Attribution to anthropogenic causes helps prevent adverse events

Florian Diekert, Timo Goeschl, Christian König-Kersting

Angaben zur Veröffentlichung / Publication details:

Diekert, Florian, Timo Goeschl, and Christian König-Kersting. 2025. "Attribution to anthropogenic causes helps prevent adverse events." *Journal of the Association of Environmental and Resource Economists (JAERE)* 12 (5): 1165–98. https://doi.org/10.1086/734547.

Attribution to anthropogenic causes helps prevent adverse events

Florian Diekert[a], Timo Goeschl[b,c], and Christian König-Kersting[d]

[a]Centre for Climate Resilience and Faculty of Business and Economics, University of Augsburg
[b]Alfred-Weber-Institute of Economics and Heidelberg Center for the Environment, Heidelberg University
[c]Leibniz Institute for European Economic Research (ZEW), Mannheim
[d]Department of Banking and Finance, University of Innsbruck

Post-print version November 15, 2024

Abstract

Can Attribution Science, a method for quantifying – ex post – humanity's contribution to adverse climatic events, induce pro-environmental behavioral change? We conduct a conceptual test of this question by studying, in an online experiment with 3,031 participants, whether backwards-looking attribution affects future decisions, even when seemingly uninformative to a consequentialist decision-maker. By design, adverse events can arise as a result of participants' pursuit of higher payoffs (anthropogenic cause) or as a result of chance (natural cause). Treatments vary whether adverse events are causally attributable and whether attribution can be acquired at cost. We find that ex-post attributability is behaviorally relevant: Attribution to an anthropogenic cause reduces future anthropogenic stress and leads to fewer adverse events compared to no attributability and compared to attribution to a natural cause. Average willingness-to-pay for ex-post attribution is positive. The conjecture that Attribution Science can be behaviorally impactful and socially valuable has empirical merit.

Keywords: Extreme event attribution; attribution science; behavioral change; cause dependence; online experiment.

JEL Codes: D91, Q54, C99

1 Introduction

Since the publication of a landmark paper on the *Human Contribution to the European Heatwave 2003* (Stott et al., 2004), the field of "Attribution Science" has been on the rise:[1] No fewer than 150 extreme event attribution studies have been published to date, documenting the significant and sometimes exclusive role of anthropogenic factors in explaining the occurrence and shape of recent extreme meteorological events, from the said 2003 European heatwave to the 2019 wildfires (Herring et al., 2021) and the 2023 Early Spring Heat in South America (Kew et al., 2023). The Sixth Assessment Report by the Intergovernmental Panel on Climate Change acknowledges the progress that the science of extreme event attribution (EEA) has made in the past decade (Arias et al., 2021). Attribution Science matters outside the narrow circle of climate scientists: In terms of media and public interest, EEA studies attract significant media attention (Ettinger et al., 2021). Legal scholars are discussing the potential ramifications of EEA for climate change litigation (Lusk, 2017; Marjanac et al., 2017; Stuart-Smith et al., 2021; Pfrommer et al., 2019).[2] And researchers examining man-made change in other environmental systems such as biodiversity are beginning to explore whether EEA could be meaningfully applied outside the climate change context(IPBES, 2019; James et al., 2019). In light of these developments, a better understanding of what EEA could contribute to mitigating anthropogenic environmental problems is called for.

EEA initially arose out of questions surrounding climate litigation, namely whether liability claims by victims of extreme events could be adjudicated on the basis of model-based *ex post* attribution to anthropogenic factors (Allen, 2003; Allen and Lord, 2004).[3] More recently, the focus has shifted to the possible role of EEA in preventing future extreme events by changing the mitigation behavior of the general public (Drake and Henderson, 2022; Oldenborgh et al., 2021; Ettinger et al., 2021; Osaka and Bellamy, 2020). In this view, informing people *ex post* about the contribution of anthropogenic factors to causing an extreme event just experienced could foster behavioral change towards reduced greenhouse gas emissions.

The idea of using EEA as a means to behavioral change is so far untested, but it could involve at least two distinct, yet complementary pathways. One is cognitive: The salience and recency of extreme events generates opportunities for the message

[1] For an introduction and overview of the methods, see for example Otto (2017).

[2] This is true in particular for those extreme events for which anthropogenic causation is considered a certainty. Such events have already occurred several times in the last ten years (Imada et al., 2018; Knutson et al., 2018; Walsh et al., 2018; Schiermeier, 2018).

[3] Jézéquel et al. (2020) provide a short social history of EEA.

of climate science to overcome a set of well-known inherent biases in how humans engage with climate change.[4] EEA arguably has potential to correct some of these. For example, explicit attribution of events could attenuate attention bias, increase visibility and help reduce psychological distance (Loy and Spence, 2020; McDonald et al., 2015). A familiar and trusted methodology could combat perception bias (Osaka and Bellamy, 2020), possibly enhanced by narrative storylines (Shepherd et al., 2018). EEA could thus support the cognitive linking of events to human activities and – ultimately – to active mitigation behavior (Wong-Parodi and Rubin, 2022). Employing EEA for 'cognitive bias correction' is currently being explored (Ettinger et al., 2023). If successful, this could lead to EEA playing a greater role in communicating climate change and its consequences to the general public.

A second possible pathway is affective: Coupling the experience of an extreme event with causal attribution could trigger behavioral responses driven by feelings such as regret (Loomes and Sugden, 1982), guilt (Battigalli and Dufwenberg, 2007), or experiential effects (Malmendier, 2021). Attribution could activate norms of responsibility, accountability, and ethical conduct that are present in many social settings where actions can have adverse consequences (Swim et al., 2009), including climate change (Hulme, 2014). Findings that the presence of anthropogenic risk leads to changes in risk perception (Slovic et al., 1986; Brun, 1992; Siegrist and Sütterlin, 2014; Hoogendoorn et al., 2020) and in perceived responsibility (Liu and Du, 2022; Weiner, 1995; Reser and Swim, 2011) could be seen as supportive of this idea.[5] Relatedly, environmental economists have found evidence that causal awareness of anthropogenic cause affects environmental valuation (Bulte et al., 2005; Shreedhar and Mourato, 2020; Hindsley and Ashton Morgan, 2020).

At the present time, whether EEA has behavioral impact, and why, can be inferred only from related studies and with only modest confidence.[6] Existing research

[4]See Steg (2023) and Zhao and Luo (2021) for recent reviews. See also the comprehensive 2009 report of the American Psychological Association (Swim et al., 2009) and a survey paper by Gifford (2011). Examples are attention bias to different information (Carlson et al., 2022), perception bias with respect to different information sources (Sarabia-Sanchez and Rodriguez-Sanchez, 2016), confirmation bias and attribution bias with respect to prior beliefs (Jang, 2013), and present bias, grounded in mechanisms such as motivated cognition, incomplete system understanding, imperfect learning, and negative affect.

[5]However, studies also find that individuals are able to ignore or deflect the pressures of such norms in a climate context (Gifford, 2011; Kriss et al., 2011; Jang, 2013), limiting or even negating the potential for behavioral change (Murtagh et al., 2012; Swim et al., 2009).

[6]This contrasts with a literature that has attempted to document the impact of experiencing extreme events on updating beliefs regarding climate change (e.g. Deryugina, 2013; Larcom et al., 2019; Lohmann and Kontoleon, 2023) