies from your parents, student grants, salaries, scholarships)?

- 0-300 Euro
- 301-600 Euro

- 601-900 Euro
- More than 900 Euro
- No answer

Payout information were presented to subjects once they finished the questionnaire:

Your personal payout for both parts of the experiment

Part 1 – payout-relevant round [NUMBER] was chosen at random

Your personal extraction in payout-relevant round [NUMBER]	X Points
Converted into EURO	…€

Part 2 – Your **decision [NUMBER]** was chosen at random

Number of Points, you chose for yourself in **decision [NUMBER]**	Y Points
Points, that you received from someone else	Z Points
Converted into EURO	…€
Plus the flat payment	3.00€
Total Payout	…€

Thank you very much for participating in this experiment.

Please remain seated and wait while we are preparing your payouts. A member of our team will tell you, once you can leave the room to collect your payouts.

Chapter 3 Are imprecise early warnings a potential benefit or threat to sustainable resource management?

Katharina Hembach-Stunden[a], Tobias Vorlaufer[a], Stefanie Engel[a]

[a] School of Business Administration and Economics and Institute for Environmental Systems Research (IUSF), Osnabrueck University, Germany.

Abstract: Overexploitation can change the underlying conditions of ecosystems, causing drastic shifts to unfavourable states once ecosystems reach critical thresholds. These so-called regime shifts often have far-reaching economic consequences for resource users. Experimental literature has shown that the common knowledge of a threshold helps to foster cooperation and to overcome social dilemmas. However, warning resource users of an imminent regime shift is difficult since the specificities of critical thresholds of ecosystems are often unknown. Consequently resource users can only be given relatively imprecise threshold ranges as early warnings. The effect of such imprecise early threshold warnings on cooperation amongst resource users is unclear. On the one hand, imprecise threshold knowledge could support cautious resource extraction due to the possibility of causing a regime shift through overexploitation. On the other hand, imprecise knowledge could lead to aggressive overexploitation due to a "use-it-or-lose-it"-mentality. Furthermore, receiving an imprecise early warning about a threshold could potentially distort the otherwise coordinating effect of certain threshold knowledge. Here, we assess whether an imprecise threshold warning increases or decreases cooperation amongst resource users and analyse if the effect of imprecise warnings persists after the certain threshold has been communicated. To do this, we designed a (quasi-) continuous-time common-pool resource experiment for the laboratory with two treatments that differ in the degree of threshold uncertainty at the beginning. We find that groups who merely know of the threshold's existence have no different cooperation than groups who have received an imprecise early warning in the form of a broad threshold range. Furthermore, an imprecise early warning does not affect coordination after the certain threshold level is communicated. Our results suggest that the scope of imprecise early warning signals to foster more sustainable natural resource management is limited.

Keywords: decisions under uncertainty, environmental uncertainty, resource user behaviour, CPR, resource management, lab experiment

3.1 Introduction

Overexploitation of resources can change the underlying conditions of ecosystems. As a consequence, ecosystems can drastically switch to an alternative state once they are driven to a critical threshold, which is referred to as a regime shift (Scheffer et al. 2001). Regime shifts often have drastic negative impacts on economies and societies (Biggs et al. 2009). Examples include desertification of woodlands and the collapse of fisheries (Millenium Ecosystem Assessment 2005). Communicating the possibility of critical thresholds is seen as a valuable tool to support sustainable resource use and to overcome both cooperation and coordination problems in resource management (Maas et al. 2017). However, due to scientific uncertainty about critical thresholds, the onset of a regime shift is often identified too late to react because of limited timeframes for efficient policy interventions (Biggs et al. 2009, Crépin et al. 2012).

In recent years, there has been a debate about the potential use of ecological early warning signals to inform resource users about approaching critical thresholds and impending regime shifts (Biggs et al. 2009, Boettiger et al. 2013). Yet, it remains difficult to determine the exact levels of such critical thresholds (ibid.). One example is deforestation of the Amazon rainforest that is predicted to cause a regime shift from rainforest to savannah, reducing rainfall and increasing temperatures in the area (Lovejoy and Nobre 2019). Some initial signals that the Amazonian regions are approaching a critical threshold have already been observed (Lovejoy and Nobre 2019), but, estimates of the critical deforestation threshold for the Amazonia remain imprecise, ranging from 20% to 40% deforestation (Lenton et al. 2019).

Hence, in practice, resource users likely understand that a critical threshold exists, but seldom know the exact level or probabilities of alternative levels. Additionally, external changes like climate change may affect thresholds over time, and thus, resource users likely face a very high degree of uncertainty about the threshold level, i.e. threshold ambiguity. Threshold ambiguity means that the underlying probability distribution of the threshold level is unknown (Aflaki 2013).

However, it is unclear how resource users respond to an early, but imprecise threshold warning that reduces the degree of ambiguity about the critical level. Theoretical work suggests two possible pathways. On the one hand, the imprecise threshold knowledge could motivate coordination for a cautious extraction strategy because resource users could become more aware that experimenting with high extraction incorporates the risk to reach the critical threshold (Diekert 2017, Bochet et al. 2019). On the other hand, an imprecise threshold warning providing more knowledge about the imminent risk of the regime shift could amplify overexploitation due to a “use-it-or-lose-it”-mentality (Crépin et al. 2012). As policy makers and researchers must decide if and when to pass on vague findings about critical thresholds to the public, a thorough understanding of the effects of imprecise threshold knowledge on resource management is needed.

This paper aims to address the lack of empirical evidence on this topic. We designed a novel (quasi-) continuous-time common-pool resource (CPR) experiment for the lab to analyse if an imprecise warning, in the form of threshold range knowledge, increases or decreases the level of cooperation amongst resource users. For this, we compare the case of threshold range knowledge to the case with general threshold ambiguity where only the presence of a threshold is known without any information about its level. Further, we assess if learning about the exact level of the threshold affects coordination on resource levels above the critical threshold differently depending on the degree of prior threshold ambiguity. To our knowledge we are the first to analyse the effect of different degrees of threshold ambiguity and the compound effect of threshold ambiguity followed by threshold certainty on resource management. In line with previous experimental studies on threshold uncertainty (Barrett and Dannenberg 2014a, Maas et al. 2017), we focus on uncertainty (ambiguity) of the threshold’s level. Uncertainty about the threshold’s level has been shown to be more relevant for cooperation and coordination amongst group members than uncertainty about the impact of the regime shift (Barrett and Dannenberg 2012) or uncertainty about the presence of a threshold (Schill et al. 2015, Schill and Rocha 2019, Rocha et al. 2020).

Previous studies have focused on different types of environmental uncertainty, without yielding a uniform picture how such uncertainty affects resource use. The literature focuses either on uncertainty about the resource stock size or the comparison of certainty about the threshold level to different degrees of threshold uncertainty. On the one hand, theoretical work from Aflaki (2013) outlines that ambiguous information on the resource size is likely to decrease resource extraction in a CPR in comparison to a scenario with less uncertainty. Individuals who are facing ambiguity are potentially more pessimistic about the resource size, and thus extract the resource more cautiously (ibid.). On the other hand, uncertainty about the available resource stock has been found to undermine the social norm to cooperate (Hine and Gifford 1996) compared to the case of certain knowledge, and to increase resource extraction (Budescu et al. 1990, 1992, Rapoport et al. 1992, Hine and Gifford 1996, Gustafsson et al. 1999). A critical threshold that triggers a collapse of the resource limits the available resource stock for resource users as well. If confronted with an uncertain threshold, resource users do not know the exact resource stock they can exploit without triggering a catastrophic regime shift. In threshold public good and common-pool resource games, individuals fail to coordinate on expected thresholds if the exact level is unknown. Contrastingly, a certain threshold motivates a majority of groups to coordinate successfully and avoid overexploitation (Barrett and Dannenberg 2012, 2014a, 2014b, Brown and Kroll 2017, Maas et al. 2017). None of these studies have compared the impact of different degrees of ambiguity about the threshold level, a gap we address in this paper.

Overall, our experimental results further the understanding of imprecise early warnings as policy instruments on resource management. We find that an imprecise early warning about the threshold in form of a known threshold range does not influence cooperation in comparison to a scenario without an early warning. Furthermore, we do not find evidence that an imprecise early threshold warning significantly affects resource extraction after the exact level of the threshold is known with certainty. Our results hence suggest that the scope of imprecise early warning signals to foster more sustainable natural resource management is limited.

3.2 Experimental setting

3.2.1 Experimental design

Our (quasi) continuous-time common-pool resource (CPR) game is a computerised lab experiment programmed in the experimental software SoPHIE (Hendriks 2012). The experimental design is based on the set-up presented in Chapter 2 and is in line with previous experiments that allow for the continuous-time nature of resource management (Janssen 2010, Brandt et al. 2017, Cerutti and Schlüter 2019). During the CPR game, participants in groups of four manage a joint resource, which develops continuously over time. Participants extract resource units to generate a payout for themselves. The resource and users' extraction development is updated by the second, which defines it a (quasi) continuous time game (Bigoni et al. 2015). Please see Appendix 3B for details of the implementation of the experiment.

User-resource model

The underlying resource dynamics of the game were as follows. Equation (3.1) below shows how the resource R_t changes over time t (in seconds). While the resource is above the critical threshold ($R_t > R_{min}$), the resource's natural growth changes with the resource level based on a simple logistic growth model (e.g. Perman et al. 2011, Brandt et al. 2017). The logistic growth term describes the resource's natural growth with the growth rate $g = 0.04$. Each round the resource development starts at the resource's maximum carrying capacity (MCC) $R_{max} = 2{,}000$ units. If the resource is at its MCC, the regrowth is zero. Below the MCC, the regrowth per second increases until the resource reaches the maximum sustainable yield (MSY) at a resource level of 1,000 units (regrowth of 20 units per second). Below the MSY, the regrowth per second decreases again. Once the resource reaches the critical threshold $R_{min} = 400$ units, it irreversibly collapses ($R_{t+1} = 0 \ if \ R_t \leq R_{min}$). At the same point, the resource's regrowth drops down to zero and stays at zero infinitely ($gR_t\left(1 - \frac{R_t}{R_{max}}\right) = 0 \ if \ R_t \leq R_{min}$).

Thus:

$$R_{t+1} = \begin{cases} R_t + gR_t\left(1 - \frac{R_t}{R_{max}}\right) - \sum_{i=1}^{n} E_{it} & if\ R_t > R_{min} \\ 0 & if\ R_t \leq R_{min} \end{cases} \tag{3.1}$$

where $\sum_{i=1}^{n} E_{it}$ denotes a group's joint extraction with $n = 4$ resource users per second t.

Participants' payout P is determined in points during the experiment (1 resource unit = 1 point) and later converted into Euro with an exchange rate of 100 points = 1.00 Euro. The payout P (in points) for participant i is determined by the sum of resource units that i extracts over time t, with t_{end} seconds denoting the last second of the round, which is unknown to participants:

$$P_i = \sum_{t=0}^{t_{end}} E_{it} \tag{3.2}$$

Groups that cause a collapse of the resource must wait until the end of the round without collecting any more points for their payout, thereby participants had no incentive to cause an early collapse to finish the experiment quickly (see Appendix 3A for further details of the model).

We decided against showing participants the growth function and instead explained the resource dynamics verbally in the instructions (Appendix 3H). Inspired by the instructions of Schill et al. (2015), we presented the regrowth levels corresponding to resource levels between 2,000 and 0 resource units in steps of 200 in a table to participants. Due to our research questions, participants initially only knew that reaching the critical threshold results in a collapse of the resource, but they did not know the exact level of the threshold.

Treatment design

The two treatments differed in participants' knowledge about the threshold at the beginning of the payout-relevant round. Both treatments incorporated the threat of crossing a critical threshold of the resource through overexploitation causing an immediate and irreversible shift to resource extinction. Participants in treatment “High uncertainty” (HU) only knew that there was a threshold, without any information on the resource level at which the regime shift occurred. In contrast, participants in treatment “Low uncertainty” (LU) received an imprecise early warning about the range around the possible threshold (700 to 200 resource units). The underlying probability distribution of the threshold level within this range was unknown to participants. Thus, participants in both treatments faced threshold uncertainty in form of threshold ambiguity. Please see the instructions in Appendix 3H for details.

In both treatments, the payout-relevant round consisted of a pre-pause and a post-pause part. During a previously unannounced pause in-between the two parts; participants were informed that the threshold level was now known to be with certainty at 400 units. Initially, it was unknown to participants that the certain threshold would be revealed during the game. The pause occurred after 58 seconds, which was late enough to drive the resource below MSY, but still early enough to prevent groups from reaching the threshold (see Appendix 3A for details of the parametrisation).

The game represents a social dilemma, where selfish and myopic individuals have an incentive to maximise their individual outcome by free riding and choosing the maximum individual extraction, while the social optimum differs from this. In real life, resource users face an infinite time horizon regarding their resource extraction. It is then socially optimal for the group of resource users to keep the resource at its MSY, because doing so indefinitely maximises the group’s outcome over time. However, inducing the socially optimal group extraction strategy in the lab is difficult because participants know that the experiment cannot go on for infinity and will end eventually. This is found to decrease cooperation towards the end of experiments and is known as endgame effect (Andreoni 1988). We implemented two strategies to reduce such endgame effects in our experiment.

Firstly, as in Chapter 2, we implemented a random continuation rule to induce an infinite time horizon in the lab (Dal Bó and Fréchette 2018). Beyond the certain minimum round length of 240 seconds, the payout-relevant round continued for another ten seconds with a probability of 90% in both treatments. With a 10% probability, the round ended. Every ten seconds, the random continuation rule was executed and as long as the round continued, participants could collect points for their payout. Secondly, we defined 210 seconds as the “end” of the round for our analysis to avoid any influence from potential endgame effects on our results (see Appendix 3B for details).

It should be noted that the high initial degree of threshold uncertainty in HU meant that groups could not be certain that the threshold is below the MSY. Thus, driving the resource quickly to MSY incorporated the risk to cause a collapse. Consequentially, groups in HU had a higher incentive to keep the resource above the MSY until they received the information about the certain threshold level. In contrast, groups in LU knew the threshold range (700 to 200 units) from the beginning. Therefore, they knew with certainty that it was safe to drive the resource to the MSY as quickly as possible.

Participants played two test rounds of 90 seconds in groups of four before the payout-relevant round. The test rounds allowed participants to familiarise themselves with the resource dynamics and the mechanism of resource extraction. In contrast to the payout-relevant round, the test rounds had a certain end and did not incorporate a critical threshold. In the test and payout-relevant rounds, participants chose their extraction level between 1 to 10 units (integer numbers) per second via a slider. Once chosen, the set extraction level was executed every second until the participant changed it or the round ended.

We randomised the two treatments on the session level. For the payout-relevant round, two members of each test round group were assigned to HU and two to LU. Furthermore, we implemented a perfect strangers’ matching such that participants who were group members in the test rounds did not interact with each other again in the payout-relevant round. All participants knew this. Thereby, we reduced social learning effects, and avoided strategic interactions and reputation building between test and payout-relevant rounds (Andreoni and Croson 2008).

3.2.2 Experimental procedures

The experiment was implemented at the WISO Experimental Lab of Hamburg University between November 2019 and February 2020. For 17 sessions a total of 360 participants was recruited from the pre-registered subject pool with the organisational tool hroot (Bock et al. 2014). We had 180 participants, i.e. 45 groups per treatment. Each participant only participated once in one of the two treatments. Participants' payout depended on their own and their group members' extraction choices made during the payout-relevant CPR round. On average, participants earned 15 Euro (SD=5.5, Min=2, Max=35) and each session lasted about 75 minutes (see Appendix 3B for details of the procedures).

As shown in the balance table based on individual observations, we find no statistically significant differences in participants' socio-demographic characteristics and understanding of the experiment between the two treatments (Table 3C.1). Overall, participants understood the instructions and the CPR game well (see Appendix 3B for details). Participants' average age was 25.5 years (SD=4.8), the total fraction of female participants was 63% (SD=48), and individuals' average income was between 601 and 900 Euro (average category=3, SD=1). 99% (SD=7) of all participants were university students. The total fraction of economics students was 29% (SD=45) and 89% (SD=31) have had previous experience taking part in economic or psychological experiments. The average risk taking measured on a 11-point scale (Dohmen et al. 2011) was 5 (SD=2). However, individuals' average expectation of the round's continuation beyond the certain minimum round length is significantly different between the two treatments. Individuals in LU expect the round to continue longer than groups in HU (two-sided t-test: $p<0.01$, Table 3C.1). We do not expect this to affect our results because we control for such an endgame effect by defining 210 seconds as endpoint for the analysis.

3.2.3 Formulating hypotheses

Our analysis focuses on the effect of imprecise early warnings on sustainable resource management. We measure sustainable resource management with two group outcomes:

(1) overexploitation as proxy for the failure of *cooperation* amongst group members, and (2) the collapse of the resource as an indicator of failed *coordination* amongst group members.

As shown above, previous evidence of the impact of environmental uncertainty in form of uncertainty about the resource stock size and threshold level on cooperation is inconclusive. Thus, we hypothesise:

Hypothesis 2.1: Ambiguity about the threshold affects *cooperation* measured as overexploitation of the resource. Groups in the LU treatment are either (a) *more* likely to overexploit and have a *higher* degree of overexploitation or (b) *less* likely to overexploit and have a *lower* degree of overexploitation than groups in HU.

We define overexploitation as a resource level below the MSY at the time of the pause (58 seconds) as a proxy for the failure of cooperation. In principle, non-cooperation can involve keeping the resource above or below the MSY, as under-exploitation and overexploitation both imply that group payouts are not maximised. However, because participants in our HU treatment had no information on the exact level of the threshold before the pause, they cannot be sure that the critical threshold is below the MSY. Hence it is not straightforward for the participants to define a resource level that constitutes under-exploitation in HU. Given these structural differences between treatments, our main analysis focuses on cooperation failure by overexploitation, i.e. resource levels below the MSY. Because the graphic representation of the resource development in the experiment made it difficult for participants to judge the exact level of the resource within a range of 30 units, we define the MSY as 970 (instead of 1000).

Our binary variable for cooperation failure is thus equal to 1 if the resource is below 970 resource units at 58 seconds, and 0 otherwise. Additionally, we define the degree of overexploitation as a continuous variable, which measures the distance of the resource level to the MSY (970) at 58 seconds. The maximum degree of overexploitation at the pause is 370 if the resource is at the lowest possible level at the time (600 units). The minimum value for this variable is 0.

Furthermore, we examine whether treatment differences persist after the exact threshold level is revealed. The effect of certain threshold knowledge may differ depending on the level of prior threshold ambiguity due to individuals' path-dependency of decision making (Kay 2005, Heinmiller 2009). The differences between the two treatments in the ambiguity about the threshold until the time of the pause might cause different group experiences and resource dynamics that determine the effect of certain threshold knowledge on coordination after the pause. Next, we describe these potential effects.

On the one hand, if an imprecise warning (LU) undermines cooperation and causes higher degrees of overexploitation compared to HU (Hypothesis 2.1a), mutual trust in group members' willingness to cooperate and coordinate might be lost. Lower trust in the willingness to coordinate could reinforce a "use-it-or-lose-it"-mentality (Crépin et al. 2012) and participants likely anticipate that their group members will cause a collapse of the resource, leading the individuals to increase their own extraction and overexploit the resource intentionally to not miss out on their individual gain (Maas et al. 2017). Thus, less cooperation and higher degrees of overexploitation in LU would be expected to lead to a higher rate of coordination failure in LU than in HU. It is also plausible that higher degrees of overexploitation in LU and thus lower resource levels compared to HU at the time when groups receive the certain threshold warning could cause a shock effect. On the other hand, being comparably close to the threshold when getting to know its exact level might alarm participants and motivate them to reduce their extraction. Thus, from this perspective, groups in LU would be expected to be less likely to fail coordination.

Conversely, if an imprecise warning (LU) fosters cooperation and causes lower degrees of overexploitation compared to HU (Hypothesis 2.1b), mutual trust in group members' willingness to cooperate and coordinate could be strengthened. Participants in LU would be more confident that successful coordination within their group is achievable compared to participants in HU. As a consequence, we would expect to observe less coordination failure in LU than HU after the certain threshold is revealed. Furthermore, lower degrees of overexploitation in LU would result in a higher resource level at the time when the certain threshold is revealed. A higher resource level means a greater distance to the threshold, which could have one of two opposing effects. Receiving certain threshold knowledge while holding a relatively high resource level may indicate that there is sufficient scope to exploit the resource, which could lead to a higher rate of coordination failure in LU compared to HU. By contrast, receiving certain threshold knowledge while holding a relatively high resource level may provide sufficient time to learn to coordinate before getting into the proximity of the threshold enhancing chances of successful coordination in LU.

In sum, we cannot clearly predict the direction of our treatment effect on coordination. Thus, we formulate:

Hypothesis 2.2: Differences in prior threshold ambiguity affect *coordination*, i.e. the likelihood of the resource collapsing, once the critical threshold is revealed. Groups in LU are either (a) *more* or (b) *less* likely to cause a collapse of the resource than groups in HU.

We measure coordination failure as a binary variable “collapse of the resource”, which is defined to equal 1 if the resource level reached or fell below the threshold of 400 resource units and thus collapsed, and 0 otherwise. We assess this variable at three points in time: at our defined round end for analysis purposes (210 seconds) and at two control times (90 and 150 seconds).

3.3 Results

This study was preregistered at “AsPredicted.org” (Wharton Credibility Lab, 2017). We refrain from reporting the full results of the pre-registration (see Appendix 3F for the remaining results and Appendix 3I for the pre-registration document). Our data analysis focuses on the two main group outcomes *cooperation* and *coordination*. It is conducted in STATA 15 (StataCorp 2017) and based on non-parametric and parametric tests. To account for some of the variance in the outcomes, we run robustness checks of our regression models including groups’ mean age, fraction of females, mean risk measure and mean expected continuation of the round as controls (see Appendices 3D, 3E and 3G).

3.3.1 Resource and group extraction development over time

Figure 3.1 shows the time trends of the average resource development and extraction choices per treatment. The resource development over time is almost identical for the two treatments (Fig. 3.1a). However, there are small differences in groups' chosen extraction levels over time (Fig. 3.1b). Groups in HU seem to be more cautious at the start and have a slightly lower average extraction level than groups in LU (HU: Mean=34, SD=5 vs LU: Mean=36, SD=4). However, this difference is not statistically significant (two-sample Mann–Whitney–Wilcoxon test: p=0.20).

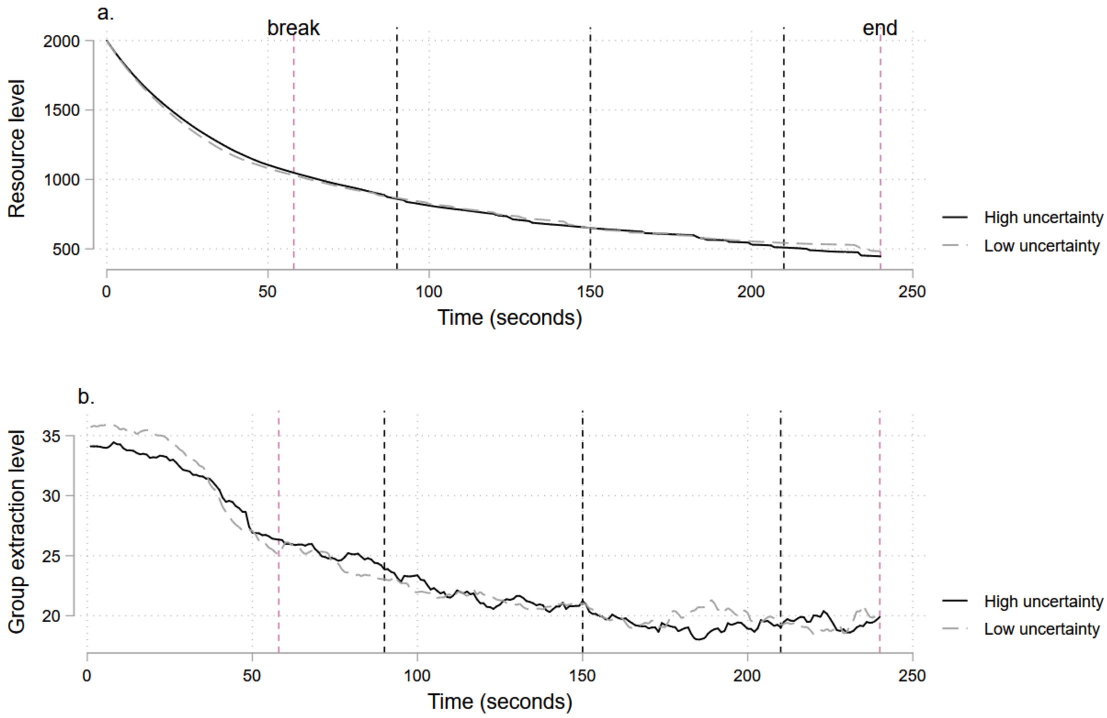


Fig. 3.1. Graph showing the resource development over time (0 to 240 seconds) (a) and the corresponding group extraction levels per second over time (b). The solid black line depicts the "High uncertainty" treatment and the dashed grey line depicts the "Low uncertainty" treatment in which participants knew the threshold range from the beginning of the round. The red vertical line at 58 seconds marks the time of the pause and the red line at 240 seconds marks the minimum round length (end). All groups in both treatments received the knowledge of the certain threshold during the pause at 58 seconds. The three blue vertical lines at 90, 150 and 210 seconds mark the three times at which we analyse groups' level of coordination.

3.3.2 Cooperation at the time of the pause

The majority of groups in both treatments successfully cooperated until the pause. Only 20% of the groups in HU (N=9, SD=40) and 24% of the groups in LU (N=11, SD=43) overexploited the resource below 970 resource units in the pre-pause part. We find no statistically significant difference in the proportion of groups that overexploited the resource between the two treatments (two-sided Fisher's exact test: p=0.8). This is also confirmed by Probit regressions (Table 3D.1).

We proceed to analyse the degree of groups' overexploitation. As explained above, this outcome measure ranges from zero if groups cooperated (keep the resource above 970) to 370 if groups failed to cooperate and drive the resource to the lowest possible resource level (600). We find no statistically significant difference in the degree of overexploitation between the two treatments (HU: Mean=19, SD=52, Min=0, Max=236; LU: Mean=17, SD=40, Min=0, Max=182; two-sample Mann–Whitney–Wilcoxon test (MWW): p=0.67). Again, Tobit regressions confirm this (Table 3D.2). Hence, we find no evidence that an imprecise early warning in form of threshold range knowledge changes cooperation amongst resource users. We thus reject Hypothesis 2.1.

Next, we add an exploratory analysis of the degree of overexploitation disregarding the structural differences between treatments. For this purpose, we define the degree of overexploitation as the exact MSY resource level (1,000 units) minus the resource level at the time of the pause. Thus, the degree of overexploitation ranges from underexploitation (negative values) to overexploitation (positive values). The box plots in Figure 3.2 show that the median (white line) is similar between HU (-28) and LU (-23) and we do not find a significant difference between the two treatments (HU: N=45, Mean=-48, SD=125, Min -306, Max=266; LU: N=45, Mean=-31, SD=118, Min=-505, Max=212; MWW: p=0.32). Yet, groups in LU are more centred around the median than in HU. In HU, the variability of the degree of overexploitation is higher amongst groups that underexploited the resource, i.e. kept the resource above the MSY (negative values). Underexploiting groups in HU have a significantly higher distance to the MSY than in LU (HU: N=22, Mean=-148, SD=72, min -306, max=-35;

LU: N=21, Mean=-117, SD=105, min=-505, max=-31; MWW: p=0.05). Thus, we find weak evidence for our expectation that the structural difference between treatments results in cooperative groups in HU being more cautious about driving the resource to the MSY compared to groups in LU.

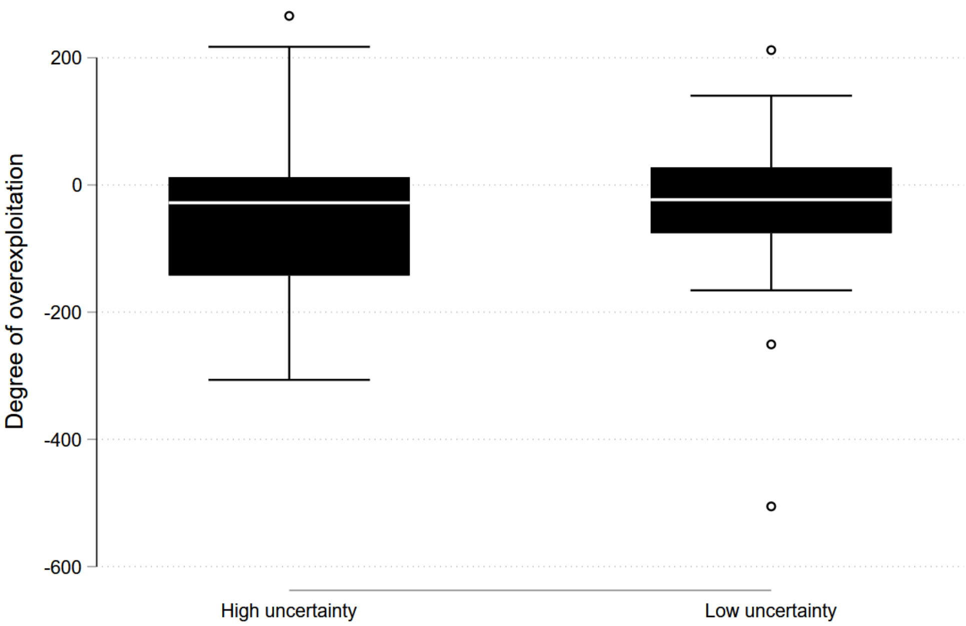


Fig. 3.2. Box plots of the degree of groups' overexploitation measured by the distance of the resource to the MSY level in treatment "High uncertainty" (HU) and "Low uncertainty" (LU) at the time of the pause (58 seconds). The white line represents the median, which is at -28 in HU and -23 in LU. The value 0 marks all groups that successfully cooperated, i.e. whose resource was exactly at the MSY of 1,000 resource units at the time of the pause. The circles mark outliers. Groups that kept the resource above 1,000 units have underexploited the resource (negative values), while groups that extracted the resource below 1,000 units have overexploited the resource (positive values).

3.3.3 Analysis of coordination

Figure 3.3 presents the percentage of groups that failed to coordinate and caused a collapse of the resource before 90, 150 or 210 seconds. At 90 seconds, only one out of 45 groups (2%, SD=0.15) in each treatment failed to coordinate causing an early collapse. At 150 seconds, more groups in LU (18%, SD=0.39) than HU (13%, SD=0.34) failed coordination. At 210 seconds however, more groups in HU (29%, SD=0.46) than LU (27%, 0.45) caused a collapse of the resource. Yet none of these differences between treatments are statistically significant (Table 3E.1). Furthermore, we find no significant treatment effects in Probit regressions with the failure of coordination as dependent variable (Table 3E.2). Thus, we find no evidence that different degrees of prior threshold ambiguity increase or decrease coordination failure after the certain threshold is known. Consequently, we reject Hypothesis 2.2.

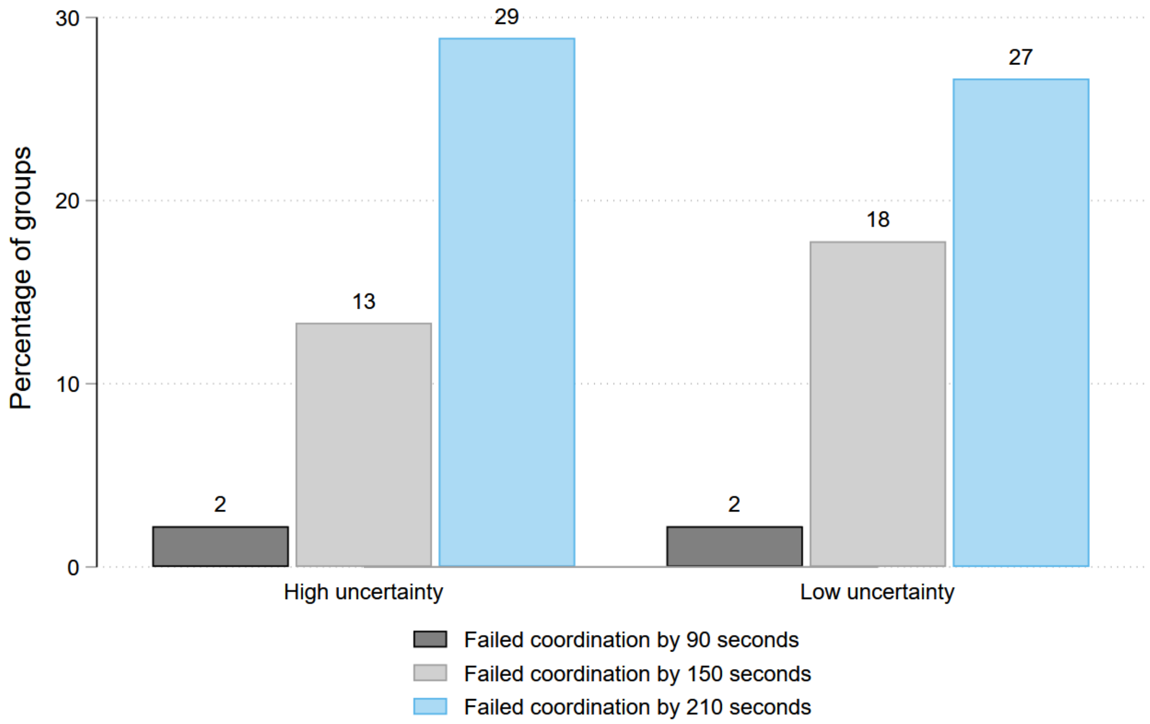


Fig. 3.3. Bar graphs showing the percentage of groups that failed to coordinate and caused a collapse of the resource by 90, 150 and 210 seconds per treatment "High uncertainty" (HU) and "Low uncertainty" (LU). At 90 seconds, only 2% of the groups in both treatments had failed to coordinate and caused an early collapse of the resource. At 150 seconds, 13% in HU and 18% in LU and at 210 seconds 29% in HU and 27% in LU had caused a collapse of the resource.

To shed further light on behavioural differences after the certain threshold is revealed, we conduct an exploratory analysis of cooperation at 90, 150 and 210 seconds. Overall, overexploitation as measure for failed cooperation increases over time, yet we find no significant differences between treatments at any of the analysed times (Table 3G.1). We further find no significant treatment effects in Tobit regressions with the degrees of overexploitation at the given times as dependent variables (Table 3G.2).

3.3.4 Impact of risk preferences on cooperation and coordination

Furthermore, in line with Aflaki (2013), we assume that the effect of higher threshold uncertainty on groups' resource extraction depends on groups' average risk measure. If groups are risk neutral, higher uncertainty potentially increases their resource extraction (ibid.). However, if groups are more risk averse, higher uncertainty could decrease their extraction (ibid.).

After the CPR game, we elicited individuals' risk preferences with the non-incentivised risk question developed by Dohmen et al. (2011). Participants rated their willingness to take on risk on a scale between 0 (risk averse) to 10 (risk loving). We ran Tobit regressions with the degree of overexploitation at 58, 90, 150 and 210 seconds as dependent variable (DV). At 58 seconds, we find no significant effect of the risk measure (Table 3D.2, p=0.12). However, at 90, 150 and 210 seconds we find positive effects of groups' average risk measure on the degree of overexploitation (Table 3G.2, Model 2, 5 and 8, p=0.03, p=0.04 and p=0.00 respectively). Furthermore, Probit regressions with coordination failure as DV only show a significant positive effect of groups' average risk measure at 210 seconds (Table 3E.2, Model 9, p=0.04). Hence, groups that were on average more risk loving were more likely to overexploit resources and less likely to coordinate successfully until the end of the round. Our findings are in contrast to Barrett and Dannenberg (2014b) and Maas et al. (2017) who find no significant effect of risk aversion.

In contrast to Aflaki (2013), we do not find evidence that the impact of the groups' average risk measure differs depending on the level of uncertainty. We do not find any significant interaction effect of groups' risk measure and treatment in the Tobit regressions with the degree of overexploitation at seconds 58, 90, 150 and 210 as DV (Table 3D.2 and 3G.2) and in the Probit regression with coordination failure as DV (Table 3E.2).

3.4 Discussion

The two treatments of our novel (quasi-) continuous-time CPR experiment for the lab varied in the degree of ambiguity about the threshold level. Under high uncertainty, participants merely knew of a threshold's existence. Under low uncertainty, participants knew the range of the potential threshold and thus, faced less ambiguity. With the latter treatment, we model an imprecise early warning. Previous studies on the effect of environmental uncertainty on resource management provided inconclusive evidence regarding the effect of higher uncertainty on cooperation and coordination amongst resource users. Moreover, these studies did not compare different levels of ambiguity. We do not find evidence that lower uncertainty about the threshold level due to threshold range knowledge has an impact on cooperation in comparison to a setting with complete ambiguity about the threshold level. Furthermore, we find no evidence that an imprecise early warning in form of threshold range knowledge affects coordination in groups once the exact threshold level is known.

In comparison to previous studies that focused on other aspects of threshold knowledge (e.g. Barrett and Dannenberg 2012, 2014a, 2014b, Dannenberg et al. 2015, Schill et al. 2015, Lindahl et al. 2016, Maas et al. 2017), we have a comparably large sample size. Assuming similar effect sizes, we would thus have sufficient power to find an effect of different degrees of ambiguity about the threshold level on cooperation and coordination if there was one. Thus, we conclude that different degrees of threshold ambiguity seem to matter less for behaviour than the comparison of no or uncertain threshold knowledge with certain threshold knowledge (e.g. Barrett and Dannenberg 2012, 2014a, 2014b, Dannenberg et al. 2015, Schill et al. 2015, Lindahl et al. 2016, Maas et al. 2017). Another explanation for our null effects could be that lower ambiguity about the threshold level can conceptually have positive and negative effects on cooperation and coordination, as we have outlined above. Thus, if both sets of conceptual arguments were true, these opposing effects could offset each other, which can well lead to a null effect on the treatment level.

It should be highlighted that even though we ran our experiment with students in the lab, our result that ambiguity of the threshold level does not cause a failure of cooperation or coordination is in line with previous evidence from lab-in-the-field experiments with actual resource users focusing on uncertainty about the threshold's presence (Rocha et al. 2015, Schill and Rocha 2019). A potential explanation could be that actual resource users who participate in lab-in-the-field experiments take context from the more complex reality of resource management into the simplified experiments. By giving the students a more complex design in our lab experiment, they may behave more like actual resource users. Thus, we see our results as an indication that the comparably complex experimental design of the (quasi-) continuous-time CPR game not only allows for a dynamic resource development and asynchronous, strategic interaction amongst group members (Pettit et al. 2014), but also mimics dynamics of actual resource management well.

We find evidence that groups who are on average more risk averse are less likely to overexploit and more likely to coordinate successfully. Our results on the impact of groups' average risk aversion somewhat contrasts that of Rocha et al. (2020) who find no significant effect of individuals' risk aversion on extraction and cooperation. They argue that group dynamics override the effect of individual risk preferences (ibid.). However, there seems to be an effect of risk preferences on resource management when a group shares specific preferences. Further research is needed to disentangle the effect of group dynamics and individuals' risk preferences on resource extraction choices with threshold uncertainty.

3.5 Conclusion

Ecosystems are under endogenous and exogenous pressure, for example through overexploitation and climate change. Thus, they become more likely to reach critical thresholds causing regime shifts to less favourable states (Scheffer et al. 2001). Ecological early warning signals could potentially help to inform resource users about approaching critical thresholds and the threat of regime shifts. However, ecological early warnings are difficult to detect and often come too late to implement targeted policies (Biggs et al. 2009, Crépin et al. 2012, Boettiger et al. 2013, Lenton et al. 2019). Therefore, resource users often face high levels of uncertainty regarding critical thresholds and approaching regime shifts.

Existing studies focused on comparing situations that are likely not faced in real world scenarios, such as the comparison of certain threshold knowledge or known probabilities vs. no threshold knowledge. Due to the underlying uncertainties in ecosystems' development and the potential of ecological early warning systems, we see different degrees of uncertainty, and particularly ambiguity, as the more policy relevant comparison for resource management. Our experiment focused on the impact of varying degrees of ambiguity about thresholds on cooperation and coordination in resource management. Our results yield no evidence that an imprecise early threshold warning in the form of broad threshold range knowledge affects overexploitation, and hence, no impact on cooperation. We also do not see an effect of the imprecise early warning on coordination after the uncertainty about the threshold level is resolved.

Based on our findings, we conclude that an early warning as we designed it makes no difference for resource management outcomes. However, in principle this could be because our chosen threshold range was too wide to be effective. Previous evidence based on a threshold public good game where participants knew the underlying probability distribution of the threshold level suggests that cooperation and coordination improve the smaller the range of the threshold level is (Barrett and Dannenberg 2014a, Wagener and de Zeeuw 2021). Further research is necessary to assess at which level of precision threshold range knowledge, as an early warning, has an effect on resource management and if such threshold warnings can be passed on to resource users without harmful consequences.

References Chapter 3

Aflaki, S. 2013. The effect of environmental uncertainty on the tragedy of the commons. *Games and Economic Behavior* 82:240–253.

Andreoni, J. 1988. Why free ride? Strategies and Learning in Public Goods Experiments. *Journal of Public Economics* 37:291–304.

Andreoni, J., and R. Croson. 2008. Partners versus Strangers: Random Rematching in Public Goods Experiments. Pages 776–783 *Handbook of Experimental Economics Results*.

Barrett, S., and A. Dannenberg. 2012. Climate Negotiations under Scientific Uncertainty. *Proceedings of the National Academy of Sciences* 109(43):17372–17376.

Barrett, S., and A. Dannenberg. 2014a. Sensitivity of collective action to uncertainty about climate tipping points. *Nature Climate Change* 4:36–39.

Barrett, S., and A. Dannenberg. 2014b. Negotiating to avoid 'gradual' versus 'dangerous' climate change: An experimental test of two prisoners' dilemmas. Pages 61–75 *in* T. L. Cherry, J. Hovi, and D. M. McEvoy, editors. *Toward a New Climate Agreement: Conflict, Resolution and Governance*. Taylor & Francis.

Biggs, R., S. R. Carpenter, and W. A. Brock. 2009. Turning back from the brink: Detecting an impending regime shift in time to avert it. *Proceedings of the National Academy of Sciences* 106(3):826–831.

Bigoni, M., M. Casari, A. Skrzypacz, and G. Spagnolo. 2015. Time Horizon and Cooperation in Continuous Time. *Econometrica* 83(2):587–616.

Bochet, O., J. Laurent-Lucchetti, J. Leroux, and B. Sinclair-Desgagné. 2019. Collective risk-taking in the commons. *Journal of Economic Behavior & Organization* 163:277–296.

Bock, O., I. Baetge, and A. Nicklisch. 2014. Hroot: Hamburg Registration and Organization Online Tool. *European Economic Review* 71:117–120.

Boettiger, C., N. Ross, and A. Hastings. 2013. Early warning signals: The charted and uncharted territories. *Theoretical Ecology* 6(3):255–264.

Brandt, G., M. M. Kulesz, D. Nissen, and A. Merico. 2017. OGUMI—A new mobile application to conduct common-pool resource experiments in continuous time. *PLoS ONE* 12(6).

Brown, T., and S. Kroll. 2017. Avoiding an uncertain catastrophe: climate change mitigation under risk and wealth heterogeneity. *Climatic Change* 141(2):155–166.

Budescu, D. V., A. Rapoport, and R. Suleiman. 1990. Resource dilemmas with environmental uncertainty and asymmetric players. *European Journal of Social Psychology* 20(6):475–487.

Budescu, D. V., A. Rapoport, and R. Suleiman. 1992. Simultaneous vs. sequential requests in resource dilemmas with incomplete information. *Acta Psychologica* 80(1–3):297–310.

Cerutti, N., and A. Schlüter. 2019. Resource changes: exogenous or endogenous, gradual or abrupt. Experimental evidence. *International Journal of Environmental Studies* 00(00):1–15.

Crépin, A.-S., R. Biggs, S. Polasky, M. Troell, and A. de Zeeuw. 2012. Regime shifts and management. *Ecological Economics* 84:15–22.

Dal Bó, P., and G. R. Fréchette. 2018. On the Determinants of Cooperation in Infinitely Repeated Games : A Survey. *Journal of Economic Literature* 56(1):60–114.

Dannenberg, A., A. Löschel, G. Paolacci, C. Reif, and A. Tavoni. 2015. On the Provision of Public Goods with Probabilistic and Ambiguous Thresholds. *Environmental and Resource Economics* 61(3):365–383.

Diekert, F. K. 2017. Threatening thresholds? The effect of disastrous regime shifts on the non-cooperative use of environmental goods and services. *Journal of Public Economics* 147(1):30–49.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner. 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association* 9(3):522–550.

Gustafsson, M., A. Biel, and T. Gärling. 1999. Overharvesting of resources of unknown size. *Acta Psychologica* 103:47–64.

Heinmiller, B. T. 2009. Path dependency and collective action in common pool governance. *International Journal of the Commons* 3(1):131–147.

Hendriks, A. 2012. *SoPHIE-Software platform for human interaction experiments*. Osnabrueck: Working Paper, Osnabrueck University.

Hine, D. W., and R. Gifford. 1996. Individual Restraint and Group Efficiency in Commons Dilemmas: The Effects of Two Types of Environmental Uncertainty. *Journal of Applied Social Psychology* 26(11):993–1009.

Janssen, M. A. 2010. Introducing Ecological Dynamics into Common-Pool Resource Experiments. *Ecology and Society* 15(2).

Kay, A. 2005. A Critique of the Use of Path Dependency in Policy Studies. *Public Administration* 83(3):553–571.

Lenton, T. M., J. Rockström, O. Gaffney, S. Rahmstorf, K. Richardson, W. Steffen, and H. J. Schellnhuber. 2019. Climate tipping points - too risky to bet against. *Nature* 575:592–595.

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.

Lovejoy, T. E., and C. Nobre. 2019. Amazon tipping point: Last chance for action. *Science Advances* 5(12):4–6.

Maas, A., C. Goemans, D. Manning, S. Kroll, and T. Brown. 2017. Dilemmas, coordination and defection: How uncertain tipping points induce common pool resource destruction. *Games and Economic Behavior* 104:760–774.

Millenium Ecosystems Assessment. 2005. *Ecosystems and Human Well-being: Synthesis*. Island Press, Washington, DC.

Perman, R., Y. Ma, M. Common, D. Maddison, and J. McGilvray. 2011. *Natural Resource and Environmental Economics*. 4th edition. Pearson Education.

Pettit, J., D. Friedman, C. Kephart, and R. Oprea. 2014. Software for continuous game experiments. *Experimental Economics* 17:631–648.

Rapoport, A., D. V. Budescu, R. Suleiman, and E. Weg. 1992. *Social dilemmas with uniformly distributed resources.* (W. B. G. Liebrand, D. M. Messick, and H. A. M. Wilke, editors). Internatio. Pergamon Press, Elmsford, NY, US.

Rocha, J. C., G. D. Peterson, and R. Biggs. 2015. Regime shifts in the anthropocene: Drivers, risks, and resilience. *PLoS ONE* 10(8):10–12.

Rocha, J. C., C. Schill, L. M. Saavedra-Díaz, R. D. P. Moreno, and J. H. Maldonado. 2020. Cooperation in the face of thresholds, risk, and uncertainty: Experimental evidence in fisher communities from Colombia. *PLoS ONE* 15(12 December):1–14.

Scheffer, M., S. Carpenter, J. A. Foley, C. Folke, and B. Walker. 2001. Catastrophic shifts in ecosystems. *Nature* 413(6856):591–596.

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).

Schill, C., and J. C. Rocha. 2019. *Uncertainty Can Help Protect Local Commons in the Face of Climate Change*. *SSRN Electronic Journal*.

StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.

Wagener, F., and de Zeeuw, A. 2021. Stable partial cooperation in managing systems with tipping points. *Journal of Environmental Economics and Management* 109: 102499.

Wharton Credibility Lab (2017) *AsPredicted*. Available at: https://aspredicted.org (Accessed: 20 October 2021).

Appendices Chapter 3

Appendix 3A: Details of the user-resource model

The experimental setting of our (quasi) continuous-time common-pool resource (CPR) game is based on the experiment presented in Chapter 2. Here, we provide more detail on the user-resource model and its parametrisation.

As in Chapter 2, the logistic growth term in equation (3A.1) below describes the natural growth $G(R_t)$ of the resource R_t that changes over time t, with the resource growth rate $g = 0.04$ and the carrying capacity $R_{max} = 2{,}000$ resource units, as long as the resource R is above the threshold $R_{min} = 400$ resource units.

$$R_{t+1} - R_t = \begin{cases} gR_t\left(1 - \frac{R_t}{R_{max}}\right) - \sum_{i=1}^{n} E_{it} & if\ R_t > R_{min} \\ 0 & if\ R_t \leq R_{min} \end{cases} \qquad (3A.1)$$

$\sum_{i=1}^{n} E_{it}$ is the group's cumulative extraction per second. There are $n = 4$ resource users per group. In our experiment, participants' extraction is cost-free and the amount of extracted resource units is constant per level of extraction. For example, an extraction level of 1 always results in 1 extracted resource unit independent of the current resource level.

We implement the regime shift by integrating an irreversible and persistent collapse of the resource once the resource level R_t reaches the certain threshold R_{min}: $R_{t+1} = 0\ if\ R_t \leq R_{min}$. The growth rate of the resource also collapses immediately ($R_{t+1} - R_t = 0$) once the threshold is reached and stays at 0 resource units infinitely.

Next, we outline the calculation of the resource level of maximum sustainable yield (MSY) and the corresponding extraction level. The dynamics of the resource R_t change over time t depending on the natural growth of the resource $G(R_t)$ and the change of the resource through groups' extraction per second $\sum_{i=0}^{n} E_{it}$. The natural growth of the resource is based on a logistic growth function (Perman et al. 2011), thus:

$$R_{t+1} - R_t = G(R_t) - \sum_{i=0}^{n} E_{it} \tag{3A.2}$$

$$\Leftrightarrow R_{t+1} - R_t = gR_t\left(1 - \frac{R_t}{R_{max}}\right) - \sum_{i=0}^{n} E_{it}$$

with the resource level R_t, the instrinsic resource growth rate g and the carrying capacity R_{max}.

Given no harvest by the group, $\sum_{i=0}^{n} E_{it} = 0$, the resource's regrowth is at its maximum when the resource level is at the level of maximum sustainable yield (MSY):

$$\max_{R} \frac{dR}{dt} = gR\left(1 - \frac{R}{R_{max}}\right) \tag{3A.3}$$

$$FOC: g - \frac{2gR}{R_{max}} = 0$$

$$\Leftrightarrow g = \frac{2gR}{R_{max}}$$

$$\Leftrightarrow R = \frac{R_{max}}{2}$$

$$\Rightarrow R_{MSY} = \frac{R_{max}}{2} = 1{,}000$$

To reach a steady state at the resource level of maximum sustainable yield $R_{MSY} = \frac{R_{max}}{2}$ where the natural growth is at its maximum, the group's extraction $\sum_{i=0}^{n} E_i$ needs to be equal to the resource's natural growth. Thus:

$$\sum_{i=0}^{n} E_i = g\frac{R_{max}}{2}\left(1 - \frac{\frac{R_{max}}{2}}{R_{max}}\right) \tag{3A.4}$$

$$\Leftrightarrow \sum_{i=0}^{n} E_i = g\frac{R_{max}}{2}\frac{1}{2}$$

$$\Leftrightarrow \sum_{i=0}^{n} E_i = g\frac{R_{max}}{4}$$

$$\Leftrightarrow E_{MSY} = g\frac{R_{max}}{4}$$

$$\Leftrightarrow E_{MSY} = 0.04\frac{2{,}000}{4} = 20$$

Therefore, the optimal group extraction at $R_{MSY} = 1{,}000$ resource units is $E_{MSY} = 20$.

While the experiment is running, the resource is updated every second such that the resource in the next second R_{t+1} is based on the resource level in the previous second R_t, its growth and the group's extraction $\sum_{i=0}^{n} E_i$ in the previous second t. Thus:

$$R_{t+1} = R_t + gR_t\left(1 - \frac{R_t}{R_{max}}\right) - \sum_{i=0}^{n} E_i$$

The continuous-time design of our CPR game has the advantage that the implemented resource development can be more dynamic than in round-based CPRs and that group members are able to interact with the resource and each other in a speedy, asynchronous way (Pettit et al. 2014). Such flexible adjustment dynamics allow experiments based on a continuous-time design to mimic long-term interactions amongst subjects in a relatively short time compared to round-based designs (ibid). However, the continuity of our CPR game is limited because technical restrictions only allow for an update of the resource development and subjects' extraction choices every second. Therefore, we define it as a quasi-continuous-time game (Bigoni et al. 2015).

Parameterisation of the user-resource model

The parameterisation of the resource development is based on the experimental set-up presented in Chapter 2:

- Carrying capacity of the resource: $R_{max} = 2000$ resource units
- Growth factor of the resource: $g = 0.04$

Under the assumption of an infinite time horizon, this parameterisation leads to:

- $R_{MSY} = \frac{R_{max}}{2} = 1{,}000$ resource units as resource level of MSY.

- $E_{MSY} = g\frac{R_{max}}{4} = 20$ resource units as maximum sustainable group extraction per second. Joint group maximisation does not mean that all group members need to have the same choice of extraction. As long as the cumulative extraction $\sum_{i=0}^{n} E_i$ is at the level of the MSY extraction $E_{MSY} = g\frac{R_{max}}{4}$, some group members can have a lower extraction rate while others might have a higher extraction rate.

In two aspects, the parametrisation is different from Chapter 2 and adapted to the experimental design presented in this paper due to the following constraints:

1. At the time of the pause (58 seconds), the current resource level R_t should be:
   i. sufficiently higher than the threshold level R_{min} to make sure that even groups that extract the resource at the maximum from the beginning cannot reach the certain threshold $R_{min} = 400$ until the pause:

      $R_t > R_{min}$ at $t = 58$

   ii. possible to be below the MSY level (defined as 970 instead of 1,000) to measure failure of cooperation:

      $R_t < R_{MSY}$ at $t = 58$

   iii. possible to be below the upper limit of the threshold range (700 resource units) known to participants in the “Low uncertainty” (LU) treatment to measure failure of early coordination on the upper limit of the known threshold range:

      $R_t < R_{upperLimit}$ at $t = 58$

   These constraints need to hold in the extreme case that each group member choses 10 as extraction level and thus, the group extracts at the maximum rate of 40 resource units per second from the start of the round. Thus, we choose 400 resource units as critical threshold, which can earliest be reached after 66 seconds of maximum extraction. Thereby, groups that extract at the maximum rate from the beginning would have the time during the pause and about 8 seconds after the pause to reassess their extraction strategy. At the same time, it is sufficiently likely that groups, which do not cooperate and do not coordinate on the upper threshold limit in LU drive the resource below $R_{MSY} = 1{,}000$ resource units and $R_{upperLimit} = 700$ resource units, respectively.

Test sessions at the LaER laboratory at Osnabrueck University in October 2019 and at the WISO Experimental Lab of Hamburg University in November 2019 showed that 8 seconds is sufficient time for groups to adapt extraction strategies to avoid the threshold if wanted after the pause. Even groups that drove the resource below the MSY at the time of the pause managed to sustain the resource beyond the earliest time of a possible collapsing resource at 66 seconds. None of the four groups that played the LU treatment in the two test sessions drove the resource below the upper limit of the known threshold range.

2. Choice of the extraction rate that maintains the MSY, once it is reached, avoids reaching the threshold and the collapse of the resource. Under that cumulative extraction choice, the resource should be sustained indefinitely.
    i. Implementation of a random continuation rule to set the incentive to choose $E_{MSY} = g\frac{R_{max}}{4}$ once the resource is at $R_{MSY} = 1{,}000$ resource units (Dal Bó and Fréchette 2018).

Appendix 3B: Implementation of the experiment

Figure 3B.1 shows a screenshot of the CPR game as it was presented to participants on screen. The graph at the top presents the resource development over time (in seconds). Participants use a slider to choose their extraction level between 0 and 10 (bottom row, left). As a control mechanism, participants can observe the development of their extraction level over time (bottom row, centre). If participants choose a higher extraction level by moving the slider to the right, the graph of their individual extraction level should increase and *vice versa* if participants choose a lower extraction level. Furthermore, participants see their individual total extraction as well as their group's total extraction (bottom row, right) at any point in time. Thus, participants receive immediate feedback about their own extraction in comparison to their group's extraction. Participants in both treatments gave the feedback that the slider was easy to handle with an average rating of 4.4 (SD=0.95, Table 3C.1).

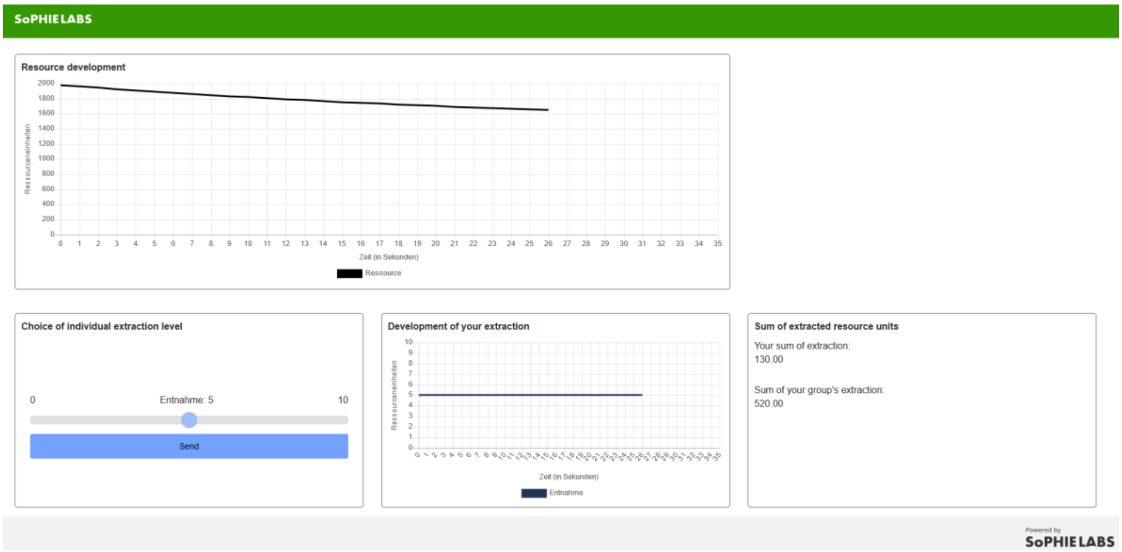


Fig. 3B.1. Screenshot of the CPR game (English translation of German original)

During the pause in the pay out-relevant round all participants had to choose a new extraction level before they could continue with the round. Participants could chose the same extraction level as they had chosen before the pause. However, we implemented the choice of the extraction level in the pause to motivate every participants to make an active choice and to prevent participants from continuing with their pre-pause extraction level simply because they preferred the status quo (*status quo bias*). Once all group members had chosen their new extraction level, the resource development continued at the pre-pause resource level.

Strategies against endgame effects

Even with the implemented random continuation rule, the experimental results presented in Chapter 2 show that some groups do not expect the experiment to continue much longer than the certain minimum round length. Even though the results analysed at 240 seconds in Chapter 2 are robust to the emerging endgame effects, we decided to adjust our experimental design in this paper to make sure that we reduce any potential endgame effects in our data analysis.

Under the assumption that the round ended immediately after the minimum round length (240 seconds), endgame effects potentially motivated increased extraction towards the 240-seconds mark. It would have maximised groups' pay out to reach the threshold exactly at 240 seconds. Coming from the MSY, groups would have needed to extract the resource at the maximum of 40 units per second from 213 seconds onwards. An earlier onset of overexploitation would have decreased groups' maximum expected pay out. Thus, no endgame effects should have affected groups' extraction before 210 seconds. Hence, we set 210 seconds as the defined end for the pay out-relevant round in this paper instead of 240 seconds as in Chapter 2.

Details of experimental procedures

On arrival to the experimental session, participants were randomly seated in an enclosed computer cabin. The experimenter randomly distributed participant codes for participants to login into the experimental software SoPHIE (Hendriks 2012). Participants' identity remained anonymous to the researcher and cannot be connected to the used participant codes. The experimental instructions consisted of two parts. The experimenter read a first part aloud and a second part was presented to participants on screen only. Participants were not allowed to communicate with each other or to use any technical devices during the experiment. However, they could ask questions to clarify their understanding of the instructions to the experimenter in private. Please see the instructions for further details (Appendix 3H).

Understanding of the instructions

We used not-payout-relevant control and feedback questions to assure and check participants' understanding of the experiment. First, participants had to answer three control questions before they played the two test rounds to secure their understanding of the resource development and the extraction mechanism. Next, they had to answer four additional control questions after they received the threshold information, i.e. before they started playing the pay out-relevant CPR round. Individuals needed, on average, one try only to answer all control questions correctly (SD=0.2), which indicates a good understanding of the experimental dynamics (Table 3C.1). Furthermore, individuals were asked to rate the instructions and the understanding of the resource development on a 5-point Likert scale as part of the post-experimental questionnaire (1 means: "strongly reject" and the value 5 means: "strongly approve"). On average, individuals' rated the instructions to be well written (Mean=4.5, SD=0.8) and stated that they understood the resource development well (Mean=4.4, SD=0.9, Table 3C.1). Again, we do not find a statistically significant difference between the two treatments (Table 3C.1). Please see the instructions for details of the design of the post-experimental questionnaire (Appendix 3H).

Appendix 3C: Details of sample's socio-economic characteristics

Table 3C.1. Balance table of individuals' socio-economic characteristics and their answers to control and feedback questions.

	(1) Total		(2) High uncertainty		(3) Low uncertainty		t-test Difference
Variables	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	**(2)-(3)**
Age (years)	360	25.511	180	25.583	180	25.439	0.144
		(4.838)		(4.907)		(4.781)	
Female (1=yes, 0=no)	360	0.628	180	0.628	180	0.628	0.000
		(0.484)		(0.485)		(0.485)	
Previous experiment experience (1=yes, 0=no)	360	0.894	180	0.878	180	0.911	-0.033
		(0.308)		(0.328)		(0.285)	
Student (1=yes, 0=no)	360	0.994	180	0.989	180	1.000	-0.011
		(0.074)		(0.105)		(0.000)	
Economics student (1=yes, 0=no)	360	0.289	180	0.294	180	0.283	0.011
		(0.454)		(0.457)		(0.452)	
Monthly income (categorical)†	343	2.843	171	2.918	172	2.767	0.151
		(1.146)		(1.150)		(1.141)	
Individual control question score‡	360	1.078	180	1.074	180	1.082	-0.008
		(0.158)		(0.143)		(0.172)	
Risk measure (0: not willing to take risk, 10: very willing to take risks)	360	5.169	180	5.289	180	5.050	0.239
		(2.213)		(2.221)		(2.205)	
Expected continuation of round (categorical)§	267	3.476	134	3.187	133	3.767	-0.580***
		(1.576)		(1.562)		(1.542)	
Well written instructions (1: strongly reject, 5: strongly approve)	360	4.478	180	4.539	180	4.417	0.122
		(0.834)		(0.779)		(0.884)	
Understood resource development well (1: strongly reject, 5: strongly approve)	360	4.394	180	4.450	180	4.339	0.111
		(0.920)		(0.867)		(0.970)	
Easy to handle slider (1: strongly reject, 5: strongly approve)	360	4.417	180	4.433	180	4.400	0.033
		(0.949)		(0.898)		(1.001)	

Table 3C.1. Continued. Balance table of individuals' socio-economic characteristics and their answers to control and feedback questions.

	(1)		(2)		(3)		t-test
	Total		High uncertainty		Low uncertainty		Difference
Variables	N	Mean/SD	N	Mean/SD	N	Mean/SD	**(2)-(3)**
Number of groups per session	360	5.467	180	5.467	180	5.467	0.000
		(0.886)		(0.887)		(0.887)	
Total payout (GBP)	360	14.739	180	14.417	180	15.062	-0.646
		(5.538)		(5.026)		(6.002)	
F-test of joint significance (F-stat)							1.206*
F-test, number of observations							259

Note: N denotes the numbers of individuals that answered the given question. The values displayed for t-tests are the differences in the means across the groups. The values displayed for F-tests are the F-statistics.
† **Monthly income** was elicited in five categories: 1. 0 to 300 Euro; 2. 301 to 600 Euro; 3. 601 to 900 Euro; 4. 901 to 1,200 Euro and 5. More than 1,200 Euro. Individuals that gave no answer were coded as missing observation. ‡ The **individual control question score** is based on individuals' answers to the seven control questions and calculated by dividing the number of tries that each individual took to answer all seven questions correctly by seven. § **Expected continuation of round** measures individuals' expected continuation of the round beyond the certain minimum round length. Individuals chose one of the following seven categories: 1. Immediate end, 2. Less than 30 seconds, 3. Another 30 seconds, 4. 30 to 60 seconds, 5. 60 to 90 seconds, 6. More than 90 seconds and 7. No expectation. Participants that answered "no expectation" (7) were coded as missing observation.
Standard deviations are presented in brackets.
* p < 0.1; ** p < 0.05; *** p < 0.01

Appendix 3D: Further details of the analysis of cooperation

Table 3D.1. Probit regression models with failure of cooperation at the time of the pause (58 seconds) as dependent variable.

	Outcome variable: Groups' failure of cooperation (coop fail) before 58 seconds.		
Variables	**(1)** **Coop fail 58**	**(2)** **Coop fail 58**	**(3)** **Coop fail 58**
Low uncertainty treatment	0.15	0.32	1.82
	[-0.43 - 0.73]	[-0.31 - 0.95]	[-1.69 - 5.34]
	(0.61)	(0.32)	(0.31)
Age (group average in years)		-0.06	-0.06
		[-0.19 - 0.08]	[-0.19 - 0.08]
		(0.40)	(0.41)
Gender (group fraction females)		-0.82	-0.64
		[-2.21 - 0.58]	[-2.10 - 0.82]
		(0.25)	(0.39)
Risk measure (group average)†		0.24	0.43
		[-0.05 - 0.54]	[-0.10 - 0.95]
		(0.10)	(0.11)
Low uncertainty treatment x Risk measure			-0.27
			[-0.90 - 0.35]
			(0.39)
Expected round continuation (group average category)‡		-0.19	-0.19
		[-0.48 - 0.10]	[-0.49 - 0.10]
		(0.20)	(0.19)
Constant	-0.84***	0.57	-0.58
	[-1.26 - -0.42]	[-3.70 - 4.84]	[-5.71 - 4.55]
	(0.00)	(0.80)	(0.82)
Observations	90	90	90
Prob > chi2	0.61	0.12	0.14

Note: Model 1 tests the treatment effect of treatment "Low uncertainty" in comparison to the control treatment "High uncertainty". Model 2 includes groups' average age, groups' fraction of females, groups' average risk measure and groups' average round continuation expectation as socio-economic controls. Model 3 additionally includes the interaction of the groups' average risk measure and treatment "Low uncertainty".
Cooperation failure (coop fail 58) is defined by the resource level at the time of the pause (58 seconds). Cooperation failure equals 1 if groups did not cooperate successfully and drove the resource below the defined level of MSY (970 resource units) before the pause and 0 in case of successful cooperation at or above the MSY.
† **Risk measure** was elicited with the non-incentivised risk question developed by Dohmen et al. (2011). Participants rated their willingness to take on risk on an 11-point Likert scale between 0 (risk averse) to 10 (risk loving). ‡ **Expected round continuation** measures participants' expected continuation of the round beyond the certain minimum round length. Participants chose one of the following seven categories: 1. Immediate end, 2. Less than 30 seconds, 3. Another 30 seconds, 4. 30 to 60 seconds, 5. 60 to 90 seconds, 6. More than 90 seconds and 7. No expectation. Participants that answered "no expectation" (7) were coded as missing before calculating the group's mean category.
Confidence intervals (at the 95% confidence level) are presented in square brackets.
P-values are presented in brackets.
* p < 0.1; ** p < 0.05; *** p < 0.01

Table 3D.2. Tobit regression models with the degree of overexploitation at the time of the pause (58 seconds) as measure of cooperation failure as the dependent variable.

Outcome variable: Groups' degree of overexploitation at 58 seconds.			
Variables	**(1) Overexploitation 58**	**(2) Overexploitation 58**	**(3) Overexploitation 58**
Low uncertainty treatment	8.79	27.03	207.84
	[-70.05 - 87.64]	[-54.70 - 108.76]	[-232.13 - 647.81]
	(0.83)	(0.51)	(0.35)
Age (group average in years)		-5.68	-5.38
		[-23.28 - 11.92]	[-22.91 - 12.16]
		(0.52)	(0.54)
Gender (group fraction females)		-103.54	-78.68
		[-292.05 - 84.97]	[-273.82 - 116.45]
		(0.28)	(0.42)
Risk measure (group average)†		30.74	51.79
		[-8.24 - 69.72]	[-13.81 - 117.39]
		(0.12)	(0.12)
Low uncertainty treatment x Risk measure			-32.95
			[-111.07 - 45.18]
			(0.40)
Expected round continuation (group average category)‡		-16.45	-16.48
		[-53.77 - 20.87]	[-53.60 - 20.63]
		(0.38)	(0.38)
Constant	-111.97***	-3.37	-142.74
	[-190.83 - -33.11]	[-562.21 - 555.47]	[-800.55 - 515.07]
	(0.01)	(0.99)	(0.67)
Observations	90	90	90
left-censored	70	70	70
right-censored	0	0	0
Prob > chi2	0.8242	0.2134	0.2513

Note: Model 1 tests the treatment effect of "Low uncertainty" in comparison to the control treatment "High uncertainty". Model 2 includes groups' average age, groups' fraction of females, groups' average risk measure and groups' average round continuation expectation as socio-economic controls. Model 3 additionally includes the interaction of the groups' average risk measure and treatment "Low uncertainty".
Degree of overexploitation is defined as 970 resource units minus the current resource level. The degree of overexploitation equals 0, if groups cooperate successfully and keep the resource above the level of maximum sustainable yield (lower limit). The degree of overexploitation equals 370, if groups do not cooperate successfully and extract the resource at the maximum rate until the time of the pause (upper limit).
† **Risk measure** was elicited with the non-incentivised risk question developed by Dohmen et al. (2011). Participants rated their willingness to take on risk on an 11-point Likert scale between 0 (risk averse) to 10 (risk loving). ‡ **Expected round continuation** measures participants' expected continuation of the round beyond the certain minimum round length. Participants chose one of the following seven categories: 1. Immediate end, 2. Less than 30 seconds, 3. Another 30 seconds, 4. 30 to 60 seconds, 5. 60 to 90 seconds, 6. More than 90 seconds and 7. No expectation. Participants that answered "no expectation" (7) were coded as missing before calculating the group's mean category.
Confidence intervals (at the 95% confidence level) are presented in square brackets.
P-values are presented in brackets.
* p < 0.1; ** p < 0.05; *** p < 0.01

Appendix 3E: Further details of the analysis of coordination

Table 3E.1. Summary statistics of coordination failure in treatment "High uncertainty" and "Low uncertainty" and the results of two-sided Fisher's exact tests of the fraction of groups that failed to coordinate and caused a collapse of the resource before 90, 150 or 210 seconds.

	High uncertainty (HU)		Low uncertainty (LU)		
Fraction of groups that failed coordination	**N**	**Mean (SD)**	**N**	**Mean (SD)**	**two-sided Fisher's exact test (p-value)**
by 90 seconds	45	0.02 (0.15)	45	0.02 (0.15)	1.0
by 150 seconds	45	0.13 (0.34)	45	0.18 (0.39)	0.7
by 210 seconds	45	0.29 (0.46)	45	0.27 (0.45)	1.0

Note: Coordination failure is defined by the resource level at the analysed time (90, 150 or 210 seconds). Coordination failure equals 1 if groups did not coordinate successfully and caused a collapse of the resource by reaching the threshold at 400 resource units and 0 in case of successful coordination on a resource level above the threshold.
N denotes the number of groups. Each group consists of four participants.
Standard deviation (SD) in brackets.
*** p<0.01, ** p<0.05, * p<0.1

Table 3E.2. Probit regression models with failure of coordination before 90, 150 or 210 seconds as dependent variables.

	Outcome variable: Groups' failure of coordination (coord fail) before 90 seconds (model 1 to 3), before 150 seconds (model 4 to 6) and before 210 seconds (model 7 to 9).								
Variables	**(1) Coord fail 90**	**(2) Coord fail 90**	**(3) Coord fail 90**	**(4) Coord fail 150**	**(5) Coord fail 150**	**(6) Coord fail 150**	**(7) Coord fail 210**	**(8) Coord fail 210**	**(9) Coord fail 210**
Low uncertainty treatment	-0.00 [-1.15 - 1.15] (1.00)	0.23 [-3.30 - 3.76] (0.90)	12.76 [-22.21 - 47.73] (0.47)	0.19 [-0.44 - 0.82] (0.56)	0.19 [-0.47 - 0.85] (0.57)	-0.05 [-3.33 - 3.23] (0.98)	-0.07 [-0.62 - 0.49] (0.81)	-0.08 [-0.67 - 0.51] (0.79)	0.96 [-2.06 - 3.98] (0.53)
Age (group average in years)		0.17 [-0.50 - 0.83] (0.62)	0.22 [-0.46 - 0.90] (0.52)		-0.10 [-0.25 - 0.06] (0.23)	-0.10 [-0.26 - 0.06] (0.23)		-0.09 [-0.23 - 0.04] (0.16)	-0.09 [-0.22 - 0.04] (0.17)
Gender (group fraction females)		1.53 [-8.37 - 11.44] (0.76)	3.67 [-9.96 - 17.30] (0.60)		0.58 [-0.95 - 2.11] (0.46)	0.54 [-1.07 - 2.15] (0.51)		1.57** [0.07 - 3.07] (0.04)	1.74** [0.14 - 3.34] (0.03)
Risk measure (group average)†		1.67 [-1.29 - 4.62] (0.27)	3.76 [-3.60 - 11.13] (0.32)		0.23 [-0.08 - 0.55] (0.14)	0.21 [-0.28 - 0.69] (0.40)		0.36** [0.06 - 0.65] (0.02)	0.48** [0.02 - 0.93] (0.04)
Low uncertainty treatment x Risk measure			-2.01 [-7.28 - 3.25] (0.45)			0.05 [-0.55 - 0.64] (0.88)			-0.20 [-0.75 - 0.36] (0.49)
Expected round continuation (group average category)‡		1.77 [-1.46 - 5.00] (0.28)	2.01 [-1.32 - 5.35] (0.24)		-0.03 [-0.33 - 0.26] (0.83)	-0.03 [-0.33 - 0.27] (0.83)		0.03 [-0.23 - 0.29] (0.82)	0.03 [-0.23 - 0.30] (0.82)
Constant	-2.01*** [-2.82 - -1.20] (0.00)	-26.89 [-81.30 - 27.51] (0.33)	-44.23 [-128.48 - 40.01] (0.30)	-1.11*** [-1.57 - -0.65] (0.00)	-0.13 [-4.93 - 4.67] (0.96)	0.05 [-5.31 - 5.40] (0.99)	-0.56*** [-0.94 - -0.17] (0.00)	-1.19 [-5.36 - 2.99] (0.58)	-1.98 [-6.80 - 2.84] (0.42)
Observations	90	90	90	90	90	90	90	90	90
Prob > chi2	1	0.0927	0.1104	0.5602	0.4891	0.6155	0.8139	0.0612	0.0877

Note: In Model 1 to 3 the dependent variable is groups' failure of coordination (Coord fail) before 90 seconds, in Model 4 to 6 before 150 seconds and in model 7 to 9 before 210 seconds. Model 2, 5 and 8 include groups' average age, groups' fraction of females, groups' average risk measure and groups' average round continuation expectation as socio-economic controls. Model 3, 6 and 9 additionally include the interaction of the groups' average risk measure and treatment "Low uncertainty".
Coordination failure (coord fail) is defined by the resource level at the analysed time (90, 150 or 210 seconds). Coordination failure equals 1 if groups did not coordinate successfully and cause a collapse of the resource by reaching the threshold at 400 resource units and 0 in case of successful coordination on a resource level above the threshold.
† **Risk measure** was elicited with the non-incentivised risk question developed by Dohmen et al. (2011). Participants rated their willingness to take on risk on an 11-point Likert scale between 0 (risk averse) to 10 (risk loving). ‡ **Expected round continuation** measures participants' expected continuation of the round beyond the certain minimum round length. Participants chose one of the following seven categories: 1. Immediate end, 2. Less than 30 seconds, 3. Another 30 seconds, 4. 30 to 60 seconds, 5. 60 to 90 seconds, 6. More than 90 seconds and 7. No expectation. Participants that answered "no expectation" (7) were coded as missing before calculating the group's mean category.
Confidence intervals (at the 95% confidence level) are presented in square brackets.
P-values are presented in brackets.
* p < 0.1; ** p < 0.05; *** p < 0.01

Appendix 3F: Additional pre-registered analysis of the degree of coordination

We preregistered the analysis of a continuous variable “distance to threshold” (correct to 2 decimal places) as measure for the degree of coordination of groups that do not cause a collapse of the resource. It is defined by the distance of the given resource level R_t to the threshold R_{min} ($R_t - R_{min}$) and is zero for all groups that reached the threshold at 400 resource units. We decided against reporting the results of this analysis in the main text because we no longer think that there is a meaningful interpretation of the variable degree of coordination. Regarding coordination it is relevant whether groups reached the threshold and thus, failed to coordinate or not. Thus, only the cut-off at the threshold at 400 resource units is relevant. Higher resource levels are not necessarily measuring higher degrees of coordination, but rather signal the degree of cooperation in the groups. Thus, instead of reporting the degree of coordination at 90, 150 and 210 seconds in the main text, we decided to report our exploratory analysis of the degree of cooperation towards the end. The degree of cooperation is indeed higher, the closer the resource is to the level of maximum sustainable yield. Thus, an analysis of the degree of cooperation gives more meaningful insights than an analysis of the degree of coordination (see Section “Analysis of coordination” and Appendix 3G for the results).

Nonetheless, we report the results of our pre-registered analysis of the degree of coordination for the interested reader in the following. We use two-sample Mann–Whitney–Wilcoxon (MWW) tests to check for differences in the groups’ degree of coordination, i.e. distances to the threshold at the times of interest (90, 150, and 210 seconds). As shown in Figures 3F.1b-d, the distance to the threshold decreases over time. At 90 seconds, groups’ degree of coordination in HU is more widely spread than in LU, which is more centred on the median. However, in general the distribution of the distance to the threshold is similar for the two treatments. We do not find any statistically significant differences in the degree of coordination between the two treatments at any of the times of interest (MWW tests, Table 3F.1).

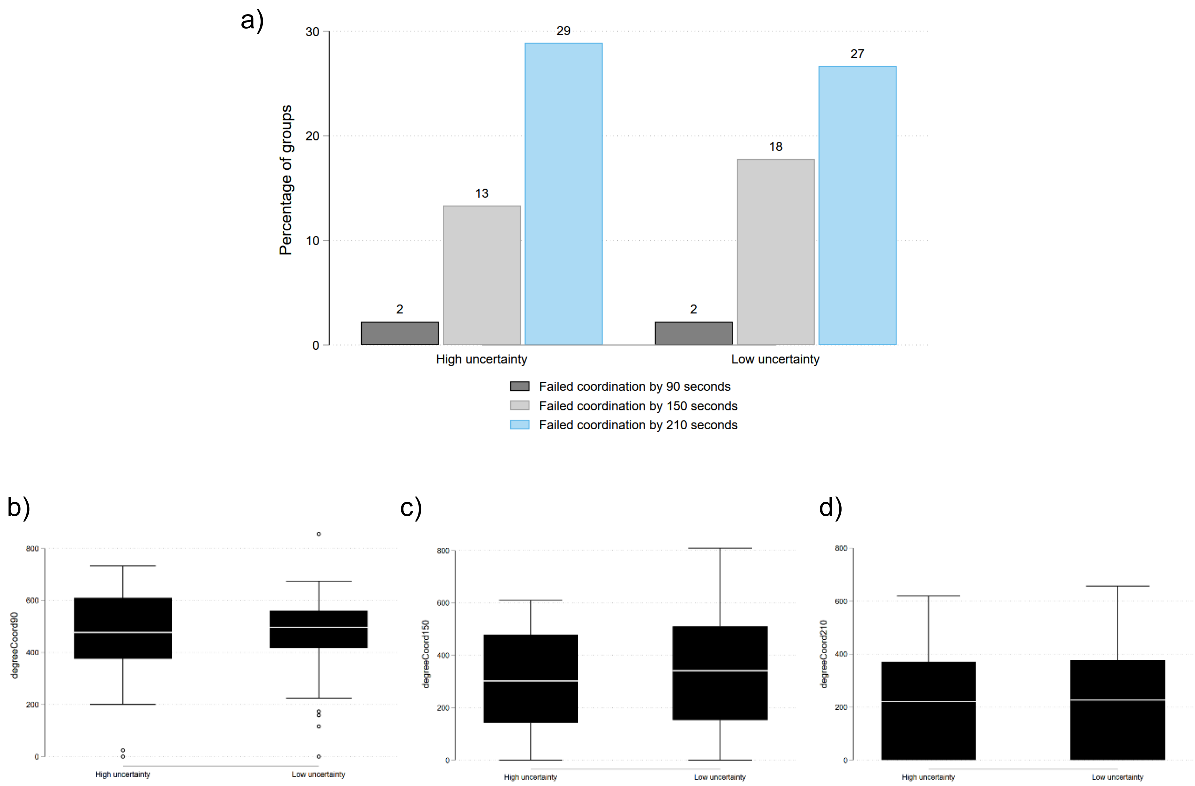


Fig. 3F.1. Bar graphs of the percentages of groups that failed to coordinate per treatment "High uncertainty" (HU) and "Low uncertainty" (LU) (a). Boxplots of distance to threshold per treatment at 90, 150 and 210 seconds (b-d). At 90 seconds, only 2% of the groups in both treatments had failed to coordinate and caused an early collapse of the resource. At 150 seconds, 13% in HU and 18% in LU and at second 210 29% in HU and 27% in LU had caused a collapse of the resource. In Fig. 3F.1b-d all groups that failed to coordinate and caused a collapse of the resource are shown with zero distance to threshold. The maximum distance to threshold would be 1,600 if groups stayed at the maximum carrying capacity of 2,000 units.

Table 3F.1. Summary statistics of degree of coordination and the results of two-sample Mann–Whitney–Wilcoxon tests at 90, 150 and 210 seconds.

	High uncertainty (HU)		**Low uncertainty (LU)**		
Degree of coordination	**N**	**Mean (SD)**	**N**	**Mean (SD)**	**Two-sample Mann–Whitney–Wilcoxon test (p-value)**
at 90 seconds	45	469 (163)	45	475 (158)	0.89
at 150 seconds	45	304 (201)	45	323 (229)	0.65
at 210 seconds	45	226 (192)	45	250 (214)	0.74

Note: Degree of coordination is defined as the current resource level minus the threshold level of 400 resource units. The degree of coordination equals 0, if groups did not coordinate successfully and caused a collapse of the resource by reaching the threshold (minimum). The degree of coordination equals 1599, if groups coordinate on keeping the resource at the highest possible level, the maximum carrying capacity of 2,000 resource units (maximum).
N denotes the number of groups. Each group consists of four participants.
Standard deviation (SD) in brackets.
*** p<0.01, ** p<0.05, * p<0.1

Further, we estimate six linear two-equation Hurdle models including three specifications with socio-economic control variables to assess the degree of coordination at 90, 150 and 210 seconds. The Hurdle models combine a selection model that determine the boundary points of the degree of coordination with the outcome model on the nonbounded values of degree of coordination (StataCorp 2017). The lower-limit of our selection model is zero, which marks coordination failure. In line with our non-parametric tests, we find no significant treatment effects on the degree of coordination at any of the analysed times (Table 3F.2).

Table 3F.2. Six linear two-equation Hurdle models with degree of coordination (degree coord) at second 90, 150 and 210 as dependent variables.

	Outcome variable: Groups' degree of coordination at second 90 (model 1 and 4), at second 150 (model 2 and 5) and at second 210 (model 3 and 6).					
Variables	**(1) Degree coord 90**	**(2) Degree coord 150**	**(3) Degree coord 210**	**(4) Degree coord 90**	**(5) Degree coord 150**	**(6) Degree coord 210**
Outcome model						
Low uncertainty treatment	6.25	48.83	27.56	-0.39	37.50	28.66
	[-54.00 - 66.51]	[-45.19 - 142.84]	[-64.81 - 119.94]	[-61.06 - 60.28]	[-56.76 - 131.75]	[-59.58 - 116.89]
	(0.84)	(0.31)	(0.56)	(0.99)	(0.44)	(0.52)
Age (group average in years)				1.38	-1.89	-13.43
				[-10.82 - 13.58]	[-20.16 - 16.37]	[-30.60 - 3.75]
				(0.82)	(0.84)	(0.13)
Gender (group fraction females)				43.24	-125.82	-89.04
				[-91.60 - 178.08]	[-331.20 - 79.57]	[-271.57 - 93.48]
				(0.53)	(0.23)	(0.34)
Risk measure (group average)†				-16.18	-37.20	-57.44***
				[-44.48 - 12.12]	[-82.20 - 7.80]	[-98.56 - - 16.32]
				(0.26)	(0.11)	(0.01)
Expected round continuation (group average category)‡				7.61	-10.68	-36.33
				[-21.69 - 36.91]	[-58.22 - 36.87]	[-83.49 - 10.83]
				(0.61)	(0.66)	(0.13)
Constant	479.29***	332.76***	300.63***	470.14**	702.96**	1,152.11***
	[436.68 - 521.91]	[263.75 - 401.78]	[231.74 - 369.52]	[66.73 - 873.55]	[74.60 - 1,331.32]	[566.67 - 1,737.55]
	(0.00)	(0.00)	(0.00)	(0.02)	(0.03)	(0.00)

Table 3F.2. Continued. Six linear two-equation Hurdle models with degree of coordination (degree coord) at second 90, 150 and 210 as dependent variables.

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	**Degree coord 90**	**Degree coord 150**	**Degree coord 210**	**Degree coord 90**	**Degree coord 150**	**Degree coord 210**
Lower-limit selection model						
Low uncertainty treatment	0.00	-0.19	0.07	-0.23	-0.19	0.08
	[-1.15 - 1.15]	[-0.82 - 0.44]	[-0.49 - 0.62]	[-3.76 - 3.30]	[-0.85 - 0.47]	[-0.51 - 0.67]
	(1.00)	(0.56)	(0.81)	(0.90)	(0.57)	(0.79)
Age (group average in years)				-0.17	0.10	0.09
				[-0.83 - 0.50]	[-0.06 - 0.25]	[-0.04 - 0.23]
				(0.62)	(0.23)	(0.16)
Lower-limit selection model						
Gender (group fraction females)				-1.53	-0.58	-1.57**
				[-11.44 - 8.37]	[-2.11 - 0.95]	[-3.07 - -0.07]
				(0.76)	(0.46)	(0.04)
Risk measure (group average)†				-1.67	-0.23	-0.36**
				[-4.62 - 1.29]	[-0.55 - 0.08]	[-0.65 - -0.06]
				(0.27)	(0.14)	(0.02)
Expected round continuation (group average category)‡				-1.77	0.03	-0.03
				[-5.00 - 1.46]	[-0.26 - 0.33]	[-0.29 - 0.23]
				(0.28)	(0.83)	(0.82)
Constant	2.01***	1.11***	0.56***	26.89	0.13	1.19
	[1.20 - 2.82]	[0.65 - 1.57]	[0.17 - 0.94]	[-27.51 - 81.30]	[-4.67 - 4.93]	[-2.99 - 5.36]
	(0.00)	(0.00)	(0.00)	(0.33)	(0.96)	(0.58)
Sigma§	4.97***	5.26***	5.16***	4.95***	5.24***	5.07***
	[4.82 - 5.12]	[5.06 - 5.46]	[4.94 - 5.38]	[4.80 - 5.10]	[5.04 - 5.44]	[4.86 - 5.28]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	90	90	90	90	90	90

Note: Model 1 to 3 focus on the treatment effects. Model 4 to 6 include groups' average age, groups' fraction of females, groups' average risk measure and groups' average round continuation expectation as socio-economic controls.

Degree of coordination at 90, 150 and 210 seconds is defined as distance of the resource level to the threshold at 90, 150 and 210 seconds respectively. The degree of coordination equals 0, if groups failed to coordinate and reached the threshold at 400 resource units (lower limit). Any value greater than zero indicates successful coordination.

The lower-limit of the selection model is zero.

† **Risk measure** was elicited with the non-incentivised risk question developed by Dohmen et al. (2011). Participants rated their willingness to take on risk on an 11-point Likert scale between 0 (risk averse) to 10 (risk loving). ‡ **Expected round continuation** measures participants' expected continuation of the round beyond the certain minimum round length. Participants chose one of the following seven categories: 1. Immediate end, 2. Less than 30 seconds, 3. Another 30 seconds, 4. 30 to 60 seconds, 5. 60 to 90 seconds, 6. More than 90 seconds and 7. No expectation. Participants that answered "no expectation" (7) were coded as missing before calculating the group's mean category. § **Sigma** presents information about the estimated standard deviation of the error term in the outcome model.

Confidence intervals (at the 95% confidence level) are presented in square brackets. Standard errors are presented in brackets.

* p < 0.1; ** p < 0.05; *** p < 0.01

Appendix 3G: Exploratory cooperation analysis

Table 3G.1. Summary statistics of degree of overexploitation as measure for cooperation in treatment "High uncertainty" and "Low uncertainty" and the results of two-sample Mann–Whitney–Wilcoxon tests at 90, 150 and 210 seconds.

	High uncertainty (HU)		**Low uncertainty (LU)**		**Two-sample Mann–Whitney–Wilcoxon test (p-value)**
Degree of overexploitation	**N**	**Mean (SD)**	**N**	**Mean (SD)**	
at 90 seconds	45	122 (141)	45	114 (135)	0.95
at 150 seconds	45	271 (193)	45	260 (210)	0.73
at 210 seconds	45	347 (187)	45	326 (204)	0.74

Note: Degree of overexploitation is defined as 970 resource units minus the current resource level. Due to structural differences between treatments, the analysis focuses on resource levels below the MSY. Thus, the degree of overexploitation equals 0, if groups cooperate successfully and keep the resource above the level of maximum sustainable yield (minimum). The degree of overexploitation equals 570, if groups do not cooperate successfully and cause a collapse of the resource by reaching the threshold at 400 resource units (maximum).
N denotes the number of groups. Each group consists of four participants.
Standard deviation (SD) in brackets.
*** p<0.01, ** p<0.05, * p<0.1

Table 3G.2. Tobit regression models with the degree of overexploitation as measure for failed cooperation as dependent variable.

	Outcome variable: Groups' degree of overexploitation at 90 seconds (model 1 to 3), at 150 seconds (model 4 to 6) and at 210 seconds (model 7 to 9).								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Variables	**Overexploitation 90**	**Overexploitation 90**	**Overexploitation 90**	**Overexploitation 150**	**Overexploitation 150**	**Overexploitation 150**	**Overexploitation 210**	**Overexploitation 210**	**Overexploitation 210**
Low uncertainty treatment	5.54	14.90	194.35	-9.10	4.23	-172.85	-33.32	-5.26	141.60
	[-73.82 - 84.90]	[-64.97 - 94.77]	[-201.65 - 590.35]	[-130.31 - 112.12]	[-115.54 - 124.00]	[-745.57 - 399.87]	[-159.47 - 92.83]	[-126.23 - 115.71]	[-441.78 - 724.98]
	(0.89)	(0.71)	(0.33)	(0.88)	(0.94)	(0.55)	(0.60)	(0.93)	(0.63)
Age (group average in years)		-4.60	-4.49		-7.76	-7.67		-2.94	-3.05
		[-20.77 - 11.56]	[-20.58 - 11.60]		[-31.66 - 16.13]	[-31.51 - 16.17]		[-26.98 - 21.10]	[-27.10 - 20.99]
		(0.57)	(0.58)		(0.52)	(0.52)		(0.81)	(0.80)
Gender (group fraction females)		-49.19	-24.60		178.88	156.50		288.34**	305.36**
		[-226.34 - 127.96]	[-208.63 - 159.42]		[-91.81 - 449.56]	[-122.55 - 435.55]		[15.94 - 560.75]	[24.74 - 585.97]
		(0.58)	(0.79)		(0.19)	(0.27)		(0.04)	(0.03)
Risk measure (group average)†		39.66**	61.36**		57.78**	36.95		91.76***	109.50**
		[3.52 - 75.79]	[1.71 - 121.01]		[3.79 - 111.77]	[-47.91 - 121.82]		[36.07 - 147.45]	[20.34 - 198.66]
		(0.03)	(0.04)		(0.04)	(0.39)		(0.00)	(0.02)
Low uncertainty treatment x Risk measure			-34.11			34.07			-28.32
			[-107.72 - 39.50]			[-73.69 - 141.83]			[-138.39 - 81.74]
			(0.36)			(0.53)			(0.61)
Expected round continuation (group average category)‡		-1.54	-1.57		-0.91	-0.90		-8.33	-8.34
		[-38.17 - 35.09]	[-38.07 - 34.93]		[-56.85 - 55.02]	[-56.68 - 54.87]		[-65.27 - 48.61]	[-65.32 - 48.64]
		(0.93)	(0.93)		(0.97)	(0.97)		(0.77)	(0.77)
Constant	83.61***	26.36	-107.87	266.82***	51.03	172.86	388.40***	-178.03	-279.36
	[26.28 - 140.93]	[-500.85 - 553.58]	[-710.67 - 494.93]	[181.40 - 352.24]	[-738.07 - 840.13]	[-702.88 - 1,048.60]	[298.47 - 478.34]	[-974.25 - 618.19]	[-1,168.74 - 610.02]
	(0.00)	(0.92)	(0.72)	(0.00)	(0.90)	(0.70)	(0.00)	(0.66)	(0.53)

Table 3G.2. Continued. Tobit regression models with the degree of overexploitation as measure for failed cooperation as dependent variable.

Variables	(1) Overexploitation 90	(2) Overexploitation 90	(3) Overexploitation 90	(4) Overexploitation 150	(5) Overexploitation 150	(6) Overexploitation 150	(7) Overexploitation 210	(8) Overexploitation 210	(9) Overexploitation 210
Observations	90	90	90	90	90	90	90	90	90
left-censored	25	25	25	15	15	15	8	8	8
right-censored	2	2	2	14	14	14	25	25	25
Prob > chi2	0.89	0.27	0.3	0.88	0.33	0.41	0.6	0.02	0.04

Note: In Model 1 to 3 the dependent variable is groups' degree of overexploitation at 90 seconds, in Model 4 to 6 at 150 seconds and in model 7 to 9 at 210 seconds. Model 2, 5 and 8 include groups' average age, groups' fraction of females, groups' average risk measure and groups' average round continuation expectation as socio-economic controls. Model 3, 6 and 9 additionally include the interaction of the groups' average risk measure and treatment "Low uncertainty".

Degree of overexploitation is defined as 970 resource units minus the current resource level. Due to structural differences between treatments, the analysis focuses on resource levels below the MSY. Thus, the degree of overexploitation equals 0, if groups cooperate successfully and keep the resource above the level of maximum sustainable yield (lower limit). The degree of overexploitation equals 570, if groups do not cooperate successfully and cause a collapse of the resource by reaching the threshold at 400 resource units (upper limit).

† **Risk measure** was elicited with the non-incentivised risk question developed by Dohmen et al. (2011). Participants rated their willingness to take on risk on an 11-point Likert scale between 0 (risk averse) to 10 (risk loving). ‡ **Expected round continuation** measures participants' expected continuation of the round beyond the certain minimum round length. Participants chose one of the following seven categories: 1. Immediate end, 2. Less than 30 seconds, 3. Another 30 seconds, 4. 30 to 60 seconds, 5. 60 to 90 seconds, 6. More than 90 seconds and 7. No expectation. Participants that answered "no expectation" (7) were coded as missing before calculating the group's mean category.

Confidence intervals (at the 95% confidence level) are presented in square brackets.

P-values are presented in brackets.

* p < 0.1; ** p < 0.05; *** p < 0.01

Appendix 3H: Instructions of the experiment

English translation of the original German experimental instructions. German original is available upon request. Additional information and explanations of experimental processes are marked in italics. The first part of the instructions was printed and read aloud by the experimenter. Horizontal lines _____ mark the switch to the next screen/step of the experiment as programmed in the experimental software SoPHIE (Hendriks 2012).

Instructions of the CPR game

Welcome to the experimental laboratory

You are now taking part in an economics experiment. This experiment will take approximately 75 minutes in total.

Please comply with the following rules during the experiment: From this moment on any kind of communication is forbidden. If you have a question, please raise your hand and we will come to your cabin. It is mandatory to turn off your mobile phones and any other technical devices or set them on silent. Please place your mobile phone in the blue bag that is attached to the curtain rod outside of your cabin. The use of your technical equipment is not allowed during the entire experiment. Please take your cell phone out of the bag once you are called to the payout by your participation code at the end of the experiment. If you fail to comply with these rules or cause undue disruptions, we reserve the right to exclude you from the experiment and all payouts.

The instructions for this experiment consist of a printed and an on-screen part. At the beginning, we will read the instructions out loud to you. Please also read the on-screen instructions that you go through independently carefully. The decisions you and the other participants make during the experiment, determine your payout at the end of the experiment. Therefore, it is important that you read the instructions carefully.

If you have any questions or problems, please raise your hand from your cabin. Asking any questions does not mean that you are confused, but mostly that we have formulated something misleading. So please do not hesitate to ask us.

All data that we collect during this experiment is stored anonymously and treated with confidentiality.

The other participants will not be able to connect your anonymously made choices to your identity. Accordingly, the people who conduct this study will not be able to connect your decisions and answers to your identity. There are no correct or incorrect decisions in this experiment. You are asked to decide based upon your own personal preferences. Please do not discuss your decisions with any of the other participants after the experiment has ended.

After the experiment your will receive your payout in cash. To receive your payout we will call out your participation code that you got for your login at the beginning of the experiment. The payouts will be handed out individually, so that no other participant will see how much you earned as payout. Only one of the experimenters and you will know your payout.

Your earnings are referred to as points during the experiment, not Euro. All your earnings are calculated in points and your total sum of points will be converted to Euro at the end of the experiment. The exchange rate for the points to Euro conversion are presented in the instructions for the individual parts of the experiment. At the end of the entire experiment, we will inform you of your total payout.

Instructions

In this experiment, you manage a renewable resource in a group of four people. The resource develops dynamically over time and your task is to decide how many resource units you want to extract from the resource during a given time period. You can change the amount that you extract at any time during the experiment.

The resource units that you extract determine the number of points that you collect for your payout. The more resource units you extract the higher your number of points and thus your payout for the corresponding round.

One resource unit equals one point (1 resource unit = 1 point).

This experiment consists of two parts.

<u>The first part consists of two test rounds:</u>

The two test rounds, allowing you to familiarize yourself with the experimental set-up. Before the start of the first test round all participants are divided into groups of four. These groups will not change for the first two test rounds. Your decisions in these **two test rounds** do **not** have any impact on your payout.

<u>The second part consist of one payout-relevant rounds:</u>

Before the start of the payout-relevant round, all participants are assigned to new groups of four.

All the participants with whom you will be in a group in the payout-relevant round are different from the people with whom you were in a group in the two test rounds.

The points you have earned based on your decisions made in that payout-relevant round are converted into Euro and given to you in cash at the end of the experiment. The exchange rate for points in Euro is 100 points = **1.00 Euro.**

The decisions that you make in one of the two test rounds will not influence the development of the resource or any probabilities in the payout-relevant round. Each round of the three rounds in the first part is independent of the others.

End of printed instructions: Subjects switch over to the on-screen part of the instructions (dividing lines mark separate screens). Before the experimenter started the on-screen instructions, subjects were asked if they have any questions regarding the first part of the *instructions. If not, subjects continued with the on-screen instructions. In contrast to the printed instructions, the on-screen instructions were not read out aloud by the experimenter.*

Info screen (visible while subjects wait for the start of the on-screen instructions):

Please read the printed instructions carefully.

Once everyone is finished with the printed instructions, the on-screen instructions and thus, the on-screen part of the experiment will be started.

---

30 seconds time for subjects to read the text, then the experimenter read the text out aloud:

Explanation of Resource Development

The resource development in this experiment takes place in real time. The development and growth of the resource is variable during each round. The number of available resource units depends on:

a) the resource extraction of all persons in your group every second, and

b) the growth of the resource every second.

Every person in your group, including you, can extract between 0 and a maximum of 10 resource units from the resource every second. This means that the extracted sum of all four people in your group will be between 0 and a maximum of 40 resource units per second.

The resource's growth per second depends on the number of available resource units. At the beginning of each round, the resource is 2,000 units. The resource cannot grow higher than 2,000 resource units. At 2,000 units, the resource stops growing and the growth per second is 0.

When the resource level is below 2,000 units, the resource's growth per second initially increases. At 1,000 units, the resource growth is at its highest with 20 resource units per second. Thus, if there are between 2,000 and 1,000 units of the resource, the resource grows between 0 and 20 resource units per second.

The resource's growth per second will decrease as soon as the resource drops below 1,000 units. As soon as 0 resource units are reached, the growth per second is also at 0 units. Thus, if there are between 1,000 and 0 units of resource, the resource grows between 20 and 0 resource units per second. As soon as the resource has reached 0 units, it remains permanently at 0 until the end of the round and you cannot extract any resource units anymore.

---

30 seconds time for subjects to read the text, then the experimenter read the text out aloud:

Explanation of Resource Development

In the table below you can see possible values of the resource in the left column, i.e. possible amounts of available resource units between 2,000 and 0 units. In the right column, the corresponding amount of growth per second of the resource between 0 and 20 units is listed.

Please note that the quantities of existing resource units shown in the table are only examples. The resource can take all possible values between 2,000 and 0. Correspondingly, the resource's growth per second also develops continuously between 0 and 20.

Level of the resource	Growth of the resource (in resource units per second)
2000	0
1800	7.2
1600	12.8
1400	16.8
1200	19.2
1000	20
800	19.2
600	16.8
400	12.8
200	7.2
0	0

---

30 seconds time for subjects to read the text, then the experimenter read the text out aloud:

Example of Resource Development

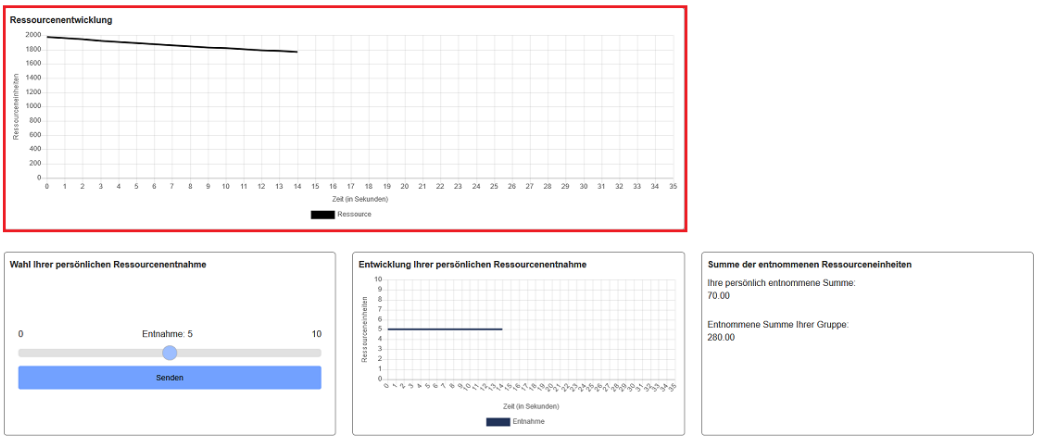
At a level of 1,000 resource units, the growth of the resource is at 20 resource units per second. If 40 units are extracted in this second, then 1,000 +20 -40 = 980 units are available in the following second.

During each round, you can monitor the development of your group's resource per second on the screen. You can also observe the development per second of your personal extraction and the development of the extraction of your group. During each round, you can change your own resource extraction as often as you like.

End of the part of the instructions that were read out aloud by the experimenter. The experimenter asked once again if there were any questions before subjects continued with the experiment.

Explanation of Resource Development

The following figure (German original) presents a screenshot of the screen during the experiment. The box **"development of the resource (Ressourcenentwicklung)"** in the top left corner of the screen shows the development of the resource.



The vertical axis with the caption “Resource units (Ressourceneinheiten)” presents the level of the resource in resource units from 0 to 2,000. The time is displayed in seconds on the horizontal axis. From 30 seconds onwards, the numbers that label the seconds on the x-axis will change over time.

Choice of your personal resource extraction level with the slider

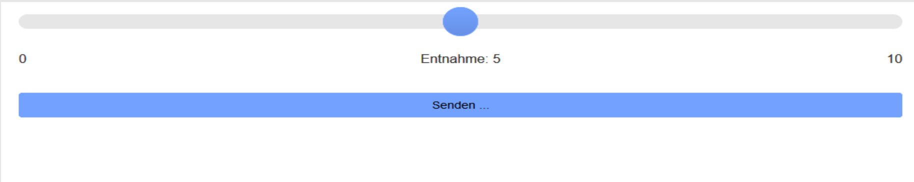
Before the start of each round, you can choose between 0 and 10 resource units as your resource extraction per second via a slider:

You have to click on the grey bar above the “Send (Absenden)” button to make a blue dot appear. You can move this blue dot to the left or right to adjust the level of your personal resource extraction to 0,1,2,3,4,5,6,7,8,9 or 10 extracted resource units per second. The text under the grey bar clearly states the exact number of resource units that are extracted when the corresponding spot on the grey bar is chosen via positioning the blue dot on it. From the first second onwards, your chosen resource extraction and the chosen extraction by the other group members will be executed.

As soon as you have chosen your preferred level of resource extraction, you have to confirm your choice by clicking “Send”. The round of the extraction game will start as soon as all members of your group clicked on “Send”.

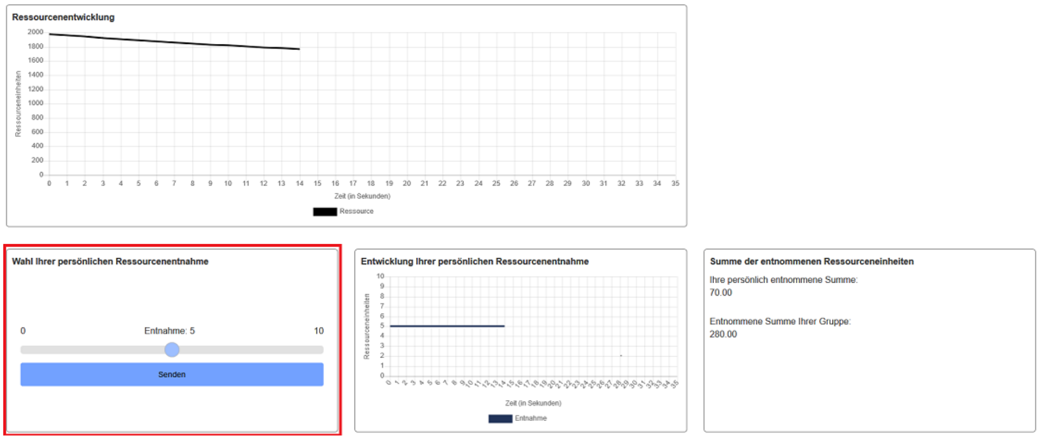
Example

“Extraction (Entnahme): 5”, means as soon as you clicked “Send” and the round starts, 5 resource units are extracted every second.



Choice of your personal resource extraction level with the slider

At the start of each round, you will see your personally chosen starting level of your extraction marked on the slider in the box in the bottom corner on the left:



You can change your chosen level of resource extraction from 0 to 10 anytime during the round by moving the blue dot on the slider to the left or the right. Your starting level of extraction is continuously executed until you make an active change.

Please be aware that you have to click “send” to confirm your choice of resource extraction and to implement the new extraction level.

Development of your personal resource extraction

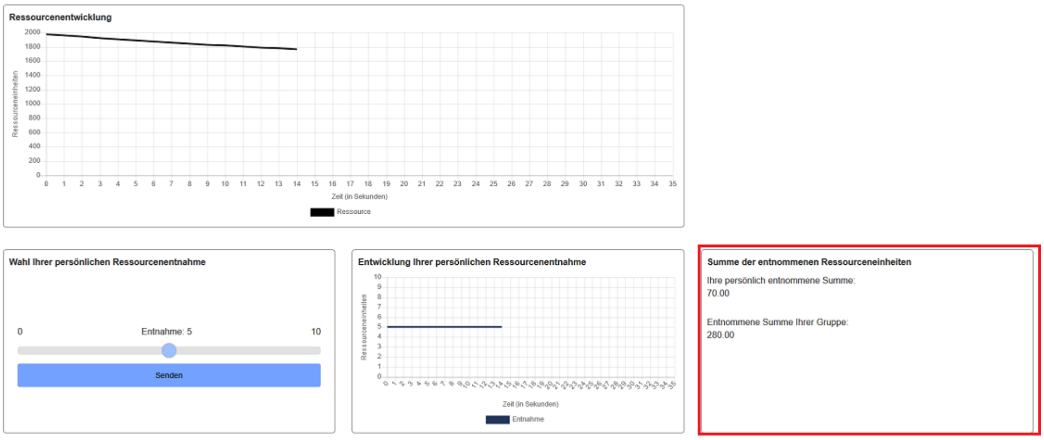
The graph “**Development of your personal resource extraction**” (box in the middle of the bottom line) shows you if your chosen extraction level is implemented:

If you move the blue dot to the right of the slider to choose a higher resource extraction per second than before and klick on “send”, the graph increases. If you move the blue dot to the left to choose a lower resource extraction per second than before and klick on “send”, the graph decreases.

Please click on “send” again, in case that you do not observe a change in the graph after you changed your level of resource extraction per second with the slider.

Presentation of the sum of resource extraction

You get an overview about the sum of resource units that you yourself and all persons in your group have extracted in total up to a given point in time by looking at the information shown in the box **“Sum of extracted resource units** (**Summe der entnommenen Ressourceneinheiten”** (bottom right corner).



The calculation of your sum of extracted resource units takes place per second. For technical reasons, however, there may be jumps and delays of a few seconds in the update of the display of the totals of the resource units extracted.

Your own individual sum (Ihre persönlich entnommene Summe) shows how many resource units have been extracted by you personally up to the given point in time.

Extracted sum of your group (Entnommene Summe Ihrer Gruppe) shows how many resource units all four persons in your group have extracted in total up to a given point in time.

Example

If **“Extracted sum of your group”** shows 280 resource unit and **“Your own individual sum”** shows 70 resource units, you know that up to that point in time, you extracted 70 resource units for yourself and the other three persons from your group extracted a total of 210 resource units.

Control Questions

Please answer the following questions to clarify your understanding of the instructions before you the start the two test rounds.

Control Questions *(1)*

At which number of resource units is the growth per second of the resource with 20 units per second the highest?

Number of resource units * [ ]

Answer: 1,000 resource units - explanation only shown if incorrect answer

Explanation: The growth of the resource every second depends on the number of available resource units. At 1,000 units, the resource growth is highest with 20 units growing every second.

Control Questions *(2)*

Please choose the correct statement:

If you choose 10 as the level of resource extraction,

c) you extract 10 resource units per second. Your choice is continuously implemented each second until you actively change it.

d) you once extract 10 resource units. You have to confirm your choice each second.

Answer: a correct – explanation only shown if incorrect answer

Explanation: The selection of your personal resource extraction remains the same until you actively change it. This means that the number of resource units that corresponds to your selected resource extraction is extracted from the resource every second. If you want to change your selection, you have to move the dot on the slider according to your wish and confirm your new selection by clicking on “Send”. Afterward the chosen level of extraction is executed automatically.

Control Questions *(3)*

How many units does the resource grow every second when there are 2,000 resource units?

Enter the answer: 0

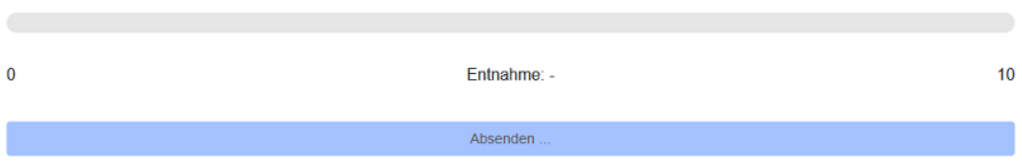
Explanation, only shown if incorrect answer: The resource's growth per second depends on the number of available resource units. At the beginning of each round, the resource is 2,000 units. The resource cannot grow beyond these 2,000 units. In other words, it stops growing at 2,000 units and the growth per second is 0.

First test round

You and the others in your group have **90 seconds** per test round to get familiar with the resource development and to try out different levels of resource extraction. Your decisions in the two test rounds have **no impact on your payout.**

Please click on "Continue" to choose the starting level of your personal extraction for the first test round As soon as all group members, including you, confirmed their choice by clicking on "send" the resource development of the first test round will start.

Please click on the grey bar to make the blue dot appear and choose your personal extraction level:



The first test round started automatically, once all group members chose their initial resource extraction level.

Once the first test round was finished, the second test round started.

Second Test Round

You have finished the first test round. For the following second test round, you will remain in the same group, which means that the other three persons in your group will be the same as before.

The resource development follows the same rules as in the first test round and will start at the level of 2,000 resource units. As soon as everyone in the group, including you, confirmed their first extraction choice by clicking on “send”. The resource extraction chosen by everyone in your group, including you, is carried out again from the first second.

Please remember:

- Your decisions in the two test rounds have no impact on your payout.
- The test round lasts **90** seconds.

Please click on “Continue” to choose the starting level of your personal extraction for the second test round. As soon as all persons in your group, including you, make your choice and confirm it by clicking on “send”, the resource development of the second test round will start.

Next step, choice of the starting extraction level for second test round (see description of first text round above). Once all group members had chosen their initial extraction level, all groups played a second test round.

Payout-relevant round

You are finished with the two test rounds now. In the following payout-relevant round, you will collect points for your payout.

Before the start of the round, that is relevant for your payout, all participants, who take part in the current experiment, are rematched in new groups of four.

The participants in your new group were **not** in your group in the test rounds. Thus, it is possible that the participants in your new group will behave differently than the participants in your old group.

Duration of the payout-relevant round

The two payout-relevant rounds are **longer** than the two test rounds. The exact end of the payout-relevant round is unknown.

The payout-relevant round lasts at least **240 seconds**. During this time, you can extract resource units and collect points for yourself.

After these 240 seconds, every 10 seconds it is randomly determined whether the round continues for another 10 seconds or ends. There is a 90 percent probability that the round will continue. There is a 10 percent probability that the round will end. That means, in 9 of 10 cases the round will continue for another 10 seconds. As long as the round continues, you can extract resource units as usual.

Examples

If you are at 240 seconds, the probability that the round will continue is 90 percent.

If you are at 300 seconds, the probability that the round will continue is also 90 percent.

There were two versions of the CPR instructions depending on the treatment “Low uncertainty” and “High uncertainty”. The instructions for the two treatments were automatically shown on-screen. Subjects only got to read the relevant instructions for their assigned treatment.

Treatment “High uncertainty”:

Resource development in the payout-relevant round

IMPORTANT: The conditions of the resource development have changed in one aspect.

In other words, there is **one** difference in the development of the resource in the payout-relevant round compared to the two test rounds.

In the payout-relevant round, reaching a certain level of the resource, the so-called threshold, causes a sudden end of the resource development.

This means that as soon as the resource level reaches the threshold, which means the resource level is **smaller or equal** to the threshold, the resource immediately drops down to 0 resource units. Additionally, once the resource is at 0 resource units, it stops growing and the growth rate is changed to 0 resource units per second. In this case, the growth of the resource drops abruptly to a growth per second of 0. Once the threshold is reached, **no further points** can be collected until the time of the round is up.

The exact value of the threshold is unknown. The three other persons in your group receive the same information as you.

As long as the available amount of resource units and thus, the value of the resource is above the threshold, the conditions for the growth per second of the resource are the same as in the test rounds.

Treatment "Low uncertainty":

Resource development in the payout-relevant round

IMPORTANT: The conditions of the resource development have changed in one aspect.

In other words, there is **one** difference in the development of the resource in the payout-relevant round compared to the two test rounds.

In the payout-relevant round, reaching a certain level of the resource, the so-called threshold, causes a sudden end of the resource development.

This means that as soon as the resource level reaches the threshold, which means the resource level is **smaller or equal** to the threshold, the resource immediately drops down to 0 resource units. Additionally, once the resource is at 0 resource units, it stops growing and the growth rate is changed to 0 resource units per second. In this case, the growth of the resource drops abruptly to a growth per second of 0. Once the threshold is reached, **no further points** can be collected until the time of the round is up.

The threshold is somewhere in between 700 and 200 resource units. The exact value of the threshold is unknown. The three other persons in your group receive the same information as you.

As long as the available amount of resource units and thus, the value of the resource is above the threshold, the conditions for the growth per second of the resource are the same as in the test rounds.

Control questions

Please answer a few questions to clarify your understanding of the instructions before you start the payout-relevant round and to help understand for the differences between the test rounds and the payout-relevant round.

Control questions *(4)*

How long does each of the two payout-relevant rounds last?

a) The rounds lasts a minimum of 240 seconds and after these first 240 seconds, it is randomly determined every 10 seconds whether the round continues. The probability for continuing is 90 percent.
b) The rounds stop immediately after 240 seconds.

Answer a correct – explanation only shown if incorrect answer

Explanation: The payout round is longer than the two test rounds. In the payout-relevant round, you have at least 240 seconds to extract units and thereby collect points for yourself. After these 240 seconds, the round continues for an unknown time.

Every 10 seconds it is randomly determined whether the round continues for another 10 seconds or ends. There is a 90 percent probability that the round will continue. There is a 10 percent probability that the round will end.

Control questions *(5)*

Under the assumption that person A and person B are in the same group in the two test rounds.

Is it possible that person A and person B are also in the same group in the payout-relevant round?

a) Yes, if person A and B were in the same group during the test rounds, they will also be in the same group in the payout-relevant round.
b) No, if person A and person B were in the same group in the test rounds, they are definitely not in the same group in the payout-relevant round.

Answer b – explanation, only shown if incorrect answer:

Explanation: Before the start of the round, which is relevant for the payout, all participants in the experiment are divided into new groups of four. The people with whom you are in a group in the payout round will be different from the people with whom you were in a group in the test rounds.

Control questions *(6)*

What happens, when the resource level reaches the threshold?

a) Nothing further, the growth of the resource and the level of the resource drops to 0, but if the group reduces their resource extraction to 0, the resource starts growing again and all participants can continue to extract units.

b) The development of the resource comes to an irreversible stop. The growth of the resource and the level of the resource drop to 0. This means that no more units can be extracted until the end of the round.

Answer: b correct – explanation for incorrect answer

Explanation:

As soon as the resource level reaches the threshold, which means the resource level is **smaller or equal** to threshold, the resource immediately drops down to 0 resource units. In this case, the growth of the resource drops abruptly to a growth per second of 0. Once the threshold is reached, no further points can be collected until the time of the round is up.

Control questions *(7)*

How many resource units do you extract per second if the level of the resource and the growth of the resource are at 0?

a. As many resource units as I chose via the slider.

b. I can no longer collect any resource units if the resource has dropped to 0.

Answer b correct - explanation for incorrect answer

Explanation: Once the threshold is reached, the resource growth remains at 0 until the end of the round and the resource remains at 0 units. Changes of your extraction level do not have any impact on the development of the resource after reaching the threshold.

This means that neither you nor any other person from your group can extract further resource units until the end of the round if the threshold of the resource has been reached. In this case, no more points can be collected in this round.

Start of the payout-relevant round

Please choose your personal level of resource extraction for the payout-relevant round in the following step.

As soon as all persons in your group, including yourself, made their choices and confirm them by clicking on “send”, the resource development of the first payout-relevant round will start.

For your payout, the sum of resource units from the following payout-relevant round displayed under “Your own individual sum” is converted into Euros.

Choice of the starting extraction level and start of the payout-relevant round.

Pause of the round after 58 seconds in both treatments (unknown to subjects). During the pause the presented information on the screen switched from the CPR game screen to the following explanation on screen. Subjects in both treatments “High” and “Low uncertainty” received the same information during the pause.

This interruption is part of the experiment.

During the pause, the resource development has stopped, and nobody can extract resource units. The resource extraction will not continue until all of the people in your group have read the following text.

Information on the threshold level of the resource

You now get more information about the threshold level of the resource:

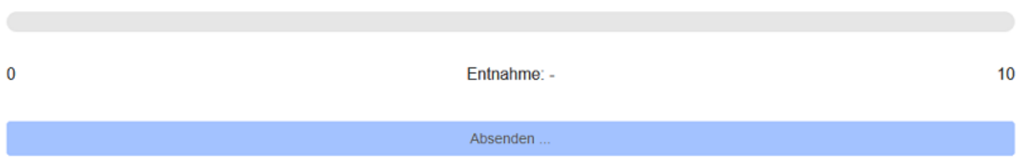
The threshold level is at 400 resource units.

When the resource level reaches the threshold of 400 resource units, which means the resource level is **smaller or equal to 400 resource units**, the resource immediately drops down to 0 resource units. In this case, the growth per second also immediately drops down to 0.

As long as the available amount of resource units is above the threshold and thus, the more than 400 resource units are available, the conditions for the growth per second of the resource are the same as in the test rounds. The resource grows every second, until the threshold is reached. Once the threshold is reached, the resource does not grow anymore and stays at the resource level of 0 until the end of the round.

Please choose the extraction level with which you would like to continue your personal resource extraction after the interruption:

As soon as all persons in your groups decide on an extraction level by using the slider below and confirm their choice by clicking “send” the round will continue.



Post-experimental questionnaire

Once the CPR game was finished, all subjects answered a post-experimental questionnaire. In the first part of the questionnaire, we asked questions regarding subjects understanding of the experiment and experimental procedures. In the second part, subjects were asked questions about their socio-economic background.

Questionnaire

To finalize the experiment, we kindly ask you to answer a few questions. In the meantime, our team is preparing your payout.

It is mandatory to answer all questions marked with an asterisk *.

All of your answers are anonymised and treated confidentially. Your answers to the following questions have no effect on the amount of your payout.

(Q1) – risk measure

How do you see yourself: are you generally a person who is prepared to take risks, or do you try to avoid taking risks?

Please tick a box on the scale, where the value 0 means: “not at all willing to take risks” and the value 10 means: “very willing to take risks”.

You can use the values in between to grade your assessment.

not at all willing to take risks										very willing to take risks
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
○	○	○	○	○	○	○	○	○	○	○

(Q2)

What were your expectations regarding the length of the **payout-relevant** round?

I expected that after the first 240 seconds the rounds *

a. would end immediately.
b. would end after less than 30 seconds.
c. would end after another 30 seconds.
d. would end after 30 to 60 more seconds.
e. would end after 60 to 90 seconds.
f. would end after more than 90 seconds.
g. I had no expectation regarding the end of the rounds.

(Q3)

Assume that you knew exactly when the payout-relevant rounds would have ended. Would you have chosen a different level of personal resource extraction?

- Yes
- No
- I don’t know.

(Q4), only if yes to Q3

To which level between 0 and 10 would you have changed your personal resource extraction?*

(integer numbers between 0 and 10)

(Q5), only if yes to Q3

How many seconds before the end of the payout-relevant round would you have changed your resource extraction per second? *(free text answer)**

(Q6)

Did the two test rounds help you to understand the development of the resource and your task during the experiment?*

a. No, I already understood the experiment by reading the instructions. The test rounds were unnecessary.
b. Yes, the test rounds were a good help to familiarize oneself with the development of the resource and the effects of resource extraction.
c. No, the test rounds did not help me to understand the experiment better.

(Q7)

Have the experiences you made in the two test rounds influenced your behavior in the payout-relevant round?

- Yes
- No
- I don't know.

(Q8)

If yes: In what ways have the experiences you made in the two test rounds influenced your behavior in the payout-relevant round? *(free text answer)*

(Q9)

At the beginning of the payout-relevant round, did you already have a guess about the level of the threshold?

- Yes
- No

(Q10), only if yes to Q9

At which resource level did you suspect the threshold?

(Enter number between 0 and 2,000)

To what extent do you agree with the following statements?

Please tick a box on the scale from 1 to 5, the value 1 means: “strongly reject” and the value 5 means: “strongly approve”.

You can use the values in between to grade your approval respectively.

1. The instructions of this experiment were well written and easy to understand. *(Q11)*
2. I understood the development of the resource in a good way. *(Q12)*
3. It was important to me to prevent the resource from reaching its threshold. *(Q13)*
4. It was easy to adjust my personal extraction level via the slider during the round. *(Q14)*
5. It was important to me to distribute the resource fairly among all persons in the group. *(Q15)*

(Q16)

Were there moments during the experiment when you felt stressed?*

- Yes
- No

(Q17)

Do you have any <u>further</u> comments on the experiment?

You can also explain here why you made certain decisions during the experiment.

(free text answer)

Please answer the following questions.

(S1) Which gender are you?* *(female)*

- female
- male
- divers

(S2) How old are you?* *(insert age in years)*

(S3)

Have you participated in one or more economic or psychological experiments before this experiment?

- Yes
- No

(S4), only if yes to S3

How many times have you participated in economic or psychological experiments?

(open number field)

(S5)

Are you a student?

- Yes
- No

(S6), only if yes to S5

Where do you study?

- University of Hamburg
- Hamburg University of Technology (TUHH)
- University of Applied Sciences (HAW)
- Other: *(free text field)*

(S7)

Which field of study are you studying (predominantly)?

a. Humanities *(Humanwissenschaften)*
b. Cultural studies *(Geisteswissenschaften)*
c. Social sciences (not economics)
d. Economics
e. Law
f. Natural Science

g. Engineering

h. Medicine

i. Other *(free text)*

(S8)

What type of degree is it?

a. Bachelor

b. Master

c. PhD

d. State examination

e. Other

(S9)

What is your monthly income (including subsidies from your parents, student grants, salaries, scholarships)?

a. 0-300 Euro

b. 301-600 Euro

c. 601-900 Euro

d. 901-1,200 Euro

e. More than 1,200 Euro

f. No answer

Next all subjects were informed about their individual payout:

Your personal payout for the experiment

Your personal earnings are …. Points

Converted into EURO …. €

Total payout …. € (Rounded to full 0.10 Euro)

Thank you for participating in this experiment.

Please stay in your cabin and wait a moment; we are currently preparing your payout.

We will call you for the payout based on the first letter of your participation code.

Appendix 3I: Pre-registration of data analysis

Text of the submitted pre-registration. The original document provided by https://aspredicted.org is available upon request. In the pre-registration, we speak of "Range treatment" which is the "Low uncertainty" treatment and "Control treatment" which is the "High uncertainty" treatment.

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

Research Questions:

1. Does knowledge of a threshold range ("Range" treatment) decrease the level of cooperation amongst group members in comparison to groups that experience a higher level of threshold ambiguity ("Control" treatment)?

2. Does the effect of a certain threshold warning differ depending on the level of uncertainty about the threshold location that resource users were exposed to before?

Hypotheses:

Based on two different strands of the literature, we formulate two different alternative hypotheses for RQ1:

H1a: Groups in the "Range" treatment cooperate less (i.e. higher level of extraction) than groups in "Control". (Lower uncertainty leads to less cooperation.)

H1b: Groups in the "Range" treatment cooperate more (i.e. lower level of extraction) than groups in "Control". (Lower uncertainty leads to more cooperation.)

RQ2 is exploratory: We are not aware of any literature that analyses the effect that an early warning in form of range knowledge has on the effect of certain threshold knowledge. We formulate the following hypothesis:

H2: Knowledge of a certain threshold value affects coordination differently depending on the level of prior uncertainty regarding resource dynamics, i.e. threshold knowledge.

Potentially, the effect of certain threshold knowledge may differ depending on prior uncertainty regarding resource dynamics because:

a) Different levels of uncertainty may result in different resource levels at the point when certain threshold knowledge is introduced (see H1). Receiving the knowledge of the certain threshold value at a relatively high resource level may on the one hand indicate that there is sufficient scope to exploit the resource, which may eventually lead to a failure of coordination. On the other hand, it may indicate that there is sufficient time to learn to coordinate before getting into the proximity of the threshold value, which may enhance chances of successful coordination.

b) Group members may perceive the threshold information differently depending on their experience within their group, in particular the degree of past cooperation. Groups that cooperated successfully may have created sufficient trust and mutual expectations of the willingness to coordinate. They are thus more likely to coordinate above the threshold in comparison to groups that did not manage to cooperate in the past.

Our design only allows to measure the compound effect of prior uncertainty on threshold knowledge. As such we cannot disentangle the different mechanism that may be present.

3) Describe the key dependent variable(s) specifying how they will be measured.

Subjects in groups of 4 extract units from a resource during 1 payout-relevant round. The CPR has 2 parts: 1st the pre-break and 2nd the post-break part. The game lasts for 240 seconds with certainty. Beyond this certain minimum round length, we implement a random continuation rule.

To RQ1: The distance of the resource level to the level of maximum sustainable yield (MSY) at the break-time is taken as proxy for the level of cooperation in the group. MSY is at exactly 1,000 units. Keeping the resource at the MSY is socially optimal and a sign for efficient cooperation since the regrowth of the resource is highest. Since it is difficult for subjects to judge the exact level of the resource when playing the game, resource levels between 1,030 and 970 count as levels of cooperation. Groups that keep the resource at levels above or below MSY do not cooperate efficiently. Given the design of the 2 treatments, subjects in “Control” cannot be sure that the critical threshold is below the MSY. Hence, cautious resource extraction and keeping the resource above the MSY is rational. Given this structural difference, we focus on the failure of efficient cooperation below the MSY (970). 1. We measure the failure of cooperation as a binary variable (1, if the level of the resource is below MSY, and 0, otherwise). 2. We take the distance of the current resource level to the MSY as measure of the degree of failure in cooperation (continuous variable, correct to 2 decimal places). All groups that kept the resource above 970 are coded as cooperative.

To RQ2: We analyse the level of the resource at the defined end of the round (210 seconds to control for an endgame effect) and at two times in between (90 and 150 seconds). We assume that groups extract the resource at higher levels once they get closer to the known minimum round length of 240. We measure the failure of coordination with two outcome variables: 1. The binary variable “crash of the resource” which is 1, if the resource crashed prior to 210 seconds (the group failed to coordinate), and 0, otherwise. 2. The continuous variable “distance to threshold” (correct to 2 decimal places) measures the degree of coordination of groups that do not crash the resource. It is defined by the distance of the given resource level to the threshold.

4) How many and which conditions will participants be assigned to?

The experiment is a (quasi) continuous-time common-pool resource game. Participants are randomly assigned to either one of two treatments that differ in the degree of uncertainty about the threshold location in the CPR game. The exact threshold location or the underlying probability distribution of the threshold value is unknown to participants in both treatments.

Participants in treatment “Control” experience a high level of uncertainty about the exact threshold location and merely know of its existence. Whereas participants in treatment “Range” experience a lower level of uncertainty since they receive an imprecise, early warning in form of knowledge about the upper and lower limit of the threshold. In both treatments, participants get certain knowledge about the threshold location during the unannounced break in the game.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

Analysis focuses on group outcomes. We use Fisher’s exact tests and Wilcoxon Ranksum/Mann-Whitney U Tests to test for structural differences in groups’ averages between treatments with respect to subjects’ understanding of the experiment and their (socio-economic) background. If we find significant differences between treatments, we will control for them in the regression models.

Analysis RQ1: We use the Fisher’s exact tests to analyse the difference in the proportion of groups that cooperated in the first part (until the break) and the Wilcoxon Ranksum/Mann-Whitney U Test as a non-parametric test to test for differences in the groups’ distances to the MSY at the time of the break between treatments. Further, we analyse the probability of cooperation by running a probit/logit model and analyse the level of cooperation by running a Tobit model on the distance of the resource level to the MSY at the time of the break. In the Tobit model, the distance is censored from below at 0 (0 = cooperation) and also censored from above at the maximum distance to MSY that is achievable in 58 seconds (time span until the break).

Analysis RQ2: We use Fisher’s exact tests to analyse differences in the proportion of groups that cause a crash of the resource and the Wilcoxon Ranksum/Mann-Whitney U Test as a non-parametric test to test for differences in the groups’ distances to the threshold at the times of interest (90, 150 and 210 seconds). Further, we estimate three two-equation Hurdle models to assess the level of coordination at 90, 150 and 210 seconds: 1) a binomial probit model of the binary variable crash and 2) a linear regression model of the distance of the resource level to the threshold.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We exclude groups if one or more of the participants' computers loose connection during the data elicitation in one of the two test rounds or the CPR round.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

Due to the programming, we can run the experiment with either 16 or 24 subjects per session. We aim to run 15 sessions with 24 subjects each to collect 45 group observations per treatment (180 participants per treatment) which results in 360 participants in total. We will collect even numbers of observations per treatment. The total number of sessions increases if we have to run a number of sessions with 16 participants instead of the preferred number of 24 participants. We will stop running sessions prior to reaching our aimed observation numbers if the tentative budget for the experiment is exceeded.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Exploratory analysis: The time trends of resource development and groups' extraction choices will be analysed graphically to check for specific behavioural patterns in groups depending on the treatment. We want to assess if the differences between treatments change over time. For that reason, we will also see if we can find a significant difference in pattern of resource extraction and the timing of the crash between the two treatments.

Secondary hypothesis: We elicit subjects' risk preferences with the risk question based on Dohmen et al. 2011 and use the group averages as control variable in additional regression models. We assume that groups with more risk-liking individuals are more likely to extract the resource at high rates and thus, drive the resource to lower levels.

Further, we will do exploratory analysis to see whether experience in test rounds has an effect on the behaviour in the main round.

Chapter 4 False and missed alarms in seasonal forecasts affect individual adaptation choices[13]

Katharina Hembach-Stunden[a], Tobias Vorlaufer[a], Stefanie Engel[a]

[a] School of Business Administration and Economics and Institute for Environmental Systems Research (IUSF), Osnabrueck University, Germany.

Abstract: Climate change increases the risk of extreme weather events. Seasonal forecasts and early warning signals can warn the public of upcoming climate extremes. Being confronted with too many inaccurate forecasts may undermine individuals' responsiveness to forecasts in the long run. Using an online experiment, we assess how the exposure to *false alarm-prone* and *missed alarm-prone forecast systems* influences individuals' adaptation investments. We show that experiencing *false alarms* more frequently decreases individuals' investments if a warning is issued (the "cry-wolf effect"), but does not influence adaptation in the absence of a warning. A history of more frequent *missed alarms* increase individuals' adaptation investments in both cases, if no warning, but also if a warning is issued. However, these effects are relatively small compared to the effect of the forecasted probabilities. Individuals still react to increasing forecasted probabilities of upcoming extremes even if they previously experienced false or missed alarms more frequently.

Keywords: decisions under uncertainty, economic online experiment, extreme weather events, forecast design, early warning signals

[13] A version of this chapter was submitted to the journal Climatic Change in mid-February 2022 and is currently under review.

4.1 Introduction

Climate change globally increases the frequency and intensity of weather and climate extremes such as heatwaves, droughts and heavy precipitation (IPCC 2019). Individual behaviour that is solely guided by past experiences runs the risk of severely underestimating the need to adapt to these new conditions. In this context, seasonal forecasts and early warning signals are increasingly recognised as important guides for government, private sector and households' adaptation actions (Bruno Soares et al. 2018; Knudson and Guido 2019; Webber 2019). Forecasts and warnings can be especially useful to inform adaptation behaviour in the context of recurring decisions (e.g. cropping choices in agriculture) or temporary behavioural responses (e.g. in response to hurricane or flood warnings).

However, seasonal forecasts, as well as early warning signals, are often highly uncertain and inaccurate, which poses a challenge for both the communication of their predictions and their use (Zommers 2012; Taylor et al. 2015, 2018; National Institute of Water and Atmospheric Research (NIWA) 2016). Individuals who base their adaptation decisions on possible inaccurate warnings face two types of potential errors (Losee and Joslyn 2018): Firstly, they may experience a *false alarm*, where a warning is issued but an extreme event does not occur. Decision makers may comply with the warning and adapt their behaviour, but then experience normal conditions. On the other hand, individuals may experience a *missed alarm* where no warning is issued but extreme conditions strike. In this case, decision makers might rely on the forecast and decide against adaptation, but experience losses from extremes they were unprepared for (Losee and Joslyn 2018).

Inaccurate forecasts may also have long-term effects, eroding decision makers' trust in the system, leading them to ignore future forecasts or warnings. Experiencing false alarms more frequently in the past may decrease individuals' responsiveness, i.e. adaptation, if they receive warnings (also known as the *cry-wolf effect*) (LeClerc and Joslyn 2015). Similarly, experiencing missed alarms more frequently could lead to adaptation investments even if no warning is issued. If this is the case, policy makers and agencies ideally consider the undermining effect that false or missed alarm experiences have on the responsiveness to future forecasts; in particular deciding when to issue a warning or when designing a forecast/warning system. If not, they could risk limiting the overall contribution of a forecast systems to climate change adaptation.

We report the results of an incentivised online experiment to study what effect experiencing more frequent false and missed alarm has on individual responses to seasonal forecasts. To our knowledge, we are the first to conduct a systematic, experimental study of these effects. Over ten experimental rounds, individuals received a seasonal forecast with an unknown level of accuracy and decided how much to invest in individual adaptation. We exogenously varied the accuracy of the forecast system and thereby controlled individuals' false and missed alarm experiences. Most previous studies regarding forecasts or extreme event warnings are observational and rely on self-reported data, making it difficult to control for individuals' prior experiences and confounding factors (Ripberger et al. 2015; Taylor et al. 2015; Trainor et al. 2015; Lim et al. 2019). Moreover, previous evidence regarding the consequences of false and missed alarms experiences for individuals' responses to forecasts is inconclusive (Trainor et al. 2015; Lim et al. 2019). Some studies find that a high rate of prior false alarms reduced individuals' compliance with extreme weather warnings, leading individuals in some studies to be less likely to seek shelter in response to tornado warnings (Simmons and Sutter 2009; LeClerc and Joslyn 2015; Trainor et al. 2015). However, other observational studies of behavioural responses to hurricanes (Dow and Cutter 1998) and tornadoes warnings (Schultz et al. 2010; Lim et al. 2019) do not find clear evidence of a *cry-wolf effect*.

Experimental studies examining individuals' responses to warning systems are mostly conducted in the context of automated machine alerts, such as life support systems in hospitals, that require fast reactions (Manzey et al. 2014; Chancey et al. 2015, 2017; Wiczorek and Meyer 2016). These studies found that experiencing multiple false alarms reduces the responsiveness to warnings (i.e. the *cry-wolf effect*), whereas experiencing multiple missed alarms reduces the responsiveness to no-warnings (Manzey et al. 2014; Chancey et al. 2015, 2017). We define these as the two *main effects* of forecast inaccuracy. It is furthermore conceivable that multiple missed and false alarms affect how individuals react to the opposite signal, due to a general decrease in trust in the forecast system. Multiple false alarms may affect how individuals respond in the no-warning case, and multiple missed alarms may affect the response in the warning case. We define these two effects as the *cross-effects* of inaccurate forecast systems. Empirical evidence concerning these *cross-effects* is more mixed. A few previous studies find evidence of negative cross-effects of false (Wiczorek and Meyer 2016) and missed alarms (LeClerc and Joslyn 2015; Ripberger et al. 2015), whereas Manzey et al. (2014) find no evidence for such cross-effects. Overall, we believe that caution is warranted when generalising findings from automated machine alerts to another behavioural domain. Machine alerts require immediate attention, so it is an intuitive decision made within seconds and often not based on conscious deliberation (e.g. alarms of life support machines in hospitals) (Kahneman 2011). In contrast, responses to seasonal forecasts or extreme weather warnings are typically slower and more deliberate (ibid.).

4.2 Material and methods

4.2.1 Experimental design

In our experiment, participants were confronted with the decision to protect themselves against losses caused by an upcoming season with an extreme climate. Participants received a fixed payment of £2 and a variable bonus between £0 and £5 depending on both chance and their decision making during the experiment. If participants had not invested in adaptation and experienced an extreme season, they lost their full £5 bonus. During the experiment, we used points instead of Great British Pounds (£1.00 = 100 points). Over 10 rounds (representing 10 seasons), individuals received probabilistic forecasts that the next season could be extreme or normal, without knowing the true underlying risk.

This experimental setup allows to systematically and randomly manipulate participants' false and missed alarm experiences, which is an advantage over observational data. With observational data, the frequency of accurate and inaccurate forecasts likely correlates with the location of residence, which would likely correlate with many other confounding factors that also influence individual adaptation decisions (such as socio-economic characteristics, background risk, risk preferences, etc.). It is consequently challenging to identify the causal effect of repeated false and missed alarms with observational data.

4.2.2 Forecast design

The design of our experimental forecast is based on common seasonal forecasts of precipitation or temperature, which present the probability whether the upcoming season is likely to be normal, below or above normal (see for example the seasonal forecasts provided by the International Research Institute for Climate and Society at Columbia University, United States of America, https://iri.columbia.edu/, accessed June 29 2021). Seasonal forecasts are inevitably probabilistic due to the uncertainties in climate models and imperfect knowledge of the atmosphere and climate system (Smith et al. 2019). For our experiment, we therefore decided to implement a probabilistic forecast system. If individuals are only presented with a deterministic warning, they cannot judge the underlying uncertainty of the forecasts and thus cannot act on their individual probabilistic threshold (Fundel et al. 2019). Previous experimental studies compared individuals' response to deterministic and probabilistic forecast systems (LeClerc and Joslyn 2015; Losee and Joslyn 2018). They found that probabilistic forecasts reduce the perceived inconsistency of warnings and have a positive effect on trust in the forecast system in general (ibid.).

To simplify the understanding of our experiment, we only introduced two forecast categories, normal and extreme seasons. The forecast showed the probabilities of both extreme and normal seasons for that upcoming season (Fig. 4.1). If the forecasted probability for an extreme season was 60% or higher, participants received a *warning forecast* (Fig. 4.1a). In contrast, if the forecasted probability for an extreme season was 40% or lower, participants received a standard forecast message (i.e. *no-warning forecast*, Fig. 4.1b).

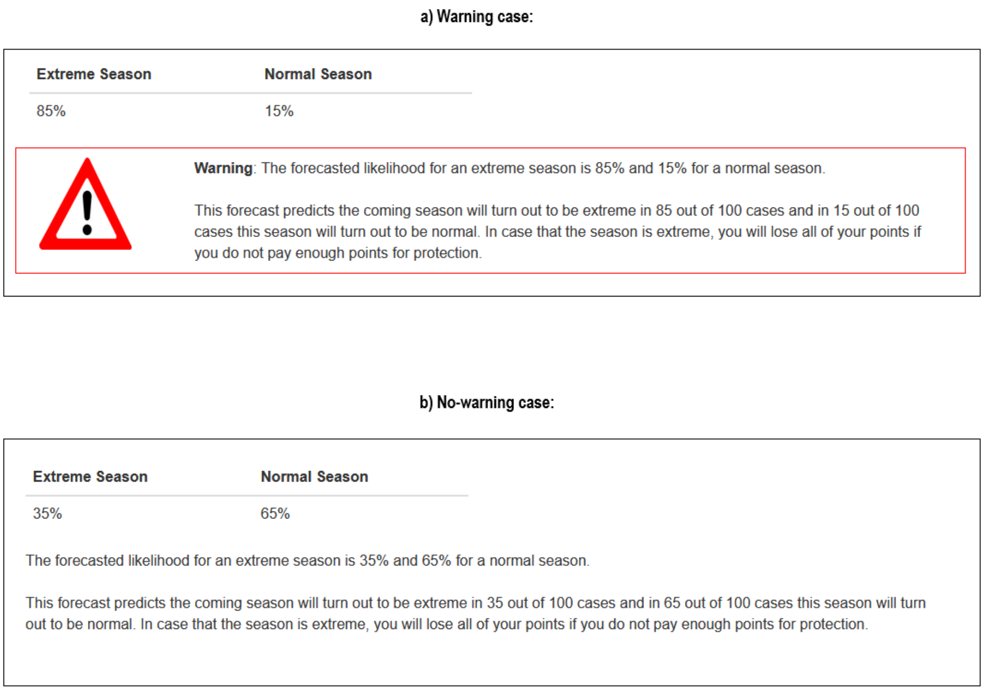


Fig. 4.1 Screenshots of exemplary forecasts in the experiment. Fig. 4.1a presents the forecast design of the experiment in the warning case, Fig. 4.1b in the no-warning case. The forecast probabilities for an extreme and normal season were presented in both cases.

4.2.3 Elicitation of adaptation behaviour

After each forecast, we elicited individuals' adaptation behaviour as their willingness to pay (WTP) for protection from the loss due to a potential extreme season. Individuals stated how much of their bonus they were willing to invest in adaptation (between 0 and 500 points).

We opted for eliciting individuals' WTP to yield a continuous measure for adaptation behaviour instead of simply eliciting a binary decision of whether to adapt (i.e. follow the warning) or not. With this, not only can we measure fine, more nuanced treatment effects, but also the continuous measurement of behaviour resembles many real-life adaptation decisions.

We used the Becker–DeGroot–Marschak (BDM) method to elicit individuals' WTP (Becker et al. 1964), which has the advantage that it is incentivised and individuals are motivated to state their actual WTP because it does not influence the actual price (Schmidt and Bijmolt 2019). Following the BDM method (Becker et al. 1964), the individuals' stated amount indicated the maximum they were prepared to pay for protection, not the actual price that protection would cost. Participants knew that only one season would be randomly selected as payout-relevant at the end. For the selected season, the price was randomly determined between 0 and 500 points. If participants had indicated a WTP equal to or higher than the price, they purchased the protection for the determined price. They thus received the rest of their bonus as payout, irrespective of an extreme or normal season. However, if participants had indicated a WTP lower than the randomly determined price, they did not buy protection and were not protected from extreme climate-related losses. They would then lose their full bonus if the season was extreme, but keep their full bonus if the season was normal.

4.2.4 Treatments

We implemented three treatments that differed in the forecast systems' accuracy but shared a common probability distribution for the underlying risk of an extreme season. The true underlying risk of an extreme season was randomly drawn each season from the same twelve pre-defined risk options (probabilities) in all treatments. In half of the options, the probability of an extreme season was below 50% (a1=0.15 to a6=0.4), and in the other half above 50% (a7=0.6 to a12=0.85, Table 4A.1). The *a priori* likelihood of an extreme season was 50%. Overall, half of all 10 seasons (M=0.5, SD=0.16, N=2,000) in the experiment were of extreme climate, which shows that the intended balance was achieved.

In the *control treatment (CTRL)*, the forecast system was accurate, and forecasted probabilities represented the true underlying risk of an extreme season. Thus, on average, half of the forecasts would issue a warning of an extreme season (i.e. probability of an extreme season 60% or higher) and the other half would not (i.e. probability of an extreme season 40% or lower). In the *false alarm treatment (FA)* and the *missed alarm treatment (MA)*, respectively, each of the twelve underlying risk options were matched with an over- or underrated forecast probability such that the false alarm rate in FA was set to 0.5 while the missed alarm rate was set to 0.17 (*vice versa* in MA). Thus, in FA, the presented probabilities in the forecasts overstated the risk of an extreme season and too many warnings were issued. While in MA, the presented probabilities understated the risk of an extreme season, with too few warnings being issued (see Table 4A.1).

We can verify our treatment designs regarding the forecast inaccuracy by analysing the frequency at which accurate forecasts, false and missed alarms occurred in the first 9 seasons (Table 4A.1). In CTRL (N=667), 38% of the warning forecasts and 35% of the no-warning forecasts were accurate, 13% were false alarms and 13% were missed alarms. In FA (N=667), 35% were accurate warnings, 3% were accurate “no-warnings”, 48% were false alarms and 14% were missed alarms. Whereas in MA (N=666), 3% were accurate warnings, 38% were accurate “no-warnings”, 13% were false alarms and 47% were missed alarms. Thus, the *a posteriori* probabilities for the four different forecast cases match the *a priori* probabilities that we aimed for and the false and missed alarm-prone forecast systems led to the desired rate of inaccuracy.

Participants knew that they were randomly allocated to a forecast system of varying accuracy (i.e. the treatments) at the beginning of the experiment and that they would receive forecasts from the same system throughout the experiment. However, participants did not know which forecast system they had been assigned to. Participants thus received a forecast with a probability of an extreme season without knowing the true underlying risk of an extreme season in the specific season.

In all treatments, at the end of each season, the computer randomly determined the outcome of the season based on the underlying risk and participants were informed whether the last season was extreme or normal. Thereby, participants received indirect feedback on the accuracy of their forecast system.

We implemented seasons one to nine to create experiences with accurate, false or missed alarm-prone forecasts for participants depending on their assigned treatment. Our outcome variable of interest is individuals’ WTP for protection in the final season (season 10). In order to have a clean comparison between treatments, the forecasts in season 10 in all treatments were based on CTRL (unknown to participants). Thus, in season 10, participants in all three treatments had the same likelihood to receive a warning or no-warning forecast with the corresponding forecast probabilities. In the analysis, we test for treatment effects in the two sub-samples who received and did not receive a warning in this last round.

4.2.5 Hypotheses

We hypothesise that *experiencing false alarms more frequently decreases adaptation investments in response to a warning forecast* ("cry-wolf effect", Hypothesis 3.1). That is, individuals that received a warning in season 10 are expected to state, on average, a lower WTP in FA than in CTRL. Furthermore, we hypothesise that *experiencing missed alarms more frequently increases adaptation investments in response to a no-warning forecast* (Hypothesis 3.2). That is, individuals who received no warning in season 10 are expected to state, on average, a higher WTP in MA than in CTRL. Besides these two main effects, we analyse the two cross-effects. Thus, we hypothesise that *experiencing false alarms more frequently increases adaptation investments in response to a no-warning forecast* (Hypothesis 3.3) and *experiencing missed alarms more frequently decreases adaption investments in response to a warning forecast* (Hypothesis 3.4).

4.3 Data collection

For our experiment, we recruited 2,000 residents of the United Kingdom (UK) via the online crowdsourcing platform Prolific (Prolific 2021) in July 2020. The average completion time of the experiment and post-experimental questionnaire was 14 minutes (SD=7, N=1,996) of which 10 minutes (SD=5, N=2,000) were spent on the experiment itself. The average payout including the bonus was £4.95 (SD=1.99, N=2,000).

We decided to use an online experiment instead of a lab experiment because they are known to be more cost-efficient and allow for much larger samples with higher power than lab experiments (Peer et al. 2017; Palan and Schitter 2018). In addition, participants of online experiments are usually from more diverse social backgrounds and represent a greater proportion of the general population than students participating in lab experiments (ibid.). In an online experiment, we cannot control for the accuracy of participants' identities (Palan and Schitter 2018) and their socio-economic characteristics. However, we have no reason to believe that participants had an incentive to provide inaccurate information because they knew that their answers to these questions were not payout-relevant and that all answers were anonymous.

Our study was programmed using the "Software Platform for Human Interaction Experiments" (SoPHIE) (Hendriks 2012). The study consisted of two parts, an economic experiment and a post-experimental questionnaire. Participation was voluntary and we followed the common ethical standards of experimental economics. Participants' responses and payments were treated with confidentiality and participants' identities remained anonymous. All participants gave their consent at the beginning of the study. We preregistered our study at "AsPredicted.org" (Wharton Credibility Lab, 2017) (see Appendix 4D for details).

We ran two separate sessions for female and male participants with 1,000 participants each. The results of our pre-test sessions showed that female participants were more active on Prolific, so we used Prolific's pre-screening function to successfully assure gender balance. We limited geographical variation by only allowing individuals that currently live in the United Kingdom to participate in the study. We asked participants if they had previously experienced any weather extremes such as blizzards, floods, droughts, etc, and the size of their financial damage caused by these events. The majority had no previous experience with extreme weather events (91%) (Table 4B.1), and overall, we found no significant effect that a previous experience of extreme events had on WTP (Table 4C.2).

4.3.1 Treatment assignment

The computer assigned participants to one of the three treatments based on the order in which participants finished the instructions. The first participant to finish was assigned to the control treatment CTRL, the second participant to the FA treatment, the third participant to the MA treatment, the fourth to CTRL again and so on. Thereby, we were able to balance the sample such that participants that registered early or late were evenly distributed among the three treatments. Participants who registered earlier are potentially more experienced and more eager to earn money by participating in experiments. We do not find statistically significant differences at the 5%-level in the socio-economic characteristics between treatments (Table 4B.1). Further, there are no significant differences between treatments regarding participants' Prolific scores which measure the percentage of studies that participants completed previously and that were accepted by other researchers. Overall the average score is very high across all treatments (99.5, SD= 1.4, N=2,000).

4.3.2 Sample characteristics

The average age of participants in our study was 34.5 years old (SD=12.7), with the majority living in urban areas of between 10,001 to 100,000 inhabitants (49%). The average household size was three individuals (SD=1.4), 76% of all participants have a disposable monthly household income below £4, 500, and 41% are the owners of the house they live in. The majority of participants have either a college degree (27%) or a Bachelor's degree (38%). Previous studies have found that higher education, homeownership and living in a rural area increase individuals' willingness to take protective measures against extreme events (Kox and Thieken 2017). However, we do not find evidence that these characteristics influence individuals' WTP in our experiment (Table 4C.2).

In addition to the questions on participants' social backgrounds, we also elicited individuals' risk preferences and their perception of climate change. We used the risk question based on Dohmen et al. (Dohmen et al. 2010) to elicit participants' risk preferences. Participants rate how prepared they are to take risks by choosing a value from a 10-point-Likert scale from 0 meaning "not at all willing to take risks" to 10 meaning "very willing to take risks" (ibid.). We find that the average answer across all three treatments is 5.2 (SD=2.3). Overall, participants' risk preference has a statistically significant effect on their stated WTP ($p<0.01$, Table 4C.2): As expected, the higher the participants' willingness to take risks, the lower their stated WTP.

We implemented three measures to assure participants' understanding and attentiveness during the experiment. Firstly, we included four control questions in the instructions to check participants' understanding of the experiment before making payout-relevant decisions. Individuals' answers to the control questions were not payout-relevant. If participants gave a wrong answer, the relevant part of the instructions was shown again and participants could answer the question again. The majority of participants (87%) answered all four control questions correctly after a maximum of two attempts per question.

Secondly, we checked for individuals' attentiveness by adding a trap question (Berinsky et al. 2014; Malone and Lusk 2018). Trap questions help filter for inattentive participants (ibid.). Our trap question was a colour screener similar to the one presented in Berinsky et al. (2014). Initially, participants are asked to read the given instructions carefully. The instructions outline that the question is designed to control for participants' attentiveness. Next, participants are instructed to answer the following question about their favourite colour with "none". Attentive participants give the answer "none" and ignore the question about their favourite colour, while inattentive participants state their favourite colour. Overall, 94% of participants answered "none" indicating they paid attention to the study.

Thirdly, we added a question regarding individuals' environment at the time of their participation and which device they were using because no control over the environment is often mentioned as a disadvantage of online experiments (Peer et al. 2017; Palan and Schitter 2018). The majority of participants (49%) used their laptop to participate in the study. Another 23% used a desktop computer and 22% used their mobile phone. The vast majority of participants, 89%, stated that they participated while being at home without any distractions around them. This comparatively high percentage could be a result of the UK government's coronavirus measures including local lockdowns that were in place at the time of the experiment (July 2020).

Overall, the results of our control, attention, environment, and device questions point out that the majority of participants read the instructions carefully, paid attention to the experiment, and were not distracted by their surroundings while making their decisions. Furthermore, we ran Tobit regression models on the sub-sample that took at most two attempts to answer all four control questions correctly and did not fail the attention check as robustness checks of our treatment effects (Table 4B.2). The results of these robustness checks are in line with our main findings.

Given the framing of our experiment in the context of seasonal forecasts and extreme climate conditions, we anticipated that participants' belief in climate change could potentially affect their decision-making. Therefore, we asked them how they perceive the impact of climate change on people across the world and on their own life (Phillips et al. 2018). Participants could choose from a 10-point-Likert scale where 0 meant “extremely bad” to 10 “extremely good” (ibid.). Participants indicated an average score of 2.1 (SD=1.9) for the impact on people across the world and 3.6 (SD=1.6) for the impact on their own lives. Thus, individuals see the impact on the people across the world worse than on their own lives. However, we do not find a significant effect of climate change perception on participants' WTP (Table 4C.2). Please see Table 4B.1 for further details of participants' socio-economic characteristics.

4.4 Results

The average WTP for protection across treatments was 329 points (SD=128, N=1,011) in the warning case and 180 points (SD=139, N=989) in the no-warning case. The average WTP per treatment is shown in Table 4.1.

Table 4.1 Summary statistics of the main outcome, WTP in season 10

	Warning		No Warning	
Treatment	**Mean WTP (SD)**	**N**	**Mean WTP (SD)**	**N**
CTRL	330 (120)	342	170 (130)	325
FA	305 (131)	316	170 (137)	351
MA	349 (129)	353	202 (147)	313

Note: N denotes the number of observations. Mean WTP is the mean WTP in season 10. Given the bonus of 500 points, the minimum possible WTP was 0 and the maximum 500 in all treatments. Standard deviations (SD) are presented in parentheses. “Warning” includes only the observations where individuals received a warning in round 10, while “no warning” includes only the observations of individuals receiving no warning in round 10.

To test our hypotheses, we ran four Tobit regression models with individuals' average WTP in season 10 as dependent variable. The focus is on a between-treatment analysis with CTRL being the control treatment. We used the observations of CTRL repeatedly for the comparisons to FA and MA and have two sub-samples per treatment depending on whether or not a warning was issued in the last round.

Figure 4.2 presents the treatment coefficients of the four Tobit models (see Table 4C.1 for details), one for each hypothesis. In all models, we control for the stated probability of an extreme season in the forecast. Our results are robust to specifications including socio-demographic controls and specifying the treatment variable as relative frequency of false and missed alarms in prior seasons (Table 4C.2 and 4C.3).

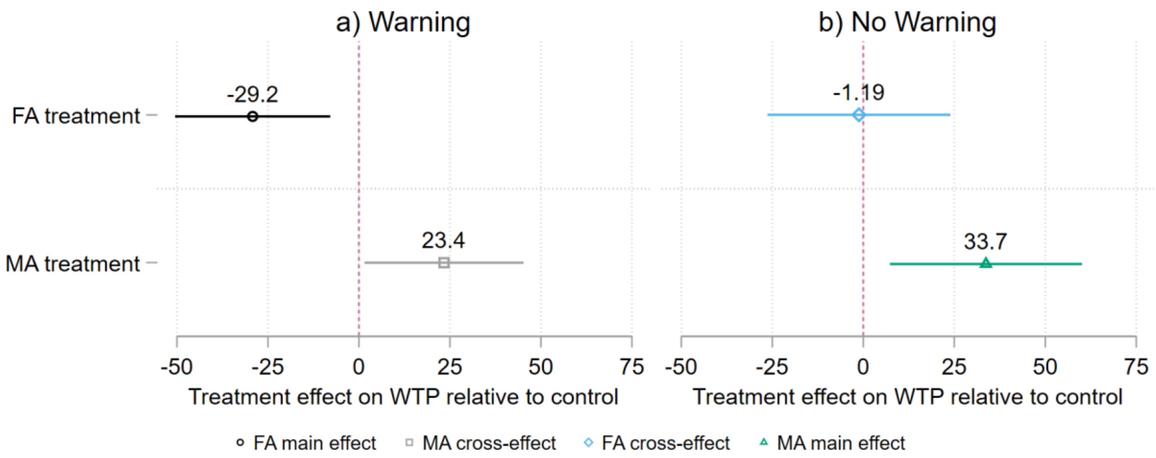


Fig. 4.2 Coefficient plots based on the four Tobit regression models with the dependent variable "WTP in season 10". The coefficient plots display the point estimates for the coefficients "False alarm treatment" (FA) and "Missed alarm treatment" (MA) with their 95%-confidence intervals along the x-axis. These coefficients represent the treatment effect on WTP and are estimated relative to the control treatment CTRL. Fig. 4.2a shows the coefficient plots for the warning case (Hypotheses 1 and 4), also defined as the FA main and MA cross-effect. Fig. 4.2b shows the coefficient plots for the no-warning case (Hypotheses 2 and 3), also known as MA main and FA cross-effect. The dotted, vertical line at zero is a reference line to visualise which coefficients are significantly different from zero at the 0.05 level. See Table 4C.1 for the corresponding Tobit regression models.

4.4.1 Hypothesis tests

Experiencing false alarms more frequently decreases average adaptation investments in response to a warning. Individuals' WTP in the warning case is on average significantly lower in FA than in CTRL (Fig. 4.2a), confirming Hypothesis 1. However, the effect is only small (d = - .20). Our finding of a *cry-wolf effect* is in line with previous studies focusing on automated machine alerts (Chancey et al. 2015; Wiczorek and Meyer 2016) and binary decision-making in response to weather forecasts (LeClerc and Joslyn 2015).

Experiencing missed alarms more frequently increases adaptation investments in the absence of a warning. We find a significant increase in the WTP among individuals that experience no warning in MA compared to CTRL (Fig. 4.2b). We therefore also confirm Hypothesis 2, and again, the treatment effect is small (d = .231). This finding adds to previous studies that focus only on deterministic warning systems (Chancey et al. 2015; Wiczorek and Meyer 2016). We also find that with a probabilistic forecast, individuals rely less on no-warning forecasts if they experienced missed alarms *more frequently*.

There is no evidence of a negative cross-effect on adaptation investments in the absence of a warning from experiencing false alarms more frequently. We do not observe a significant difference in individuals' WTP when comparing FA to CTRL in the no-warning case (Fig. 4.2b). Our data thus does not support Hypothesis 3, namely that experiencing false alarms more frequently increases adaptation investments if no warning is issued. This result agrees with Manzey et al. (2014), but is in contrast to the experimental study on deterministic machine warnings by Wiczorek and Meyer (2016) who found a negative cross-effect.

Contrary to Hypothesis 4, we find evidence of a positive cross-effect on adaptation investments if a warning is issued from experiencing missed alarms more frequently. Average WTP in the warning case is significantly higher among individuals in MA compared to CTRL (Fig. 4.2a). Nonetheless, this MA cross-effect is smaller than the two main effects (d=.15). Our result does not concur with previous studies on deterministic warning systems that find some evidence for a negative cross-effect (LeClerc and Joslyn 2015; Ripberger et al. 2015) or evidence for no effect (Manzey et al. 2014; Wiczorek and Meyer 2016).

4.4.2 Trust in the forecast system

Experiencing more frequent false or missed alarms negatively impacts trust in forecast systems. We further examine the correlation between participants' trust in forecast systems and their stated WTP to shed some light on the underlying mechanism of the observed treatment effects. Previous studies suggest that trust affects the relationship between forecast accuracy and individuals responsiveness to warning and no-warning forecasts (Buontempo et al. 2014; Chancey et al. 2015, 2017; Ripberger et al. 2015; Trainor et al. 2015; DeYoung et al. 2019). In the post-experimental questionnaire, we asked participants to rate their trust in the last forecast from 1 (strong mistrust) to 6 (strong trust). We find that trust is significantly higher in CTRL (M=4.2, SD=1.1, N=642) than in FA (M=3.7, SD=1.3, N=643; two-sided t-test: t(1,283)=7.64, p=0.000) and MA (M=3.8, SD=1.3, N=636; two-sided t-test: t(1,276)=6.73, p=0.000). There is no significant difference between FA and MA (two-sided t-test: t(1,277)= -0.84, p=0.4) (see Table 4C.4 for regression results). These results support previous findings that inaccurate forecasts decrease trust in the warning system (LeClerc and Joslyn 2015; Chancey et al. 2017). However, in contrast to Chancey et al. (2017) we do not find that experiencing false alarms more frequently has a stronger impact on trust than experiencing more frequent missed alarms.

Trust in the forecast system affects adaptation investments only if a warning is issued. If a warning was issued, trust is positively correlated with WTP (Spearman correlation coefficient, r_s=0.2, p=0.000, see Table 4C.5 for regression results). In the no-warning case, trust is weakly negatively correlated with WTP (r_s=-0.06, p=0.086), but has no statistically significant effect on WTP in regression models (p>0.1, see Table 4C.5).

4.4.3 Treatment effects relative to forecasted probabilities

Even though we observe statistically significant treatment effects for three of the four hypotheses, one key question is how strong these effects are relative to other determinants of adaptation investments. In this section, we therefore compare the treatment effects to the effects of the forecasted probabilities. We also assessed if experiencing false or missed alarms *more frequently* affects the sensitivity to the forecasted probabilities. To do this, we ran four Tobit models including the interaction of treatment dummy and forecasted probability (see Table 4C.6). The predicted adaptation investments by treatment in relation to the different forecast probabilities are illustrated in Figure 4.3.

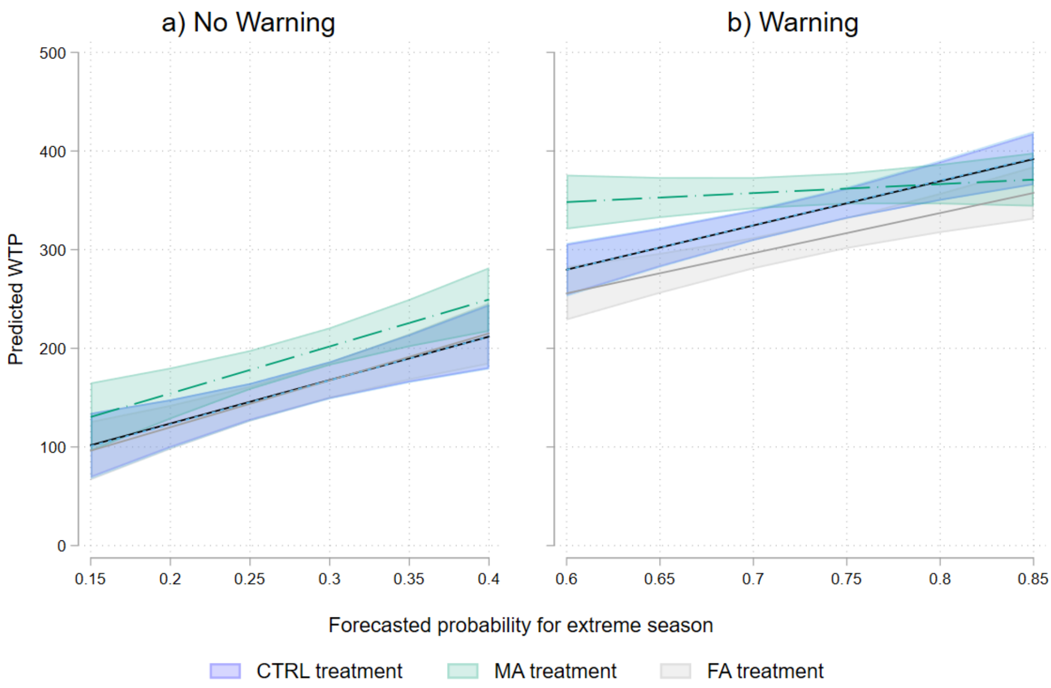


Fig. 4.3 Adjusted predictions of mean WTP in season 10 per treatment with 95% confidence interval bands. CTRL treatment is the control treatment with accurate forecasts, FA the false alarm-prone and MA the missed alarm-prone treatment. Fig. 4.3a shows the no warning case with forecasted probability for an extreme season below 0.4. In this case, the adjusted predictions of CTRL and FA are overlapping. Fig. 4.3b shows the warning case with forecasted probability for an extreme season above 0.6.

The treatment effects are relatively small compared to the effects of the forecasted probabilities. We observe that overall, the higher the forecasted probability, the higher the predicted WTP (Fig. 4.3). These relative effects of the forecasted probabilities are overall stronger than the treatment effects (.2 < d > 1.09, see Table 4C.7). The behavioural responses to the forecasted probabilities are thus stronger than the impact of false or missed alarm-prone forecast systems on individual behaviour.

More frequent false alarms do not affect the sensitivity to forecasted probabilities, while more frequent missed alarms to some extent do. In the no-warning case, the near-identical slopes of the treatments indicate that the sensitivity to the forecast probabilities is very similar across treatments (see Fig. 4.3a). Similarly, in the warning case, the predicted WTP in FA and

CTRL are nearly parallel indicating that the sensitivity to the probability in both treatments is similar (Fig. 4.3b). In contrast, the slope of MA compared to both CTRL and FA is flatter in Figure 4.3b, and thus WTP is, on average, less sensitive to the forecasted probability in MA. In the MA treatment, WTP is generally higher independent of the forecast probabilities (the MA cross-effect).

4.5 Discussion and conclusions

Based on our online experiment with 2000 participants, we find evidence that experiencing false or missed alarms more frequently affects individuals' responsiveness to forecasts. More frequent false alarms decrease individual adaptation investment if a warning is issued (the *cry-wolf effect*). More frequent missed alarms increase individual investments irrespective of whether or not a warning is issued. In both cases we observe a decrease in individuals' trust in the forecasts.

Based on our literature review, we identified three different ways through which forecast inaccuracies may affect adaptation behaviour. Firstly, individuals who observe multiple false or missed alarms are potentially less likely to follow the advice given by the forecast system since they learned that false and missed alarms occur frequently. This could be considered frequency-based probability learning (Estes 1976). Ultimately, individuals may lose trust generally in the forecast system, resulting in a decrease of individuals' willingness to respond to the information provided by the forecast system (Ripberger et al. 2015).

Secondly, individuals that experience false alarms more frequently could erroneously believe that a false alarm in the future is less likely to happen (same for missed alarms). Such an erroneous belief that events are conditional even though there are independent is called the *gambler's fallacy* (e.g. Rabin and Vayanos (2010)). Following the gambler's fallacy, individuals that experience multiple false alarms would expect that the next warning they receive is true and would be more willing to comply with future warnings. Similarly, individuals that experience multiple missed alarms would rely more on future forecasts and be less likely to invest in adaptation if no warning is issued.

Lastly, the actual loss of income and wealth in the real world might influence future decisions regarding individuals' responses to forecast systems. For example, experiencing losses due to a missed alarm potentially leads to an increase in future adaptation investments simply because a second loss might not be financially viable, and so individuals opt to invest independently of the forecast to prevent financial ruin. Conversely, experiencing losses could also limit future investment in adaptation, regardless of any issued warnings.

We assess the impact of trust in the forecast system on individuals' willingness to invest in adaptation. Our results suggest that the gambler's fallacy did not dominate participants' choice of WTP because, on average, participants decreased their WTP due to false alarms and increased their WTP due to missed alarms. If participants had erroneous beliefs in line with the gambler's fallacy, we should have observed the opposite. The design of our experiment excludes income and wealth effects as potential drivers because participants knew that only one of the ten seasons was randomly chosen to be payout-relevant in the end.

Unexpectedly, we find that experiencing missed alarms more frequently increases the responsiveness to warnings. This unexpected cross-effect potentially occurs because false and missed alarms affect two different facets of trust in the forecast system. Firstly, missed alarms potentially reduce trust in the forecasted probability itself, but not in the warning *per se*. Individuals may assume that if a warning is issued with a moderately low forecast probability, the forecast still understates the probability of an extreme season. They therefore are more willing to invest in adaptation compared to individuals in the false alarm-prone and accurate forecast treatments. It seems that individuals learn that the forecasts systematically underestimate risks and compensate accordingly. As a result, individuals are less responsive to an increase in the forecasted probability and are generally more willing to invest in adaptation. Secondly, experiencing false alarms more frequently seems to lower trust in the warning itself, but not in the forecasted probability in case that no warning is issued. Individuals are less likely to comply with an issued warning but still rely on the forecast in the absence of a warning.

We find that the observed treatment effects of experiencing false and missed alarms more frequently are relatively small in relation to the effects of the forecasted probabilities. Even if the forecast system is prone to false or missed alarms, individuals respond to an increase of the forecasted probabilities with larger adaptation investments. This can be considered conducive for a wider application of seasonal forecasts and early warning signals since the long-term costs of inaccuracies may indeed be limited. Whether more frequent false or missed alarms cause more harm in the long run inevitably also depends on the case-specific stakes at risk and adaptation costs. Policy makers must carefully assess the forecast users and their risk profiles before deciding on the communication and design of forecasts. In the future, we suggest implementing additional experiments with users of domain-specific seasonal forecasts or early warning signals (e.g. with farmers) to deepen our knowledge of the implications of forecast inaccuracy on adaptation behaviour.

References Chapter 4

Becker GM, DeGroot MH, Marschak J (1964) Measuring utility by a single-response sequential method. Behav Sci 9:226–232

Berinsky AJ, Margolis MF, Sances MW (2014) Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. Am J Polit Sci 58:739–753. https://doi.org/10.1111/ajps.12081

Bruno Soares M, Alexander M, Dessai S (2018) Sectoral use of climate information in Europe: A synoptic overview. Clim Serv 9:5–20. https://doi.org/10.1016/j.cliser.2017.06.001

Buontempo C, Hewitt CD, Doblas-Reyes FJ, Dessai S (2014) Climate service development, delivery and use in Europe at monthly to inter-annual timescales. Clim Risk Manag 6:1–5. https://doi.org/10.1016/j.crm.2014.10.002

Chancey ET, Bliss JP, Liechty M, Proaps AB (2015) False Alarms vs. Misses: Subjective Trust as a Mediator Between Reliability and Alarm Reaction Measures. In: Proceedings of the Human Factors and Ergonomics Society

Chancey ET, Bliss JP, Yamani Y, Handley HAH (2017) Trust and the Compliance–Reliance Paradigm: The Effects of Risk, Error Bias, and Reliability on Trust and Dependence. Hum Factors 59:333–345. https://doi.org/10.1177/0018720816682648

DeYoung SE, Sutton JN, Farmer AK, et al (2019) “Death was not in the agenda for the day”: Emotions, behavioral reactions, and perceptions in response to the 2018 Hawaii Wireless Emergency Alert. Int J Disaster Risk Reduct 36:. https://doi.org/10.1016/j.ijdrr.2019.101078

Dohmen T, Falk A, Huffman D, Sunde U (2010) Are Risk Aversion and Impatience Related to Cognitive Ability? Am Econ Rev 100:1238–1260

Dow K, Cutter SL (1998) Crying Wolf: Repeat Responses to Hurricane Evacuation Orders. Coast Manag 26:237–252

Estes WK (1976) The cognitive side of probability learning. Psychol Rev 83:37–64. https://doi.org/10.1037/0033-295X.83.1.37

Fundel VJ, Fleischhut N, Herzog SM, et al (2019) Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. Q J R Meteorol Soc 145:210–231. https://doi.org/10.1002/qj.3482

Hendriks A (2012) SoPHIE-Software platform for human interaction experiments. Osnabrueck: Working Paper, Osnabrueck University

IPCC, 2019. Summary for Policymakers, Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. https://doi.org/10.4337/9781784710644

Kahneman D (2011) Thinking, Fast and Slow. Farrar Straus & Giroux, New York

Knudson C, Guido Z (2019) The missing middle of climate services: layering multiway, two-way, and one-way modes of communicating seasonal climate forecasts. Clim Change 157:171–187. https://doi.org/10.1007/s10584-019-02540-4

Kox T, Thieken AH (2017) To act or not to act? Factors influencing the general public's decision about whether to take protective action against severe weather. Weather Clim Soc 9:299–315. https://doi.org/10.1175/WCAS-D-15-0078.1

LeClerc JE, Joslyn S (2015) The cry wolf effect and weather-related decision making. Risk Anal 35:385–395. https://doi.org/10.1111/risa.12336

Lim JR, Fisher Liu B, Egnoto M (2019) Cry Wolf Effect ? Evaluating the Impact of False Alarms on Public Responses to Tornado Alerts in the Southeastern United States. Weather Clim Soc 11:549–563. https://doi.org/10.1175/WCAS-D-18-0080.1

Losee JE, Joslyn S (2018) The need to trust: How features of the forecasted weather influence forecast trust. Int J Disaster Risk Reduct 30:95–104. https://doi.org/10.1016/j.ijdrr.2018.02.032

Malone T, Lusk JL (2018) Consequences of Participant Inattention with an Application to Carbon Taxes for Meat Products. Ecol Econ 145:218–230. https://doi.org/10.1016/j.ecolecon.2017.09.010

Manzey D, Gérard N, Wiczorek R (2014) Decision-making and response strategies in interaction with alarms : the impact of alarm reliability , availability of alarm validity information and workload. Ergonomics 57:1833–1855

National Institute of Water and Atmospheric Research (NIWA) (2016) Improving seasonal climate forecasts. [Online]. In *NIWA News and Publications*, 2016-02-08. https://niwa.co.nz/news/improving-seasonal-climate-forecasts. Accessed 11 Mar 2020

Palan S, Schitter C (2018) Prolific.ac—A subject pool for online experiments. J Behav Exp Finance 17:22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Peer E, Brandimarte L, Samat S, Acquisti A (2017) Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. J Exp Soc Psychol 70:153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Phillips D, Curtice J, Phillips M, Perry J (2018) British Social Attitudes: The 35th Report. London

Prolific (2021) Prolific. https://www.prolific.co/. Accessed 14 May 2021

Rabin M, Vayanos D (2010) The gambler's and hot-hand fallacies: Theory and applications. Rev Econ Stud 77:730–778. https://doi.org/10.1111/j.1467-937X.2009.00582.x

Ripberger JT, Silva CL, Jenkins-Smith HC, et al (2015) False Alarms and Missed Events: The Impact and Origins of Perceived Inaccuracy in Tornado Warning Systems. Risk Anal 35:44–56. https://doi.org/10.1111/risa.12262

Schmidt J, Bijmolt THA (2019) Accurately measuring willingness to pay for consumer goods: a meta-analysis of the hypothetical bias. J Acad Mark Sci. https://doi.org/10.1007/s11747-019-00666-6

Schultz DM, Gruntfest EC, Hayden MH, et al (2010) Decision making by Austin, Texas, residents in hypothetical tornado scenarios. Weather Clim Soc 2:249–254. https://doi.org/10.1175/2010WCAS1067.1

Simmons KM, Sutter D (2009) False alarms, tornado warnings, and tornado casualties. Weather Clim Soc 1:38–53. https://doi.org/10.1175/2009WCAS1005.1

Smith DM, Eade R, Scaife AA, et al (2019) Robust skill of decadal climate predictions

Taylor AL, Dessai S, Bruine De Bruin W (2015) Communicating uncertainty in seasonal and interannual climate forecasts in Europe. Philos Trans R Soc A 373:. https://doi.org/10.1098/rsta.2014.0454

Taylor AL, Kox T, Johnston D (2018) Communicating high impact weather: Improving warnings and decision making processes. Int J Disaster Risk Reduct 30:1–4. https://doi.org/10.1016/j.ijdrr.2018.04.002

Trainor JE, Nagele D, Philips B, Scott B (2015) Tornadoes , Social Science , and the False Alarm Effect. Weather Clim Soc 7:333–352. https://doi.org/10.1175/WCAS-D-14-00052.1

Webber S (2019) Putting climate services in contexts: advancing multi-disciplinary understandings: introduction to the special issue. Clim Change 157:1–8. https://doi.org/10.1007/s10584-019-02600-9

Wharton Credibility Lab (2017) AsPredicted. https://aspredicted.org. Accessed: 20 October 2021.

Wiczorek R, Meyer J (2016) Asymmetric effects of false positive and false negative indications on the verification of alerts in different risk conditions. Proc Hum Factors Ergon Soc 289–292. https://doi.org/10.1177/1541931213601066

Zommers Z (2012) Climate early warning system feasibility report: Early warning systems and hazard prediction. United Nations Environment Programme (UNEP)

Appendices Chapter 4

Appendix 4A: Details of the treatment design

Table 4A.1. Forecast probability design and outcome in the three treatments

CTRL - accurate (= true underlying risk)				**FA - false alarm-prone**				**MA - missed alarm-prone**			
	forecast probability				forecast probability				forecast probability		
option	extreme	normal	warning	option	extreme	normal	warning	option	extreme	normal	warning
a 1	0.15	0.85	no	f 1	0.6	0.4	yes	m 1	0.6	0.4	yes
a 2	0.2	0.8	no	f 2	0.65	0.35	yes	m 2	0.65	0.35	yes
a 3	0.25	0.75	no	f 3	0.7	0.3	yes	m 3	0.25	0.75	no
a 4	0.3	0.7	no	f 4	0.75	0.25	yes	m 4	0.3	0.7	no
a 5	0.35	0.65	no	f 5	0.8	0.2	yes	m 5	0.35	0.65	no
a 6	0.4	0.6	no	f 6	0.85	0.15	yes	m 6	0.4	0.6	no
a 7	0.6	0.4	yes	f 7	0.6	0.4	yes	m 7	0.15	0.85	no
a 8	0.65	0.35	yes	f 8	0.65	0.35	yes	m 8	0.2	0.8	no
a 9	0.7	0.3	yes	f 9	0.7	0.3	yes	m 9	0.25	0.75	no
a 10	0.75	0.25	yes	f 10	0.75	0.25	yes	m 10	0.3	0.7	no
a 11	0.8	0.2	yes	f 11	0.35	0.65	no	m 11	0.35	0.65	no
a 12	0.85	0.15	yes	f 12	0.4	0.6	no	m 12	0.4	0.6	no
a priori probabilities of season 1 to 9[a]	accurate warning		0.36	**a priori probabilities of season 1 to 9**[a]	accurate warning		0.36	**a priori probabilities of season 1 to 9**[a]	accurate warning		0.03
	false alarm		0.14		false alarm		0.47		false alarm		0.14
	accurate no-warning		0.36		accurate no-warning		0.03		accurate no-warning		0.36
	missed alarm		0.14		missed alarm		0.14		missed alarm		0.47
a posteriori probabilities of season 1 to 9[b]	accurate warning		0.38 (0.17)	**a posteriori probabilities of season 1 to 9**[b]	accurate warning		0.35 (0.16)	**a posteriori probabilities of season 1 to 9**[b]	accurate warning		0.03 (0.05)
	false alarm		0.13 (0.11)		false alarm		0.48 (0.17)		false alarm		0.13 (0.11)
	accurate no-warning		0.35 (0.16)		accurate no-warning		0.03 (0.06)		accurate no-warning		0.38 (0.15)
	missed alarm		0.13 (0.12)		missed alarm		0.14 (0.12)		missed alarm		0.47 (0.17)

Note: Each season the computer randomly chooses one of the 12 options from a1 to a12, which determine the communicated forecast probabilities and whether or not individuals receive a warning. In CTRL, the values from a1 to a12, which are the true underlying risks are shown as forecasts. Participants in treatment FA and MA are shown the matching false or missed alarm-prone probabilities (f1 to f12 or m1 to m12 respectively) instead of the accurate, true underlying risks (a1 to a12).

[a]**A priori probabilities** refer to the a priori probabilities to receive an accurate warning, an accurate no-warning, a false alarm or a missed alarm in season 1 - 9. A priori probabilities for season 10 are the same for all treatments and follow the shown a priori probabilities in the CTRL treatment. [b]**A posteriori probabilities** are calculated based on the seasons one to nine, excluding season 10.

Appendix 4B: Details of sample characteristics

Table 4B.1. Balance table of socio-economic characteristics

Variable	Total (SD)	CTRL (SD)	FA (SD)	absolute difference (p-value)[1]	MA (SD)	absolute difference (p-value)[1]
Age (years)	34.475	34.228	34.348	-0.120	34.850	-0.622
	(12.666)	(12.782)	(12.050)	(0.860)	(13.153)	(0.381)
Female (fraction)	0.500	0.507	0.499	0.007	0.494	0.013
	(0.500)	(0.500)	(0.500)	(0.784)	(0.500)	(0.642)
Experienced extreme	0.086	0.088	0.079	0.009	0.092	-0.003
weather event (fraction)	(0.281)	(0.284)	(0.271)	(0.554)	(0.289)	(0.842)
Household size	2.920	2.890	2.931	-0.041	2.937	-0.047
	(1.388)	(1.350)	(1.336)	(0.584)	(1.474)	(0.551)
Owner of house (fraction)	0.412	0.408	0.409	-0.001	0.419	-0.011
	(0.492)	(0.492)	(0.492)	(0.956)	(0.494)	(0.680)
Living area (fraction)						
rural	0.212	0.210	0.204	0.006	0.221	-0.011
	(0.408)	(0.408)	(0.403)	(0.787)	(0.415)	(0.631)
urban	0.493	0.489	0.496	-0.007	0.494	-0.005
	(0.500)	(0.500)	(0.500)	(0.784)	(0.500)	(0.848)
metropolitan	0.274	0.270	0.286	-0.016	0.266	0.004
	(0.446)	(0.444)	(0.452)	(0.502)	(0.442)	(0.866)
Monthly disposable household income (GBP, fractions)						
Less than £1,499	0.230	0.259	0.217	0.042*	0.215	0.045*
	(0.421)	(0.439)	(0.413)	(0.072)	(0.411)	(0.055)
£1,500 to £2,999	0.331	0.321	0.328	-0.007	0.342	-0.022
	(0.471)	(0.467)	(0.470)	(0.770)	(0.475)	(0.405)
£3,000 to £4,499	0.197	0.175	0.205	-0.030	0.210	-0.035
	(0.398)	(0.381)	(0.404)	(0.163)	(0.408)	(0.108)
£4,500 to £5,999	0.075	0.067	0.090	-0.022	0.066	0.001
	(0.263)	(0.251)	(0.286)	(0.127)	(0.249)	(0.919)
£6,000 to £7,499	0.034	0.028	0.037	-0.009	0.036	-0.008
	(0.181)	(0.166)	(0.190)	(0.358)	(0.187)	(0.436)
Over £7,500	0.036	0.031	0.040	-0.009	0.036	-0.005
	(0.186)	(0.175)	(0.197)	(0.378)	(0.187)	(0.646)
Highest Educational Degree (fraction):						
Secondary school	0.139	0.150	0.123	0.027	0.143	0.007
	(0.346)	(0.357)	(0.329)	(0.151)	(0.350)	(0.707)
College	0.267	0.276	0.262	0.013	0.263	0.013
	(0.443)	(0.447)	(0.440)	(0.579)	(0.440)	(0.590)
Bachelor’s degree	0.384	0.370	0.396	-0.025	0.387	-0.017
	(0.487)	(0.483)	(0.489)	(0.339)	(0.488)	(0.521)
Post-graduate degree	0.198	0.195	0.207	-0.012	0.192	0.003
	(0.399)	(0.396)	(0.405)	(0.585)	(0.394)	(0.900)
Risk Aversion	5.182	5.306	5.097	0.208*	5.143	0.163
(0 = not willing to take risk, 10 = very willing to take risk)	(2.330)	(2.273)	(2.299)	(0.096)	(2.413)	(0.204)
Impact of Climate Change (0=extremely bad, 10=extremely good)						
on people across the world	2.077	2.075	2.124	-0.049	2.032	0.043
	(1.934)	(1.883)	(2.046)	(0.646)	(1.870)	(0.673)
on personal life	3.567	3.562	3.565	-0.003	3.574	-0.011
	(1.624)	(1.599)	(1.623)	(0.973)	(1.654)	(0.899)

Table 4B.1 Continued. Balance table of socio-economic characteristics

Variable	Total (SD)	CTRL (SD)	FA (SD)	absolute difference (p-value)[1]	MA (SD)	absolute difference (p-value)[1]
Failed Attention Check (fractions)	0.061	0.073	0.051	0.022*	0.059	0.015
	(0.239)	(0.261)	(0.220)	(0.089)	(0.235)	(0.274)
Control Questions (fractions)						
all correct at first try	0.439	0.430	0.411	0.019	0.476	-0.046*
	(0.496)	(0.495)	(0.492)	(0.471)	(0.500)	(0.094)
max two tries	0.431	0.426	0.465	-0.039	0.402	0.023
	(0.495)	(0.495)	(0.499)	(0.152)	(0.491)	(0.387)
more than three tries	0.130	0.144	0.124	0.019	0.122	0.022
	(0.336)	(0.351)	(0.330)	(0.297)	(0.327)	(0.230)
Participation Device (fractions)						
Desktop computer	0.230	0.217	0.235	-0.018	0.237	-0.020
	(0.421)	(0.413)	(0.425)	(0.433)	(0.426)	(0.388)
Laptop	0.491	0.498	0.495	0.003	0.479	0.019
	(0.500)	(0.500)	(0.500)	(0.913)	(0.500)	(0.493)
Tablet	0.053	0.046	0.055	-0.009	0.057	-0.011
	(0.224)	(0.211)	(0.229)	(0.456)	(0.232)	(0.384)
Mobile phone	0.223	0.235	0.208	0.027	0.227	0.009
	(0.417)	(0.425)	(0.406)	(0.236)	(0.419)	(0.708)
Participation Environment (fractions)						
home, no distractions	0.887	0.891	0.889	0.001	0.881	0.009
	(0.317)	(0.312)	(0.314)	(0.930)	(0.324)	(0.599)
home, distractions	0.079	0.082	0.070	0.012	0.083	-0.000
	(0.269)	(0.275)	(0.256)	(0.410)	(0.275)	(0.993)
not home, no distractions	0.031	0.025	0.037	-0.012	0.030	-0.005
	(0.173)	(0.158)	(0.190)	(0.210)	(0.171)	(0.614)
not home, distractions	0.001	0.000	0.000	0.000	0.003	-0.003
	(0.032)	(0.000)	(0.000)	(0.000)	(0.055)	(0.157)
Prolific score[a]	99.479	99.538	99.463	0.075	99.434	0.104
	(1.348)	(1.081)	(1.330)	(0.259)	(1.587)	(0.161)
Participation duration	14.205	14.347	13.945	0.402	14.322	0.025
(in minutes)	(6.987)	(7.184)	(6.676)	(0.291)	(7.093)	(0.950)
Joint F-Test				0.79		0.58
P-value				0.768		0.964
Observations	2,000	667	667	1,334	666	1,333

Note: Total in first column describes the overall average of the sample. CTRL denotes the control treatment with an accurate forecast system, FA the false alarm-prone treatment and MA the missed alarm-prone treatment.
Standard deviations (SD) in brackets. Fractions of missing values are not presented.
[a]***Prolific score*** states the percentage of studies that participants have participated in and that have been accepted by the researcher.
[1]**p-values of two-sample t-tests** (in brackets) between the control treatment CTRL and treatment FA or MA respectively.
* p < 0.1; ** p < 0.05; *** p < 0.01

Table 4B.2. Robustness checks based on understanding and attention check. Tobit regressions on WTP using the sub-sample that took at most two attempts to answer any of the four control questions correctly and that did not fail the attention check as robustness checks of main and cross-effects. The results are in line with the findings presented in the main text.

	Outcome variable: Participants' stated willingness to pay for protection from a potential extreme season in season 10 (WTP).							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Variables	warning	warning	no-warning	no-warning	no-warning	no-warning	warning	warning
False Alarm Treatment	-19.90* (-42.77 - 2.969)	-22.75* (-46.68 - 1.183)			1.271 (-24.23 - 26.78)	-4.710 (-29.79 - 20.37)		
Missed Alarm Treatment			39.86*** (11.34 - 68.38)	32.41** (3.078 - 61.74)			25.27** (1.957 - 48.58)	25.96** (1.852 - 50.08)
Forecasted Probability[a] (in season 10)	23.40*** (16.78 - 30.02)	23.02*** (15.98 - 30.05)	25.92*** (17.45 - 34.40)	27.71*** (18.93 - 36.49)	25.20*** (17.71 - 32.69)	26.04*** (18.73 - 33.35)	13.66*** (6.928 - 20.39)	13.12*** (6.129 - 20.12)
Individual Controls								
Age		-0.175 (-1.350 - 0.999)		1.688** (0.306 - 3.071)		1.651*** (0.401 - 2.902)		-0.914 (-2.026 - 0.199)
Gender (female = 1)		13.54 (-11.49 - 38.57)		8.827 (-21.96 - 39.61)		40.99*** (15.11 - 66.88)		30.98** (5.139 - 56.82)
Extreme Event Experience[b] (1 = yes, 0 = no)		-90.41 (-271.2 - 90.42)		378.3** (66.81 - 689.9)		245.4* (-24.08 - 514.8)		-118.1 (-270.3 - 34.23)
Damages caused by event[b] base category = No loss								
Less than £1,000		76.54 (-112.7 - 265.8)		-402.8** (-721.6 - -83.90)		-217.9 (-493.3 - 57.50)		122.5 (-38.33 - 283.3)
More than £1,000		107.2 (-83.98 - 298.5)		-358.3** (-686.3 - -30.32)		-260.6* (-542.3 - 21.00)		114.1 (-49.87 - 278.0)
Household Size		-1.722 (-12.16 - 8.715)		5.130 (-6.890 - 17.15)		2.819 (-6.649 - 12.29)		-2.934 (-12.67 - 6.799)
House Owner (yes = 1; no = 0)		0.300 (-29.42 - 30.02)		-23.27 (-58.96 - 12.43)		-13.87 (-44.50 - 16.75)		2.688 (-25.88 - 31.26)
Participants' Living Area[c] base category = rural								
urban		-28.03* (-59.88 - 3.823)		1.786 (-35.98 - 39.55)		19.85 (-13.44 - 53.13)		-17.13 (-48.69 - 14.43)
metropolitan		-22.23 (-58.67 - 14.20)		-5.978 (-47.26 - 35.31)		18.20 (-18.60 - 55.00)		-0.780 (-36.97 - 35.41)
Household Income[d] base category = Less than £1,499								
£1,500 to £2,999		-15.73 (-49.31 - 17.86)		-32.13* (-70.30 - 6.036)		-12.18 (-43.98 - 19.62)		8.031 (-24.25 - 40.31)
£3,000 to £4,499		-13.25 (-51.80 - 25.30)		27.87 (-16.28 - 72.02)		27.76 (-9.228 - 64.74)		0.743 (-35.34 - 36.83)

Table 4B.2. Continued. Robustness checks based on understanding and attention check.

Variables	(1) warning	(2) warning	(3) no-warning	(4) no-warning	(5) no-warning	(6) no-warning	(7) warning	(8) warning
Household Income[d] base category = Less than £1,499								
£4,500 to £5,999		-10.71 (-60.39 - 38.97)		4.513 (-60.94 - 69.96)		-13.71 (-65.67 - 38.26)		15.12 (-36.59 - 66.83)
£6,000 to £7,499		-42.18 (-107.8 - 23.48)		-50.96 (-136.2 - 34.26)		-27.53 (-96.86 - 41.79)		3.312 (-64.08 - 70.71)
Over £7,500		0.373 (-73.51 - 74.25)		41.76 (-46.03 - 129.6)		72.40* (-0.988 - 145.8)		-49.06 (-113.4 - 15.32)
Highest Educational Degree base category = Secondary school								
College		-18.90 (-60.68 - 22.88)		27.23 (-22.33 - 76.78)		34.67 (-8.233 - 77.58)		7.358 (-36.15 - 50.87)
Bachelor's degree		-7.642 (-47.98 - 32.70)		38.40 (-9.508 - 86.30)		43.80** (2.884 - 84.71)		10.24 (-31.30 - 51.78)
Post-graduate degree		-3.311 (-47.76 - 41.14)		42.08 (-11.51 - 95.67)		33.50 (-12.75 - 79.74)		4.377 (-41.83 - 50.58)
Risk Preference (0 = risk averse, 10 = risk liking)		-5.959** (-11.43 - -0.485)		-10.28*** (-16.93 - -3.622)		-8.642*** (-14.61 - -2.672)		-6.336** (-12.00 - -0.671)
Impact of Climate Change (0 = bad, 10 = good)								
on people across the world		-4.866 (-12.85 - 3.117)		1.222 (-8.650 - 11.09)		1.799 (-7.195 - 10.79)		-5.079 (-14.04 - 3.886)
on personal life		1.910 (-7.663 - 11.48)		0.466 (-11.03 - 11.96)		-3.319 (-13.78 - 7.142)		2.140 (-7.386 - 11.67)
Control Questions[e] base category = all correct at first try								
maximum of two tries		13.18 (-12.16 - 38.51)		-8.617 (-38.66 - 21.43)		5.485 (-19.90 - 30.87)		15.94 (-8.596 - 40.47)
Participation Device base category = Desktop computer								
Laptop		10.97 (-20.07 - 42.01)		4.668 (-34.51 - 43.85)		-6.742 (-39.73 - 26.24)		-25.42* (-55.60 - 4.760)
Tablet		-26.75 (-90.59 - 37.08)		-38.10 (-108.1 - 31.85)		-72.89** (-136.3 - -9.522)		-11.53 (-81.89 - 58.83)
Mobile phone		24.97 (-13.29 - 63.23)		28.89 (-16.72 - 74.50)		14.71 (-24.48 - 53.90)		-38.11** (-75.67 - -0.544)
Participation Environment base category = at home, no distractions								
at home, with distractions		5.492 (-42.06 - 53.04)		-22.01 (-79.54 - 35.53)		27.29 (-21.66 - 76.25)		-5.463 (-52.01 - 41.08)
not at home, no distractions		-90.31** (-175.9 - -4.731)		39.64 (-42.14 - 121.4)		31.83 (-29.81 - 93.46)		11.45 (-65.09 - 87.99)

Table 4B.2. Continued. Robustness checks based on understanding and attention check.

Variables	(1) warning	(2) warning	(3) no-warning	(4) no-warning	(5) no-warning	(6) no-warning	(7) warning	(8) warning
Participation Environment base category = at home, no distractions								
not at home, with distractions				-76.08 (-389.0 - 236.8)				-38.24 (-304.7 - 228.2)
Constant	109.6*** (44.39 - 174.9)	179.7*** (78.24 - 281.1)	58.45*** (22.79 - 94.12)	13.45 (-102.1 - 129.0)	62.67*** (30.76 - 94.59)	-13.94 (-107.5 - 79.59)	202.9*** (136.5 - 269.4)	282.9*** (179.8 - 386.1)
Observations	539	469	531	456	561	489	570	497
left-censored	27	23	95	79	111	88	20	16
right-censored	45	38	23	18	13	9	72	63
Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009

Note: All participants that needed more than two attempts to answer the control questions correctly and that failed the attention check are excluded from these robustness checks of our treatment effects. Furthermore, to estimate the 8 models only the relevant sub-sample depending on the treatment and whether participants received a warning in the last season was considered: **Model 1 and 2** test **Hypothesis 1** (sub-sample of treatments ACC and FA that received a warning). **Model 3 and 4** test **Hypothesis 2** (sub-sample of treatments ACC and MA that did not receive a warning). **Model 5 and 6** test **Hypothesis 3** (sub-sample of treatments ACC and FA that did not receive a warning). **Model 7 and 8** test **Hypothesis 4** (sub-sample of treatments ACC and MA that received a warning). Censoring limits for the dependent variable WTP are 0 points as lower and 500 points as upper limit.

[a]**Forecasted Probability** was between 0.6 and 0.85 in the warning case and between 0.15 and 0.4 in the no-warning case. [b]**Extreme event experience** is a binary variable indicated if participants had previously experienced an extreme weather event, for example a drought, flood, heavy thunderstorm etc. **Damages caused by event** refers to any losses that participants might have experienced due to these extreme weather events. [c]**Categories** of participants' living areas: rural = less than 10,000 inhabitants, urban = 10,001 to 100,000 inhabitants and metropolitan = more than 100,001 inhabitants. [d]**Household income** is the monthly disposable income. [e]**Control questions** were part of the instructions to ensure participants' understanding of the experiment. Participants were able to answer each question as many times as they needed to get the correct answer. If the answer was incorrect, the relevant part of the instructions was repeated. Participants that took more than two attempts are excluded from the analysis.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$, Confidence intervals (at the 95% confidence level) are presented in brackets.

Appendix 4C: Details of the results

Table 4C.1. Tobit estimations of the main and cross treatment effects

	Outcome variable: Participants' stated willingness to pay for protection from a potential extreme season in season 10 (WTP)			
Variables	**(1)** warning	**(2)** no-warning	**(3)** no-warning	**(4)** warning
False Alarm Treatment	-29.19***		-1.191	
	(-50.51 - -7.867)		(-26.33 - 23.95)	
Missed Alarm Treatment		33.71**		23.38**
		(7.310 - 60.10)		(1.525 - 45.23)
Forecasted Probability[a] (in season 10)	21.43***	23.00***	22.98***	13.76***
	(15.23 - 27.63)	(15.10 - 30.89)	(15.61 - 30.35)	(7.318 - 20.19)
Constant	132.2***	75.84***	76.32***	205.4***
	(71.50 - 192.8)	(42.44 - 109.2)	(44.68 - 108.0)	(142.4 - 268.5)
Observations	658	638	676	695
left-censored	30	107	131	24
right-censored	60	32	25	96
Prob > chi2	0.000	0.000	0.000	0.000

Note: To estimate the 4 models only the relevant sub-samples depending on the treatment and whether participants received a warning in the last season were considered: **Model 1 tests Hypothesis 1** (sub-sample of treatments CTRL and FA that received a warning). **Model 2 tests Hypothesis 2** (sub-sample of treatments CTRL and MA that did not receive a warning). **Model 3 tests Hypothesis 3** (sub-sample of treatments CTRL and FA that did not receive a warning). **Model 4 tests Hypothesis 4** (sub-sample of treatments CTRL and MA that received a warning). Censoring limits for the dependent variable WTP are 0 points as lower and 500 points as upper limit.

[a] **Forecasted Probability** was between 0.6 and 0.85 in the warning case and between 0.15 and 0.4 in the no-warning case.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$, Confidence intervals (at the 95% confidence level) are presented in brackets.

Table 4C.2. Tobit estimations of the main and cross treatment effects with additional control variables

	Outcome variable: Participants' stated willingness to pay for protection from a potential extreme season in season 10 (WTP)			
Variables	**(1)** warning	**(2)** no-warning	**(3)** no-warning	**(4)** warning
False Alarm Treatment	-34.19***		-4.818	
	(-56.78 - -11.60)		(-30.22 – 20.59)	
Missed Alarm Treatment		30.06**		23.66**
		(2.746 – 57.38)		(1.065 – 46.26)
Forecasted Probability[a]	20.28***	24.76***	24.43***	12.89***
(in season 10)	(13.62 - 26.94)	(16.57 - 32.94)	(17.05 – 31.82)	(6.199 – 19.58)
Individual Controls				
Age	0.305	0.695	1.369**	-0.946*
	(-0.809 - 1.419)	(-0.562 - 1.952)	(0.123 - 2.615)	(-2.001 – 0.109)
Gender	12.49	-0.972	40.53***	31.47**
(female = 1)	(-11.28 – 36.27)	(-30.05 - 28.11)	(14.24 – 66.83)	(7.142 – 55.79)
Extreme Event Experience[b]	-93.86	93.84	29.32	-116.7
(1 = yes, 0 = no)	(-280.5 – 92.81)	(-142.4 - 330.1)	(-195.8 – 254.5)	(-273.2 – 39.91)
Damages caused by event[b]				
base category = No loss				
Less than £1,000	73.38	-112.9	-5.347	135.3
	(-120.9 - 267.6)	(-359.0 - 133.2)	(-238.5 - 227.9)	(-28.14 - 298.7)
More than £1,000	131.3	-34.53	-11.70	132.4
	(-62.95 - 325.6)	(-282.3 - 213.3)	(-245.7 - 222.3)	(-33.22 - 298.1)
Household Size	2.013	1.111	-1.022	-6.576
	(-7.341 - 11.37)	(-10.22 - 12.44)	(-10.52 - 8.475)	(-14.60 - 1.444)
House Owner	-12.51	-11.80	-13.43	-3.222
(yes = 1; no = 0)	(-40.14 – 15.13)	(-44.62 - 21.02)	(-44.24 – 17.39)	(-29.92 – 23.48)
Participants' Living Area[c]				
base category = rural				
urban	-23.35	6.512	39.13**	0.852
	(-52.98 - 6.286)	(-28.42 - 41.45)	(5.545 - 72.72)	(-28.45 - 30.15)
metropolitan	-10.90	1.339	38.62**	8.095
	(-44.88 - 23.07)	(-37.22 - 39.90)	(1.595 - 75.64)	(-25.31 - 41.50)
Household Income[d]				
base category = Less than £1,499				
£1,500 to £2,999	-8.836	-25.20	-20.03	7.774
	(-39.98 - 22.31)	(-60.29 - 9.894)	(-52.25 - 12.19)	(-22.51 - 38.06)
£3,000 to £4,499	-14.79	23.45	0.810	-0.0565
	(-50.59 - 21.00)	(-17.27 - 64.18)	(-36.47 - 38.10)	(-34.09 - 33.98)
£4,500 to £5,999	1.798	4.064	-24.54	26.13
	(-44.68 - 48.27)	(-55.49 - 63.62)	(-74.71 - 25.63)	(-21.89 - 74.15)
£6,000 to £7,499	-31.46	-31.38	-22.95	4.717
	(-95.18 - 32.26)	(-111.9 - 49.13)	(-93.41 - 47.51)	(-58.28 - 67.72)
Over £7,500	-46.94	40.18	68.43*	-44.14
	(-110.5 - 16.66)	(-46.07 - 126.4)	(-1.012 - 137.9)	(-103.6 - 15.35)
Highest Educational Degree				
base category = Secondary school				
College	-13.89	-6.473	24.34	7.753
	(-52.27 - 24.49)	(-51.96 - 39.02)	(-19.69 - 68.36)	(-31.32 - 46.83)
Bachelor's degree	-0.986	11.99	33.96	13.89
	(-38.39 - 36.42)	(-32.17 - 56.16)	(-7.784 - 75.71)	(-24.03 - 51.81)
Post-graduate degree	5.996	22.03	25.18	-2.547
	(-35.41 - 47.40)	(-26.62 - 70.68)	(-21.19 - 71.54)	(-45.06 - 39.96)
Risk Preference	-5.389**	-8.070***	-3.491	-7.765***
(0 = risk averse, 10 = risk liking)	(-10.40 - -0.379)	(-14.14 - -2.002)	(-9.367 – 2.385)	(-12.99 - -2.536)

Table 4C.2. Continued. Tobit estimations of the main and cross treatment effects with additional control variables

Variables	(1) warning	(2) no-warning	(3) no-warning	(4) warning
Impact of Climate Change				
(0 = bad, 10 = good)				
on people across the world	-4.416	-0.613	2.418	-4.376
	(-11.49 - 2.657)	(-9.611 - 8.385)	(-6.405 - 11.24)	(-11.97 - 3.222)
on personal life	3.099	3.573	-1.800	0.209
	(-5.484 - 11.68)	(-7.134 - 14.28)	(-11.98 - 8.376)	(-8.098 - 8.516)
Failed Attention Check	6.273	72.77**	27.67	50.04**
(1=yes, 0=no)	(-40.89 – 53.44)	(14.38 – 131.2)	(-26.23 – 81.57)	(3.715 – 96.36)
Control Questions[e]				
base category = all correct at first try				
maximum of two tries	14.95	-6.542	5.173	16.31
	(-10.07 - 39.98)	(-36.59 - 23.50)	(-22.24 - 32.59)	(-8.262 - 40.88)
more than three tries	-3.040	-23.84	28.13	21.92
	(-42.21 - 36.13)	(-68.44 - 20.76)	(-13.26 - 69.52)	(-17.62 - 61.45)
Participation Device				
base category = Desktop computer				
Laptop	12.40	4.734	-10.41	-14.41
	(-16.90 - 41.69)	(-32.37 - 41.84)	(-43.98 - 23.16)	(-43.54 - 14.73)
Tablet	8.050	-15.88	-46.91	-1.400
	(-46.64 - 62.74)	(-79.18 - 47.42)	(-108.3 - 14.53)	(-62.01 - 59.21)
Mobile phone	29.04	12.96	-2.448	-27.73
	(-6.332 - 64.41)	(-29.14 - 55.06)	(-41.53 - 36.64)	(-62.61 - 7.145)
Participation Environment				
base category = at home, no distractions				
at home, with distractions	8.185	-27.62	8.597	7.815
	(-38.57 - 54.94)	(-79.40 - 24.15)	(-37.84 - 55.03)	(-35.14 - 50.77)
not at home, no distractions	-87.93**	5.066	43.87	15.93
	(-164.5 - -11.38)	(-71.58 - 81.71)	(-19.71 - 107.4)	(-55.40 - 87.26)
not at home, with distractions		-66.87		-30.49
		(-385.5 - 251.8)		(-304.0 - 243.0)
Constant	-372.9	71.11	-15.16	290.0***
	(-1,392 - 646.6)	(-32.75 - 175.0)	(-108.6 - 78.32)	(193.4 - 386.6)
Observations	564	549	591	598
left-censored	26	89	105	19
right-censored	49	24	19	83
Prob > chi2	0.000	0.000	0.000	0.000

Note: To estimate the 4 models only the relevant sub-samples depending on the treatment and whether participants received a warning in the last season were considered: **Model 1 tests Hypothesis 1** (sub-sample of treatments CTRL and FA that received a warning). **Model 2 tests Hypothesis 2** (sub-sample of treatments CTRL and MA that did not receive a warning). **Model 3 tests Hypothesis 3** (sub-sample of treatments CTRL and FA that did not receive a warning). **Model 4 tests Hypothesis 4** (sub-sample of treatments CTRL and MA that received a warning). Censoring limits for the dependent variable WTP are 0 points as lower and 500 points as upper limit.
[a] **Forecasted Probability** was between 0.6 and 0.85 in the warning case and between 0.15 and 0.4 in the no-warning case. [b] **Extreme event experience** is a binary variable indicated if participants had previously experienced an extreme weather event, for example a drought, flood, heavy thunderstorm etc. **Damages caused by event** refers to any losses that participants might have experienced due to these extreme weather events. [c] Categories of **participants' living areas**: rural = less than 10,000 inhabitants, urban = 10,001 to 100,000 inhabitants and metropolitan = more than 100,001 inhabitants. [d] **Household income** is the monthly disposable income. [e] **Control questions** were part of the instructions to ensure participants' understanding of the experiment. Participants were able to answer each question as many times as they needed to get the correct answer. If the answer was incorrect, the relevant part of the instructions was repeated.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$, Confidence intervals (at the 95% confidence level) are presented in brackets.

Table 4C.3. Robustness checks based on frequencies. Tobit regressions on WTP using the accumulated false and missed alarm frequencies of season 1 to 9 instead of treatment variables as robustness checks of our main and cross-effects. The results are in line with the findings presented in the main text.

Variables	**Outcome variable:** Participants' stated willingness to pay for protection from a potential extreme season in season 10 (WTP)			
	(1) Warning	**(2)** Warning	**(3)** No-warning	**(4)** No-warning
False Alarm Rel. Frequency (0 – 1)	-107.2*** (-154.4 - -59.94)	-122.1*** (-171.3 - -72.83)	6.974 (-48.10 - 62.05)	7.371 (-48.00 – 62.74)
Missed Alarm Rel. Frequency (0 – 1)	40.62* (-7.637 - 88.89)	42.71* (-7.460 – 92.89)	130.4*** (73.42 - 187.4)	123.0*** (64.94 – 181.1)
Forecasted Probability[a] (in season 10)	16.00*** (10.74 - 21.27)	15.80*** (10.28 – 21.32)	22.90*** (16.64 - 29.16)	24.90*** (18.55 – 31.25)
Individual Controls				
Age		-0.341 (-1.235 - 0.552)		0.880* (-0.148 - 1.908)
Gender (female = 1)		23.48** (3.810 – 43.15)		24.41** (1.775 – 47.05)
Extreme Event Experience[b] (1 = yes, 0 = no)		-108.9 (-245.2 – 27.52)		150.5 (-33.33 – 334.3)
Damages caused by event[b] base category = No loss				
Less than £1,000		130.2* (-12.33 - 272.8)		-138.6 (-329.4 - 52.29)
More than £1,000		130.2* (-13.84 - 274.3)		-111.0 (-304.2 - 82.08)
Household Size		-2.118 (-9.079 - 4.843)		2.925 (-5.534 - 11.38)
House Owner (yes = 1; no = 0)		-4.519 (-26.88 – 17.85)		-9.194 (-35.03 – 16.65)
Participants' Living Area[c] base category = rural				
urban		-11.76 (-35.93 - 12.40)		14.84 (-13.20 - 42.88)
metropolitan		2.474 (-25.20 - 30.15)		8.412 (-22.61 - 39.43)
Household Income[d] base category = Less than £1,499				
£1,500 to £2,999		2.984 (-22.40 - 28.37)		-22.97 (-50.75 - 4.806)
£3,000 to £4,499		-8.025 (-37.17 - 21.12)		8.571 (-23.12 - 40.26)
£4,500 to £5,999		18.78 (-20.62 - 58.18)		-17.53 (-61.57 - 26.51)
£6,000 to £7,499		-27.12 (-78.99 - 24.76)		-12.24 (-72.69 - 48.22)
Over £7,500		-52.00** (-103.3 - -0.716)		54.44* (-5.565 - 114.4)
Highest Educational Degree base category = Secondary school				
College		-4.259 (-36.87 - 28.35)		11.16 (-25.25 - 47.56)
Bachelor's degree		6.416 (-24.95 - 37.78)		29.77* (-4.981 - 64.52)
Post-graduate degree		-0.808 (-35.52 - 33.90)		32.81* (-5.845 - 71.47)

Table 4C.3. Continued. Robustness checks based on frequencies.

Variables	(1) Warning	(2) Warning	(3) No-warning	(4) No-warning
Risk Preference		-7.450***		-6.669***
(0 = risk averse, 10 = risk liking)		(-11.65 - -3.251)		(-11.52 - -1.817)
Impact of Climate Change (0 = bad, 10 = good)				
on people across the world		-3.625		1.722
		(-9.602 - 2.352)		(-5.481 - 8.924)
on personal life		1.205		0.676
		(-5.703 - 8.112)		(-7.700 - 9.052)
Failed Attention Check		21.05		51.57**
(1=yes, 0=no)		(-17.31 – 59.41)		(3.494 – 99.65)
Control Questions[e] base category = all correct at first try				
maximum of two tries		16.87		1.490
		(-3.488 - 37.24)		(-22.09 - 25.07)
more than three tries		4.093		10.88
		(-28.39 - 36.58)		(-24.14 - 45.91)
Participation Device base category = Desktop computer				
Laptop		0.455		1.146
		(-23.62 - 24.52)		(-27.23 - 29.52)
Tablet		-0.375		-16.90
		(-47.11 - 46.36)		(-67.35 - 33.55)
Mobile phone		4.548		2.925
		(-24.61 - 33.71)		(-29.95 - 35.80)
Participation Environment base category = at home, no distractions				
at home, with distractions		7.125		-4.828
		(-29.66 - 43.91)		(-45.99 - 36.34)
not at home, no distractions		-27.09		29.72
		(-86.71 - 32.52)		(-26.13 - 85.57)
not at home, with distractions		-29.58		-143.8
		(-303.3 - 244.2)		(-456.3 - 168.7)
Constant	198.9***	251.3***	52.88***	-2.104
	(144.4 - 253.4)	(169.1 - 333.6)	(19.69 - 86.07)	(-85.75 - 81.54)
Observations	1,011	877	989	864
left-censored	42	36	182	150
right-censored	119	102	46	35
Prob > chi2	0.000	0.000	0.000	0.000

Note: To estimate Model 1 and 2 (3 and 4) only the sub-sample that (did not) received a warning in the last season was considered. Censoring limits for the dependent variable WTP are 0 points as lower and 500 points as upper limit.

[a]**Forecasted Probability** was between 0.6 and 0.85 in the warning case and between 0.15 and 0.4 in the no-warning case. [b]**Extreme event experience** is a binary variable indicated if participants had previously experienced an extreme weather event, for example a drought, flood, heavy thunderstorm etc. **Damages caused by event** refers to any losses that participants might have experienced due to these extreme weather events. [c]Categories of **participants' living areas**: rural = less than 10,000 inhabitants, urban = 10,001 to 100,000 inhabitants and metropolitan = more than 100,001 inhabitants. [d]**Household income** is the monthly disposable income. [e]**Control questions** were part of the instructions to ensure participants' understanding of the experiment. Participants were able to answer each question as many times as they needed to get the correct answer. If the answer was incorrect, the relevant part of the instructions was repeated.

* p < 0.1; ** p < 0.05; *** p < 0.01, Confidence intervals (at the 95% confidence level) are presented in brackets.

Table 4C.4. Ordinary least square (OLS) estimations of treatment effects on trust as robustness checks. *The results are in line with the findings presented in the main text.*

	Outcome variable: Participants' stated trust level in the forecast of the last season (post-experimental)			
Variables	**(1)**	**(2)**	**(3)** warning	**(4)** no-warning
False Alarm Treatment	-0.523***	-0.555***	-0.729***	-0.435***
	(-0.661 - -0.385)	(-0.704 - -0.407)	(-0.942 - -0.517)	(-0.647 - -0.224)
Missed Alarm Treatment	-0.460***	-0.491***	-0.435***	-0.582***
	(-0.599 - -0.322)	(-0.640 - -0.341)	(-0.642 - -0.229)	(-0.801 - -0.363)
Forecasted Probability[a] (in season 10)		0.0167* (-0.000765 - 0.0342)	0.0304 (-0.0203 - 0.0812)	-0.00745 (-0.0587 - 0.0438)
Individual Controls				
Age		-0.00860***	-0.0119***	-0.00598
		(-0.0143 - -0.00287)	(-0.0200 - -0.00386)	(-0.0142 - 0.00228)
Gender (female = 1)		0.100	0.177*	0.0267
		(-0.0270 - 0.227)	(-0.00205 - 0.357)	(-0.156 - 0.209)
Extreme Event Experience[b] (1 = yes, 0 = no)		0.443	0.0317	0.879
		(-0.573 - 1.460)	(-1.394 - 1.457)	(-0.574 - 2.331)
Damages caused by event[b] base category = No loss				
Less than £1,000		-0.550	-0.172	-0.910
		(-1.603 - 0.503)	(-1.648 - 1.303)	(-2.419 - 0.599)
More than £1,000		-0.443	-0.117	-0.848
		(-1.506 - 0.619)	(-1.603 - 1.369)	(-2.374 - 0.678)
Household Size		-0.00411	-0.0134	-0.00780
		(-0.0497 - 0.0415)	(-0.0757 - 0.0489)	(-0.0756 - 0.0600)
House Owner (yes = 1; no = 0)		0.0955	0.134	0.0712
		(-0.0489 - 0.240)	(-0.0684 - 0.337)	(-0.137 - 0.280)
Participants' Living Area[c] base category = rural				
urban		0.0638	0.0181	0.117
		(-0.0931 - 0.221)	(-0.202 - 0.238)	(-0.109 - 0.343)
metropolitan		-0.0280	-0.00217	-0.0418
		(-0.205 - 0.149)	(-0.254 - 0.249)	(-0.292 - 0.209)
Household Income[d] base category = Less than £1,499				
£1,500 to £2,999		0.0665	0.0495	0.136
		(-0.0945 - 0.227)	(-0.183 - 0.282)	(-0.0892 - 0.361)
£3,000 to £4,499		0.151	0.0644	0.266**
		(-0.0331 - 0.335)	(-0.203 - 0.332)	(0.00827 - 0.523)
£4,500 to £5,999		-0.0710	-0.265	0.105
		(-0.321 - 0.179)	(-0.621 - 0.0899)	(-0.252 - 0.462)
£6,000 to £7,499		0.130	0.125	0.254
		(-0.208 - 0.469)	(-0.343 - 0.593)	(-0.242 - 0.750)
Over £7,500		-0.256	-0.397*	-0.0785
		(-0.596 - 0.0833)	(-0.858 - 0.0641)	(-0.586 - 0.430)
Highest Educational Degree base category = Secondary school				
College		-0.281***	-0.505***	-0.0601
		(-0.488 - -0.0733)	(-0.801 - -0.209)	(-0.353 - 0.233)
Bachelor's degree		-0.118	-0.291**	0.0163
		(-0.316 - 0.0802)	(-0.576 - -0.00585)	(-0.264 - 0.296)
Post-graduate degree		-0.0162	-0.0677	0.0236
		(-0.237 - 0.204)	(-0.384 - 0.249)	(-0.288 - 0.336)
Risk Preference (0 = risk averse, 10 = risk liking)		0.0266*	-0.00652	0.0586***
		(-0.000457 - 0.0536)	(-0.0445 - 0.0315)	(0.0195 - 0.0977)

Table 4C.4. Continued. OLS estimations of treatment effects on trust as robustness checks.

Variables	(1)	(2)	(3) warning	(4) no-warning
Impact of Climate Change				
(0 = bad, 10 = good)				
on people across		0.0493**	0.0612**	0.0456
the world		(0.00976 - 0.0889)	(0.00626 - 0.116)	(-0.0125 - 0.104)
on personal life		-0.0193	-0.0462	0.000507
		(-0.0653 - 0.0267)	(-0.110 - 0.0172)	(-0.0668 - 0.0678)
Failed Attention				
Check		-0.0190	0.183	-0.310
(1=yes, 0=no)		(-0.280 - 0.242)	(-0.170 - 0.537)	(-0.702 - 0.0821)
Control Questions[e]				
base category = all correct at first try				
maximum of two		0.0397	0.134	-0.0378
tries		(-0.0925 - 0.172)	(-0.0510 - 0.319)	(-0.229 - 0.154)
more than three		-0.0626	-0.0646	-0.0671
tries		(-0.268 - 0.143)	(-0.364 - 0.235)	(-0.353 - 0.218)
Participation Device				
base category = Desktop computer				
Laptop		-0.00600	-0.0381	0.0171
		(-0.164 - 0.152)	(-0.258 - 0.181)	(-0.212 - 0.246)
Tablet		0.0604	-0.106	0.171
		(-0.230 - 0.351)	(-0.527 - 0.314)	(-0.234 - 0.576)
Mobile phone		0.0670	0.131	-0.00707
		(-0.121 - 0.255)	(-0.135 - 0.398)	(-0.273 - 0.259)
Participation Environment				
base category = at home, no distractions				
at home, with		-0.0826	-0.279*	0.105
distractions		(-0.316 - 0.151)	(-0.611 - 0.0531)	(-0.226 - 0.436)
not at home, no		-0.128	0.164	-0.249
distractions		(-0.470 - 0.213)	(-0.381 - 0.709)	(-0.692 - 0.195)
not at home, with		0.00156	2.658**	-2.537**
distractions		(-1.759 - 1.762)	(0.178 - 5.138)	(-5.070 - -0.00466)
Constant	4.223***	4.243***	4.659***	3.863***
	(4.125 - 4.320)	(3.806 - 4.680)	(3.928 - 5.389)	(3.205 - 4.521)
Observations	1,921	1,676	845	831
R-squared	0.033	0.064	0.106	0.075
F-Stat:				
False Alarm = Missed Alarm	0.79	0.74	7.71	1.82
P-Value	0.374	0.389	0.006	0.177

Note: Ordinary least squares (OLS) were used to estimate the parameters of the regression models. The dependent variable **Trust Level** is participants' stated level of trust in the last season's forecast as a proxy for their trust in the forecast system. Individuals stated their level of trust by choosing a value from a 6-point Likert scale with 1 meaning "strong mistrust" and 6 "strong trust". To estimate Model 1 and 2, the whole sample was considered. Model 3 only considers participants that received a warning in the last season, while Model 4 only considers participants that did not receive a warning in the last season.

[a]**Forecasted Probability** was between 0.6 and 0.85 in the warning case and between 0.15 and 0.4 in the no-warning case. [b]**Extreme event experience** is a binary variable indicated if participants had previously experienced an extreme weather event, for example a drought, flood, heavy thunderstorm etc. **Damages caused by event** refers to any losses that participants might have experienced due to these extreme weather events. [c]Categories of **participants' living areas**: rural = less than 10,000 inhabitants, urban = 10,001 to 100,000 inhabitants and metropolitan = more than 100,001 inhabitants. [d]**Household income** is the monthly disposable income. [e]**Control questions** were part of the instructions to ensure participants' understanding of the experiment. Participants were able to answer each question as many times as they needed to get the correct answer. If the answer was incorrect, the relevant part of the instructions was repeated.

* p < 0.1; ** p < 0.05; *** p < 0.01, Confidence intervals (at the 95% confidence level) are presented in brackets.

Table 4C.5. Tobit regressions including trust measure and individual controls

	Outcome variable: Participants' stated willingness to pay for protection from a potential extreme season in season 10 (WTP)			
Variables	**(1)** warning	**(2)** no-warning	**(3)** no-warning	**(4)** warning
False Alarm Treatment	-17.37 (-40.59 - 5.856)		-6.272 (-32.25 - 19.71)	
Missed Alarm Treatment		29.81** (1.434 - 58.19)		33.64*** (11.03 - 56.26)
Forecasted Probability[a] (in season 10)	20.01*** (13.41 - 26.60)	25.66*** (17.42 - 33.91)	25.24*** (17.80 - 32.67)	12.76*** (6.114 - 19.40)
Trust Level[b]	26.45*** (17.10 - 35.81)	-1.888 (-12.99 - 9.215)	-3.389 (-13.79 - 7.007)	22.20*** (12.64 - 31.76)
Individual Controls				
Age	0.609 (-0.483 - 1.701)	0.605 (-0.655 - 1.865)	1.187* (-0.0565 - 2.431)	-0.852 (-1.882 - 0.177)
Gender (female = 1)	7.804 (-15.59 - 31.20)	-1.978 (-31.15 - 27.19)	38.34*** (11.88 - 64.81)	30.81** (6.929 - 54.69)
Extreme Event Experience[c] (1 = yes, 0 = no)	-90.38 (-270.2 - 89.47)	90.3 (-143.4 - 324.0)	22.64 (-199.7 - 245.0)	-77.45 (-262.6 - 107.7)
Damages caused by event[c] base category = No loss				
Less than £1,000	68.32 (-118.8 - 255.5)	-111.3 (-354.8 - 132.2)	3.668 (-226.6 - 233.9)	93.14 (-98.06 - 284.3)
More than £1,000	136.5 (-50.95 - 324.0)	-26.87 (-272.0 - 218.3)	-3.773 (-234.8 - 227.3)	92.88 (-99.23 - 285.0)
Household Size	2.173 (-6.927 - 11.27)	-0.375 (-11.70 - 10.95)	-2.092 (-11.54 - 7.356)	-6.416 (-14.20 - 1.363)
House Owner (yes = 1; no = 0)	-15.47 (-42.62 - 11.67)	-13.93 (-46.77 - 18.91)	-12.99 (-44.00 - 18.03)	-2.580 (-28.77 - 23.61)
Participants' Living Area[e] base category = rural				
urban	-19.55 (-48.58 - 9.472)	11.52 (-23.53 - 46.57)	41.32** (7.753 - 74.89)	-0.760 (-29.45 - 27.93)
metropolitan	-1.769 (-35.07 - 31.54)	-4.471 (-43.29 - 34.35)	33.13* (-3.973 - 70.23)	1.052 (-31.56 - 33.66)
Household Income[f] base category = Less than £1,499				
£1,500 to £2,999	-5.495 (-36.23 - 25.24)	-20.60 (-56.10 - 14.91)	-17.35 (-49.85 - 15.15)	12.70 (-17.22 - 42.62)
£3,000 to £4,499	-18.14 (-53.49 - 17.21)	27.17 (-14.12 - 68.46)	-1.130 (-38.81 - 36.55)	4.922 (-28.68 - 38.52)
£4,500 to £5,999	12.59 (-33.43 - 58.61)	11.23 (-49.18 - 71.64)	-26.88 (-78.03 - 24.27)	31.36 (-15.30 - 78.03)
£6,000 to £7,499	-26.52 (-88.27 - 35.22)	-1.335 (-84.56 - 81.89)	-10.15 (-81.35 - 61.06)	3.525 (-57.55 - 64.60)
Over £7,500	-43.31 (-105.0 - 18.33)	49.82 (-43.48 - 143.1)	68.47* (-1.685 - 138.6)	-29.76 (-87.53 - 28.00)
Highest Educational Degree base category = Secondary school				
College	-2.554 (-40.17 - 35.06)	-10.60 (-56.59 - 35.40)	21.61 (-22.87 - 66.10)	23.93 (-14.39 - 62.25)
Bachelor's degree	12.14 (-24.57 - 48.84)	6.080 (-38.94 - 51.10)	32.69 (-9.597 - 74.97)	25.59 (-11.38 - 62.57)
Post-graduate degree	3.822 (-36.62 - 44.26)	13.57 (-35.64 - 62.79)	21.79 (-25.13 - 68.71)	-4.614 (-46.29 - 37.06)
Risk Preference (0 = risk averse, 10 = risk liking)	-4.678* (-9.635 - 0.279)	-7.577** (-13.71 - -1.449)	-2.224 (-8.168 - 3.719)	-7.433*** (-12.54 - -2.324)

Table 4C.5. Continued. Tobit regressions including trust measure and individual controls

Variables	(1) warning	(2) no-warning	(3) no-warning	(4) warning
Impact of Climate Change (0 = bad, 10 = good)				
on people across the world	-5.242	1.773	4.476	-4.791
	(-12.36 - 1.871)	(-7.289 - 10.83)	(-4.421 - 13.37)	(-12.35 - 2.768)
on personal life	3.590	0.546	-1.818	1.055
	(-4.964 - 12.14)	(-10.21 - 11.30)	(-11.98 - 8.344)	(-7.226 - 9.336)
Failed Attention Check (1=yes, 0=no)	-4.619	61.42**	25.03	45.42*
	(-52.08 - 42.84)	(2.555 - 120.3)	(-30.05 - 80.11)	(-0.487 - 91.34)
Control Questions[g] base category = all correct at first try				
maximum of two tries	8.979	-7.392	0.721	13.04
	(-15.57 - 33.53)	(-37.84 - 23.06)	(-26.89 - 28.34)	(-11.03 - 37.11)
more than three tries	2.325	-19.93	29.80	30.56
	(-36.70 - 41.35)	(-64.97 - 25.11)	(-12.24 - 71.83)	(-8.822 - 69.95)
Participation Device base category = Desktop computer				
Laptop	12.72	-0.0256	-15.39	-13.05
	(-16.09 - 41.52)	(-37.25 - 37.20)	(-49.19 - 18.42)	(-41.69 - 15.60)
Tablet	11.13	-25.56	-51.82*	0.529
	(-41.91 - 64.18)	(-88.83 - 37.70)	(-113.6 - 9.957)	(-58.22 - 59.28)
Mobile phone	33.17*	-2.873	-14.57	-33.70*
	(-1.880 - 68.22)	(-45.16 - 39.42)	(-53.98 - 24.84)	(-67.97 - 0.556)
Participation Environment base category = at home, no distractions				
at home, with distractions	3.734	-16.58	12.97	10.52
	(-42.04 - 49.51)	(-69.13 - 35.97)	(-33.15 - 59.09)	(-31.30 - 52.35)
not at home, no distractions	-99.75**	2.580	44.69	4.409
	(-176.8 - -22.72)	(-73.24 - 78.40)	(-18.27 - 107.6)	(-64.54 - 73.36)
not at home, with distractions		-85.52		-90.19
		(-401.7 - 230.7)		(-354.6 - 174.3)
Constant	29.50	94.96	4.233	180.2***
	(-74.79 - 133.8)	(-18.78 - 208.7)	(-97.44 - 105.9)	(75.90 - 284.4)
Observations	545	526	569	577
left-censored	24	83	100	18
right-censored	48	23	18	80
Prob > chi2	0.0000	0.0003	0.0000	0.0000

Note: To estimate the 4 models only the relevant sub-sample depending on the treatment and whether participants received a warning in the last season was considered: **Model 1 tests Hypothesis 1** (sub-sample of treatments CTRL and FA that received a warning). **Model 2 tests Hypothesis 2** (sub-sample of treatments CTRL and MA that did not receive a warning). **Model 3 tests Hypothesis 3** (sub-sample of treatments CTRL and FA that did not receive a warning). **Model 4 tests Hypothesis 4** (sub-sample of treatments CTRL and MA that received a warning). Censoring limits for the dependent variable WTP are 0 points as lower and 500 points as upper limit.

[a]**Forecasted Probability** was between 0.6 and 0.85 in the warning case and between 0.15 and 0.4 in the no-warning case. [b]**Trust Level** is participants' stated level of trust in the last season's forecast as a proxy for their trust in the forecast system. Individuals stated their level of trust by choosing a value from a 6-point Likert scale with 1 meaning "strong mistrust" and 6 "strong trust". [c]**Extreme event experience** is a binary variable indicated if participants had previously experienced an extreme weather event, for example a drought, flood, heavy thunderstorm etc. [d]**Damages caused by event** refers to any losses that participants might have experienced due to these extreme weather events. [e]Categories of **participants' living areas**: rural = less than 10,000 inhabitants, urban = 10,001 to 100,000 inhabitants and metropolitan = more than 100,001 inhabitants. [f]**Household income** is the monthly disposable income. [g]**Control questions** were part of the instructions to ensure participants' understanding of the experiment. Participants were able to answer each question as many times as they needed to get the correct answer. If the answer was incorrect, the relevant part of the instructions was repeated.

* p < 0.1; ** p < 0.05; *** p < 0.01, Confidence intervals (at the 95% confidence level) are presented in brackets.

Table 4C.6. Analysis of the relevance of treatment effects. Tobit regressions including probability treatment interaction terms.

	Outcome variable: Participants' stated willingness to pay for protection from a potential extreme season in season 10 (WTP)							
Variables	**(1)** warning	**(2)** warning	**(3)** no-warning	**(4)** no-warning	**(5)** no-warning	**(6)** no-warning	**(7)** warning	**(8)** warning
False Alarm Treatment	-9.465 (-129.4 - 110.4)	-47.47 (-175.8 - 80.89)			-7.246 (-64.26 - 49.77)	-12.47 (-70.03 - 45.09)		
Missed Alarm Treatment			27.88 (-34.28 - 90.05)	33.54 (-32.18 - 99.26)			195.1*** (71.27 - 319.0)	155.5** (26.65 - 284.4)
Forecasted Probability[a] (in season 10)	22.43*** (13.82 - 31.04)	19.59*** (10.27 - 28.91)	22.19*** (11.08 - 33.30)	25.24*** (13.63 - 36.84)	22.04*** (11.18 - 32.90)	23.20*** (12.10 - 34.30)	22.62*** (13.63 - 31.60)	19.89*** (10.41 - 29.38)
Interaction Term False Alarm x Forecasted Probability	-2.074 (-14.48 - 10.33)	1.389 (-11.82 - 14.60)			1.747 (-13.02 - 16.51)	2.198 (-12.64 - 17.03)		
Interaction Term Missed Alarm x Forecasted Probability			1.631 (-14.14 - 17.41)	-0.977 (-17.79 - 15.83)			-18.07*** (-30.89 - - 5.244)	-13.86** (-27.19 - - 0.527)
Individual Controls								
Age		0.306 (-0.808 - 1.421)		0.698 (-0.560 - 1.956)		1.371** (0.125 - 2.617)		-0.967* (-2.019 - 0.0853)
Gender (female = 1)		12.45 (-11.33 - 36.22)		-1.005 (-30.09 - 28.08)		40.46*** (14.16 - 66.76)		31.54** (7.277 - 55.79)
Extreme Event Experience[b] (1 = yes, 0 = no)		-94.78 (-281.6 - 92.07)		95.04 (-142.2 - 332.3)		27.76 (-197.5 - 253.0)		-119.7 (-275.8 - 36.45)
Damages caused by event[b] base category = No loss								
Less than £1,000		74.17 (-120.2 - 268.5)		-114.5 (-362.2 - 133.2)		-3.317 (-236.8 - 230.2)		138.8* (-24.18 - 301.7)
More than £1,000		132.4 (-62.12 - 326.9)		-35.33 (-283.5 - 212.9)		-10.98 (-244.9 - 223.0)		131.7 (-33.41 - 296.9)
Household Size		1.964 (-7.401 - 11.33)		1.086 (-10.25 - 12.43)		-0.961 (-10.47 - 8.545)		-6.588 (-14.58 - 1.402)
House Owner (yes = 1; no = 0)		-12.46 (-40.10 - 15.17)		-11.85 (-44.68 - 20.99)		-13.35 (-44.17 - 17.46)		-3.575 (-30.20 - 23.05)
Participants' Living Area[c] base category = rural								
urban		-23.40 (-53.03 - 6.229)		6.566 (-28.38 - 41.51)		39.25** (5.657 - 72.84)		-2.467 (-31.85 - 26.92)
metropolitan		-10.92 (-44.89 - 23.05)		1.155 (-37.53 - 39.84)		39.00** (1.891 - 76.11)		3.996 (-29.54 - 37.54)

Table 4C.6. Continued. Analysis of the relevance of treatment effects. *Tobit regressions including probability treatment interaction terms.*

Variables	(1) warning	(2) warning	(3) no-warning	(4) no-warning	(5) no-warning	(6) no-warning	(7) warning	(8) warning
Household Income[d] base category = Less than £1,499								
£1,500 to £2,999		-8.900 (-40.04 - 22.24)		-25.34 (-60.53 - 9.838)		-19.59 (-51.94 - 12.77)		8.686 (-21.53 - 38.90)
£3,000 to £4,499		-14.89 (-50.69 - 20.92)		23.39 (-17.36 - 64.13)		0.901 (-36.38 - 38.19)		0.878 (-33.07 - 34.83)
£4,500 to £5,999		1.510 (-45.04 - 48.06)		4.129 (-55.44 - 63.70)		-24.23 (-74.44 - 25.99)		24.50 (-23.40 - 72.41)
£6,000 to £7,499		-31.70 (-95.44 - 32.05)		-31.35 (-111.9 - 49.17)		-22.65 (-93.12 - 47.83)		6.682 (-56.15 - 69.51)
Over £7,500		-47.33 (-111.0 - 16.37)		40.29 (-45.98 - 126.6)		68.32* (-1.120 - 137.8)		-42.00 (-101.3 - 17.35)
Highest Educational Degree base category = Secondary school								
College		-13.43 (-52.05 - 25.19)		-6.280 (-51.89 - 39.33)		24.03 (-20.04 - 68.10)		5.064 (-34.00 - 44.12)
Bachelor's degree		-0.543 (-38.18 - 37.09)		12.09 (-32.10 - 56.29)		33.73 (-8.036 - 75.50)		12.62 (-25.22 - 50.46)
Post-graduate degree		6.433 (-35.17 - 48.04)		22.19 (-26.54 - 70.91)		24.95 (-21.43 - 71.33)		-4.088 (-46.51 - 38.33)
Risk Preference (0 = risk averse, 10 = risk liking)		-5.365** (-10.38 - - 0.351)		-8.082*** (-14.15 - - 2.011)		-3.478 (-9.353 - 2.398)		-7.823*** (-13.04 - - 2.610)
Impact of Climate Change (0 = bad, 10 = good)								
on people across the world		-4.415 (-11.49 - 2.656)		-0.631 (-9.634 - 8.372)		2.482 (-6.350 - 11.32)		-4.436 (-12.01 - 3.140)
on personal life		3.096 (-5.486 - 11.68)		3.613 (-7.116 - 14.34)		-1.876 (-12.06 - 8.311)		-0.0530 (-8.341 - 8.235)
Failed Attention Check (1=yes, 0=no)		6.217 (-40.94 - 53.38)		72.48** (13.87 - 131.1)		28.07 (-25.90 - 82.03)		49.17** (2.979 - 95.36)
Control Questions[e] base category = all correct at first try								
maximum of two tries		14.89 (-10.14 - 39.92)		-6.473 (-36.54 - 23.60)		5.238 (-22.18 - 32.65)		16.34 (-8.155 - 40.84)
more than three tries		-3.338 (-42.60 - 35.93)		-23.99 (-68.67 - 20.69)		28.38 (-13.04 - 69.79)		23.72 (-15.73 - 63.17)

Table 4C.6. Continued. Analysis of the relevance of treatment effects. *Tobit regressions including probability treatment interaction terms.*

Variables	(1) warning	(2) warning	(3) no-warning	(4) no-warning	(5) no-warning	(6) no-warning	(7) warning	(8) warning
Participation Device base category = Desktop computer								
Laptop		12.28 (-17.04 - 41.59)		4.753 (-32.35 - 41.86)		-10.45 (-44.01 - 23.11)		-14.82 (-43.88 - 14.23)
Tablet		8.127 (-46.56 - 62.81)		-16.02 (-79.37 - 47.33)		-46.71 (-108.1 - 14.72)		-1.917 (-62.33 - 58.50)
Mobile phone		29.08 (-6.285 - 64.45)		13.03 (-29.09 - 55.16)		-2.733 (-41.86 - 36.40)		-27.22 (-62.00 - 7.570)
Participation Environment base category = at home, no distractions								
at home, with distractions		8.188 (-38.56 - 54.93)		-27.85 (-79.78 - 24.08)		8.970 (-37.53 - 55.47)		7.605 (-35.22 - 50.43)
not at home, no distractions		-88.32** (-164.9 - - 11.70)		5.037 (-71.61 - 81.68)		44.26 (-19.36 - 107.9)		18.95 (-52.28 - 90.18)
not at home, with distractions				-67.90 (-387.0 - 251.2)				-26.46 (-299.2 - 246.2)
Constant	122.7*** (39.67 - 205.7)	170.9*** (57.61 - 284.1)	78.69*** (35.44 - 121.9)	69.36 (-38.78 - 177.5)	79.63*** (37.38 - 121.9)	-11.20 (-108.4 - 86.00)	121.3*** (34.62 - 208.0)	229.2*** (116.6 - 341.9)
Observations	658	564	638	549	676	591	695	598
left-censored	30	26	107	89	131	105	24	19
right-censored	60	49	32	24	25	19	96	83
Prob > chi2	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000

Note: To estimate the 8 models only the relevant sub-sample depending on the treatment and whether participants received a warning in the last season was considered: **Model 1 and 2** test **Hypothesis 1** (sub-sample of treatments CTRL and FA that received a warning). **Model 3 and 4** test **Hypothesis 2** (sub-sample of treatments CTRL and MA that did not receive a warning). **Model 5 and 6** test **Hypothesis 3** (sub-sample of treatments CTRL and FA that did not receive a warning). **Model 7 and 8** test **Hypothesis 4** (sub-sample of treatments CTRL and MA that received a warning). Censoring limits for the dependent variable WTP are 0 points as lower and 500 points as upper limit.

[a]**Forecasted Probability** was between 0.6 and 0.85 in the warning case and between 0.15 and 0.4 in the no-warning case. [b]**Extreme event experience** is a binary variable indicated if participants had previously experienced an extreme weather event, for example a drought, flood, heavy thunderstorm etc. **Damages caused by event** refers to any losses that participants might have experienced due to these extreme weather events. [c]Categories of **participants' living areas**: rural = less than 10,000 inhabitants, urban = 10,001 to 100,000 inhabitants and metropolitan = more than 100,001 inhabitants. [d]**Household income** is the monthly disposable income. [e]**Control questions** were part of the instructions to ensure participants' understanding of the experiment. Participants were able to answer each question as many times as they needed to get the correct answer. If the answer was incorrect, the relevant part of the instructions was repeated.

* p < 0.1; ** p < 0.05; *** p < 0.01, Confidence intervals (at the 95% confidence level) are presented in brackets.

Table 4C.7. Effect sizes of difference in forecasted probability

Warning case – forecast probabilities 0.6 to 0.85					
Treatment	*N 0.6*	*Mean WTP (SD)*	*N 0.85*	*Mean WTP (SD)*	*Cohens' d*
CTRL	58	284 (102)	56	384 (118)	0.91 [0.52 – 1.29]
FA	50	256 (110)	55	365 (92)	1.09 [0.67 – 1.49]
MA	56	346 (145)	57	375 (126)	0.21 [-0.16 – 0.58]
No-warning case – forecast probabilities 0.15 to 0.4					
	N 0.15	*Mean WTP (SD)*	*N 0.4*	*Mean WTP (SD)*	*Cohens' d*
CTRL	54	122 (120)	45	219 (108)	0.85 [0.43 – 1.26]
FA	67	129 (127)	57	218 (134)	0.69 [0.32 – 1.05]
MA	40	163 (157)	57	231 (139)	0.47 [0.06 – 0.88]

Note: N denotes the number of observations that received the stated probability as forecast. Cohen's d is used to calculate the effect size for the difference between mean WTP of the two group that the forecast for an extreme season stating the lowest probability and the group that received the forecast stating the highest probability. Standard deviations (SD) are presented in brackets. Confidence intervals (at the 95% confidence level) are presented in square brackets.

Appendix 4D: Pre-registration of data analysis

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

Research objective: Assess if the experience of false alarm-prone or miss-prone forecasts decrease individuals' willingness to respond to probabilistic warnings of extreme climate in comparison to accurate forecasts.

H1: Willingness to pay for protection (WTP) if a warning is received is, on average, lower if the forecast system is false alarm-prone than if the forecast system is accurate.

H2: WTP if no warning is received is, on average higher if the forecast system is miss-prone than if the forecast system is accurate.

H3: False alarm-prone forecast systems lead to a loss of credibility of forecasts in general, such that WTP if no warning is received is, on average, higher if the forecast system is false alarm-prone than if the forecast system is accurate.

H4: Miss-prone forecast systems lead to a loss of credibility of forecasts in general, such that WTP if a warning is received is, on average, lower if the forecast system is miss-prone than if the forecast system is accurate.

3) Describe the key dependent variable(s) specifying how they will be measured.

Over 10 rounds, individuals receive probabilistic warnings that the next season could be extreme or not (without knowing the true likelihood). We elicit individuals' WTP for protection from the loss of their entire bonus payment in case that the season turns out to be extreme. Paying for protection covers the full potential loss. WTP is elicited with the BDM method based on Becker, DeGroot and Marschak (1964). Key dependent variable (DV) is individuals' stated WTP for protection in the last round of the game (Round 10).

4) How many and which conditions will participants be assigned to?

We run three treatments that differ in the accuracy of the forecast model: (1) accurate, (2) false alarm-prone and (3) miss-prone. In (1) the probabilities presented as forecast are the true underlying probabilities. In (2) the forecast model is designed such that the probability to experience an extreme season is overrated and in (3) underrated (except the last round where true probabilities are reported as forecast in all three treatments). Participants do not know how accurate the shown forecasts are. Whether the season is extreme or normal is determined in all treatments based on the true underlying probabilities at the end of each round by the computer. For the analysis, the sample per treatment is split in the sub-sample that receives a warning of extreme conditions in the last round and the sub-sample that does not.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We run Tobit regressions with individuals' average WTP in the last round as DV. The focus is on a between-treatment analysis. Treatment dummies are "False Alarm" = 1 if treatment was false-alarm prone model, and 0 otherwise. "Miss" = 1 if model was miss-prone, 0 otherwise. The base category for both variables is the accurate model treatment.

To test H1-H4 different models will be specified:

H1: Sub-sample of treatment 1 and 2 and participants that received a warning in the last round. Treatment dummy for treatment 2 "False Alarm".

H2: Sub-sample of treatment 1 and 3 and participants that did not receive a warning in the last round. Treatment dummy for treatment 3 "Miss".

H3: Sub-sample of treatment 1 and 2 and participants that did not receive a warning in the last round. Treatment dummy for treatment 2 "False Alarm".

H4: Sub-sample of treatment 1 and 3 and participants that received a warning in the last round. Treatment dummy for treatment 3 "Miss".

Two sets of control variables: (1) basic (no potential effect of treatment on answers) includes age, gender, disposable household income, household size, education, ownership of housing (owner or tenant), living area (rural, urban, metropolitann), used device (mobile, desktop, laptop, tablet), understanding of the game (1, if less than or exactly two attempts needed to answer the 4 control questions correctly; 0 otherwise), past experience of financial damage due to extreme weather events (1, if individuals choose one or more of the presented extreme event types, 0 otherwise), size of the experienced damage (dummy coding based on categorical answers from less than £1,000 to more than £35,000) and location of participation (dummy coding based on categorical answers). (2) extended (potential effect of treatment on answers) includes risk preference (elicited through survey question, Dohmen 2010), expectation of future financial damage due to extreme weather events (dummy coding based on Likert scale 1, very likely, to 5, very unlikely), belief of climate change's impact on people globally and own personal life (dummy coding based on Likert scale 0, extremely bad, to 10, extremely good) and attention check (1, if correct answer; 0, otherwise).

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Since the DV WTP is bounded between 0 and 500 points, the dataset will not contain any extreme outliers. Thus, no observations will be excluded based on the DV. For robustness-checks, we exclude subjects that took more than two attempts to answer any of the four control questions correctly and that failed the attention check. Subjects that return their submission ("drop out") at any point are excluded from the sample.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We are aiming for 2,000 observations in total, which are evenly distributed amongst the three treatments and two sub-samples (with/without warning in the last round). Our sample size is restricted by budget and we have to terminate data elicitation prior to reaching our aimed for numbers if we run out of budget. We run two separate sessions via the online platform Prolific for “female” and “male” participants to balance the sample.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Exploratory analysis: (1) If we find evidence for cross effects (Hypotheses 3 and 4), we will test if cross effects are bigger/smaller than the two main effects. (2) Analyse effect of false alarm experience and miss experience based on the frequency of experienced false alarms and misses instead of the forecast model/ treatments. (3) Analysis of individuals’ WTP development over the rounds.

“trust in forecast” as mediator variable: We analyse if trust has a mediating effect on the effect of “False alarm” and “Miss” on WTP. “trust” is elicited with a 6-point Likert-scale as part of the post-experimental questionnaire.

Appendix 4E: Instructions of the experiment

Additional information and explanations of experimental processes are provided in italics.

Horizontal lines _____ mark the switch to the next page/step of the experiment.

Instructions of the forecast experiment

Study name: Study on seasonal climate forecasts

Pre-study information provided on Prolific. Based on this information subjects decide to participate or not:

We would like to invite you to participate in a research study led by Osnabrueck University (Germany). This study is part of a doctoral thesis that examines individuals' responses to environmental change.

This study consists of two sections. In the first section, you can earn **up to £5.00 as a bonus payment**. The exact size of your bonus payment (between £0.00 and £5.00) depends on the decisions that you make during the first section of the study and chance.

This first section of the study takes approximately 15 minutes. You will receive detailed instructions regarding the decision-making task and the bonus payment and must answer four comprehension questions before making your decisions.

The second section is a short survey, which takes about 5 minutes. You will be expected to answer some questions regarding your experiences related to the topic of the study and demographic questions. Your answers to the survey <u>will not</u> influence the size of your bonus.

You are only eligible for both the reward for submission and any bonus payment if you finish both the first and second section of the study.

If individuals wished to participate in the study, the transfer of subjects to the experimental software SoPHIE was combined with an automated integration of the individuals' ProlificIDs.

Declaration of Consent:

All of your responses and decisions made in this study will be treated confidentially. All information collected will be used in an aggregated form for academic publications and educational purposes only. The dataset will be anonymised after data collection, by permanently deleting the Prolific IDs from the dataset.

Your participation in this study is voluntary. You may, if you so choose, withdraw from this study by returning your submission at any time before receiving any payments. You will not be asked for any reason to withdraw. If you return your submission or do not complete the study, you forfeit receiving any and all payments. If you return your submission or do not complete the study, we will retain the information you have given thus far, unless you message us via the Prolific "contact researcher" services with explicit instructions to delete it.

By clicking on **"I agree and wish to participate."**, you declare that you understand the information stated above, along with your rights and commitments during your participation in this study. You also understand that you can return your submission and withdraw your participation at any time, giving up payment related to your participation in the study.

If you have any questions or concerns regarding this study, please contact us through the "contact researcher" services provided by Prolific.

If you are willing to take part in this study, please click the following button to give your consent.

Button to click: I agree and wish to participate.

Welcome

Thank you for agreeing to participate in this study about seasonal forecasting.

In the first part of this study, you make decisions in 10 consecutive rounds, which each represents one season. The amount you receive as your bonus payment will depend on chance and the decisions you make during one of the 10 seasons, chosen at random. Thus, any one of the 10 decisions you make will be randomly used to determine the size of your bonus payment. The size of the bonus payment can be any amount between £0.00 and £5.00.

During the study, we use points instead of GBP, with the exchange rate being: £1.00 = 100 points.

Section 1 - Seasonal Forecast

At the start of each of the 10 seasons, you have 500 points. Each of the 10 seasons can be of two different types: "**extreme**" or "**normal**" climate conditions. You will **not** know the precise risk that the coming season is of extreme or normal conditions. Instead, you will receive a **forecast of the likelihood** that the coming season will be extreme or normal.

Before the first season, you will be randomly assigned a forecast model, which will generate **all** forecasts you receive for all 10 seasons. You will not be informed of how accurate (or inaccurate) the forecasts generated by your assigned model are.

In every season, you are at risk of losing all of your 500 points through extreme climate conditions. Extreme climate conditions could come in the form of heatwaves, droughts, heavy storms and flash flooding events. However, you can protect yourself from extreme climate-related losses by paying with your points for protection from extreme climate conditions.

Example of a seasonal forecast:

Extreme Season	Normal Season
50%	50%

The forecasted likelihood for an extreme season is 50% and 50% for a normal season. This forecast predicts the coming season will turn out to be extreme in 50 out of 100 cases and in 50 out of 100 cases this season will turn out to be normal. In case that the season is extreme, you will lose all of your points if you do not pay enough points for protection.

Section 1 – Payment for Protection

Once you receive the forecast for a given season, you may decide how many of your 500 points you are prepared to pay for protection from extreme climate conditions. You can state any amount between 0 and 500 points:

- Stating **0 points** means that you are **not prepared to pay anything** for protection.
- Stating **500 points** means that you are **prepared to pay all** of your points for protection.

The amount that you state indicates the maximum number of points you are prepared to pay for protection, not the price that protection will cost.

Once you have completed all 10 seasons, the price, in points, for protection will be randomly generated. The randomly generated price can be any amount between 0 and 500 points.

- If you have nominated to pay **equal or more** than the randomly determined price, **you will pay for protection** at the randomly generated price.
- If you nominate to pay **less** than the randomly generated price, **you will not buy protection** and thus you are not protected from extreme climate-related losses.

All prices for protection, between 0 and 500, are equally likely and none of the events during the 10 seasons will influence the randomly determined price set by the computer.

At the end of each season, the climate condition (extreme or normal) of the past season is revealed to you. Please remember that the events in one season will not affect any other season. The risk of extreme climate conditions varies from season to season. The risk of extreme conditions in a given season is independent from the previous seasons.

Section 1 – Payment for Protection

At the end of the study, **one** of the 10 seasons is chosen at random. The randomly chosen season will be used to determine how many points you keep (and are converted into your bonus payment). This means that **only one of your 10 decisions** regarding the points that you are prepared to pay for protection is relevant for your bonus payment.

If you are not prepared to pay enough points for protection in a season that is extreme, you will lose all of your points due to extreme climate-related losses. However, if you are prepared to pay enough points for protection, you will pay the price for protection and are protected from any losses due to an extreme season.

If a season was normal, you will not experience any season-related losses and will keep all of your 500 points. However, if you have stated that you are prepared to pay enough points for protection (i.e. an amount of points above the randomly determined price for protection) in a season that is normal, you have spent your points on unnecessary protection. Thus, you forfeit part of your points by paying for unnecessary protection.

Section 1 – Determination of Bonus Payment

Here are four scenarios that describe how the amount of points you keep for your bonus payment is determined depending on the following factors:

- the amount of points you are prepared to pay for protection in the randomly chosen season
- the type of climate conditions in that season
- the randomly determined price for protection

1. You are prepared to pay more than or equal to the price for protection, and the season was extreme:

   Since you are prepared to pay **more than/equal** to the price for protection, you will pay **the price** (not your stated amount) and receive protection from losses due to the extreme season.

   You keep: 500 points minus “the price for protection”

2. You are prepared to pay less than the price for protection and the season was extreme:

   Since you are prepared to pay **less** than the price for protection, you have not met the criteria to be protected from losses due to the extreme season. Therefore, you will not pay for protection, and lose all of your points.

   You keep: **0 points**

3. You are prepared to pay more than or equal to the price for protection and the season was normal:

   Since you are prepared to pay **more than/equal** to the price for protection, you will pay **the price** (not your stated amount) and receive protection. However, as the season is normal, the protection is unnecessary.

   You keep: 500 points minus “the price for protection”

4. You are prepared to pay less than the price for protection and the season was normal:

   Since you are prepared to pay **less** than the price, you will not pay for protection and you are not protected. As the season is normal however, you will not incur any extreme season losses, and keep all of your points.

   You keep: **500 points**

Section 1 – Determination of Bonus Payment

Here are two examples to **demonstrate**:

Example 1: The randomly chosen season was extreme, and therefore if you do not pay for protection, you will lose all of your points. If you were prepared to pay 250 points in this season and the price for protection is 300 points, you are not able to pay for protection because your stated amount of 250 points is lower than the price of 300 points. Since the season was extreme, it follows that you lose all of your points.

In this scenario, the amount you were prepared to pay is <u>lower</u> than the randomly determined price for protection and the season was <u>extreme</u>.

Example 2: The randomly chosen season is a normal season. If you were prepared to pay 250 points this season and the price for protection is 200 points, you will pay 200 points for protection. You keep the remaining 300 points. If you had not paid for protection, you would have kept the full 500 points.

In this scenario, the amount you were prepared to pay is <u>higher</u> than the randomly determined price for protection and the season was <u>normal</u>.

Section 1 – Comprehension Questions

Please answer the following comprehension questions. You have to answer each of the questions before you can proceed with the study. If you get a question wrong, you will be able to re-read the explanation and answer the question again.

Control questions were programmed on separate pages in SoPHIE:

1. [CQ1] How many of the 10 seasons are relevant to determine the size of your bonus payment?

    - *Incorrect:* All of the 10 seasons are jointly relevant to determine how many points I keep for my bonus payment.

- *Correct:* One of the 10 seasons is randomly chosen to determine how many points I keep for my bonus payment.
- *(if incorrect is chosen)* **Explanation**: The computer randomly choses one of the 10 seasons at the end of the study. The amount of points you keep and thus the size of your bonus payment depends only on the amount you are prepared to pay for protection and the price for protection in this one randomly chosen season, and whether this randomly chosen season was extreme or not.

2. [CQ2] When do you get to know whether a specific season was extreme or normal?

- *Correct:* Whether a specific season was extreme or normal is revealed at the end of each season after I stated the amount that I am prepared to pay for protection in that given season.
- *Incorrect:* The forecast at the beginning of each season tells me with certainty whether the next season will be extreme or normal before I state the amount I am prepared to pay for protection.
- *(if incorrect is chosen)* **Explanation:** The forecast only tells you the likelihood that the season will be extreme or normal. You learn whether the season was extreme or normal only after you have stated the amount that you are prepared to pay for protection, thus in the last step of each season.

3. [CQ3] Is this statement true or false:

Regardless of the forecast model I am assigned, the forecasted likelihood for an extreme season is always equal to the actual risk that the season will be extreme.

- Incorrect: True
- Correct: False
- *(if incorrect is chosen)* **Explanation:** The forecast model you are assigned to generates an estimation of the risk that the coming season is extreme. However,

these estimated risks (i.e. the forecasted likelihoods) might be inaccurate and are not necessarily equal to the actual risk that the season will be extreme.

4. [CQ4] Imagine the following scenario:

At the end of the study, season 4 is chosen as the bonus-relevant season for Person A. Person A stated they would be prepared to pay 250 points in season 4. The computer determined that season 4 was extreme, and the price for protection is 300 points.

How many points would Person A keep in this scenario?

Points of Person A in whole numbers:

- Correct: 0 points
- if correct number: 0 is correct.
- *if incorrect number:* **Explanation:** Your answer is not correct. The stated amount of Person A is lower than the price of protection. Thus, Person A does not pay for protection, and experiences the loss of all points due to the extreme season.

Section 1 - Start of the 10 seasons

You will now start with the first of the 10 seasons. Please read the forecast before every season and make your decision about the amount that you are prepared to pay for protection.

Remember that the 10 seasons are independent from each other, and all decisions and outcomes affect only that single season. Decisions and outcomes in that single season do not affect the climate conditions and points you could obtain in the following seasons. Please base every decision that you make in the next 10 seasons solely on your own personal preferences. <u>There are no right or wrong decisions.</u>

Please click on “Continue” to start the first season.

Information provided in […] and the forecast probabilities Y and (1-Y) were automatically updated for each season 1 to 10:

Forecast for season *[1 to 10]*:

Extreme Season	Normal Season
Y%	(1-Y)%

Text A: Normal text was shown if forecast probability Y for extreme was < 50%:

The forecasted likelihood for an extreme season is Y% and (1-Y)% for a normal season.

This forecast predicts the coming season will turn out to be extreme in Y out of 100 cases and in (1-Y) out of 100 cases this season will turn out to be normal. In case that the season is extreme, you will lose all of your points if you do not pay enough points for protection.

Text B: Warning text was only shown if the forecast probability Y for extreme was > 50%. The threshold for a warning forecast was unknown to participants, they were only shown the additional message on screen:

Warning: The forecasted likelihood for an extreme season is Y% and (1-Y)% for a normal season.

This forecast predicts the coming season will turn out to be extreme in Y out of 100 cases and in (1-Y) out of 100 cases this season will turn out to be normal. In case that the season is extreme, you will lose all of your points if you do not pay enough points for protection.

Information provided in […] was automatically updated for each season 1 to 10:

Please state the amount that you are prepared to pay for protection in season *[1 to 10]*:

You can choose any amount in whole numbers between 0 and 500 points:

[field to insert number - integer numbers between 0 and 500 possible, no digits after the dot]

With click on **"Submit"**, you confirm the amount that you are prepared to pay for protection and move on to the next step. You **cannot go back** and change your amount once you have clicked on "Submit".

Information provided in […] was automatically updated for each season 1 to 10:

Information about climate conditions in season *[1 to 10]*

Season [1 to 10] was **[extreme, normal]**.

Round 1-9: Please click "Continue" to move on to the next season.

Round 10: This was the last season. Please click "Continue" to move on to the second section of the study.

Po*st-experimental questionnaire*

Section 2 – Survey

To finish the study, please answer a few questions about yourself. Following this, you will be presented with your bonus payment results.

The answers and decisions that you make in this section of the study have **no influence** on the size of your bonus payment. However, you are only eligible to receive the reward for submission and your bonus if you finish this second section of the study as well. All of your answer will be kept confidential and will only be used in aggregate. None of the following questions can be used to identify you.

If a question is marked with *, an answer is mandatory.

---

1. How do you see yourself: Are you generally a person who is prepared to take risks, or do you try to avoid taking risks?

Please tick a box on the scale, where the value 0 means: “not at all willing to take risks” and the value 10 means: “very willing to take risks”.

You can use the values in between to grade your assessment.

not at all willing to take risks										very willing to take risks
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
○	○	○	○	○	○	○	○	○	○	○

Please click on “Continue” to submit your answer.

---

2. Did you trust the seasonal forecast you received for the last season (Season 10)?*

Please tick a box on the scale, where the value **1** means: “I strongly <u>mistrusted</u> the forecast.” and the value **6** means: “I strongly <u>trusted</u> the forecasts.”. You can use the values in between to grade your assessment.

- Likert scale ranging from 1 (“Strongly mistrusted.”) to 6 (“Strongly trusted.”)
- Additional option: “I don’t know”

3.a Have you personally experienced financial damage due to one or more of the mentioned extreme weather events in the last 24 months?

[choose all that apply]

- No, I did not.
- Blizzards
- Flash flooding/Flooding
- Heavy Storms/Thunderstorms
- Drought
- Wildfires/Forest fires
- Heatwave, i.e. more than three consecutive days with temperatures above 30°C
- Other. Please specify:

Follow up if answer to 3.a was not answered with "No, I did not":

3.b How high was the financial damage of your household caused by the experienced extreme weather events?

We are only interested in a rough estimate. Please indicate the approximate size of the financial damage. [choose one only]

- Less than £1,000
- £1,001 to £2,500
- £2,501 to £5,000
- £5,001 to £10,000
- £10,001 to £35,000

- More than £35,001
- I don't know.
- Prefer not to answer.

---

4. In your opinion, how likely is it that you experience major financial damage caused by an extreme weather event within the next 24 months?*

Major financial damage means damage that costs you **more than £5,000**.

- Likert scale ranging from 1 to 5: 1 - Very likely, likely, neutral, unlikely, 5 - very unlikely
- Additional option: I don't know

---

5. Please answer the following questions by choosing a number from 0 to 10, where 0 is "extremely bad" and 10 is "extremely good". You can use the values in between to grade your assessment.*

5.a How good or bad do you think the impact of climate change will be on <u>people across the world</u>?

5.b How good or bad do you think the impact of climate change will be on <u>your life</u>?

Answer scales for 5.a and 5.b from 0 = Extremely bad, to 10 = Extremely good

---

6. Please read the instructions carefully and answer the following question:

To help us understand how people make decisions, we are interested in information about you; specifically if you have taken the time to read the instructions. To indicate this, please ignore the question below and instead write "none" in the box. Thank you very much.

What is your favourite colour?* - *free text answer.*

7. Are you:*

- Female
- Male
- Trans Male/Trans Man
- Trans Female/Trans Woman
- Genderqueer/Gender Non Conforming
- Different Identity
- Prefer not to answer.

---

8. How old are you?*

*(*insert number*)* age in years

---

9. What is the highest level of education that you have completed?*

If you are currently enrolled in school, please indicate the highest degree you have received.

[choose one only]

- Primary school
- Secondary school
- College (e.g. Diploma, BTEC, Apprenticeship)
- Bachelor’s degree
- Post-graduate degree (e.g. *Postgraduate* Diploma, Master’s *Degree*, and/or PhD)
- Prefer not to answer

- Other. Please specify: _______________

---

10. Are you the owner of the place where you live?* [choose one only]

- Yes, I am the owner of the place where I live.
- No, I am a tenant.
- I prefer not to answer.
- Other. Please specify:

---

11. In which country are you currently living?*

free text answer

---

12. What type of area are you currently living in? [choose one only]

- Rural (less than 10,000 inhabitants)
- Urban (10,001 to 100,000 inhabitants)
- Metropolitan (more than 100,001 inhabitants)
- I don't know.
- I prefer not to answer.

---

13. How much does your combined household's disposable income per month add up to?*

Disposable monthly income is your household's monthly income after tax.

- Less than £1,499

- £1,500 to £2,999
- £3,000 to £4,499
- £4,500 to £5,999
- £6,000 to £7,499
- Over £7,500
- Prefer not to answer.

---

14. <u>Including yourself</u>, how many people currently live in your household?*

- Number of adults (18 years or older) [*insert number*]
- Number of children (younger than 18 years) [*insert number*]
- I prefer not to answer.

---

15. Which device did you use to participate in this study?*

[choose one only]

- Desktop computer
- Laptop
- Tablet
- Mobile Phone
- I prefer not to answer.
- Other, please specifiy:

16. Where were you while taking part in this study?*

Your answer to this question doesn't influence your reward payment or the acceptance of your submission.

- At home, without any distractions
- At home, with distractions (TV or music was playing, children in the same room, etc.)
- Not at home, without any distractions (quiet office, etc.)
- Not at home, with distractions (on the bus, train; in a noisy restaurant etc.)
- I prefer not to answer.
- Other. Please specify:

---

17. Do you have any further comments on the study?

You can also explain here why you made certain decisions during the study.

(voluntary question: insert open text)

Depending on the determined c and p of the randomly chosen season, the relevant information for x, the season (extreme or normal), c and subjects' bonus nn was automatically filled in:

Payment information

Thank you for participating in our study. Please click **"Return to Prolific"** at the end to complete your participation in this study. If you do not click on "Return to Prolific", we cannot process your payment and you will not receive your payment.

Your total payment consists of your reward for submission: fixed amount of **£2.00** plus your bonus payment of ***£X***.

Explanation to bonus payment:

For the determination of your bonus, season *x* was randomly chosen by the computer. Climate conditions in season *x* were *[extreme; normal]*.

In season *x*, you stated that you are prepared to pay *p* points for protection and the randomly determined price for protection is *c* points.

- *if option 1 "c>p" and extreme season occurred in season x - text:*

Since p points is less than the randomly determined price for protection c points, you have not met the criteria to be protected from losses due to the extreme season. Therefore, you do not pay for protection, and you lose your entire bonus.

Your bonus is 0 points, which corresponds to **£0.00**.

- *if option 2 "c>p" and normal season in season x - text:*

Since p points is less than the randomly determined price for protection c points, you have not met the criteria to pay for protection. However, since the season is normal, you do not experience any losses and thus, you keep your entire bonus.

Your bonus is 500 points, which corresponds to **£5.00**.

- *if option 3 "c≤p" independent of which season occurred (extreme or normal):*

    Since p points is equal or higher than the randomly determined price for protection c points, you are paying c points for protection independent of the climate conditions of season *x*.

    Hence, your remaining bonus is 500 points – *c* points = *nn* points, which corresponds to *£nn/100*.

Click the button "Return to Prolific" to complete your participation and return to Prolific.

Button to click: **Return to Prolific**

Completion URL was automatically saved and activated by clicking the return to Prolific-button.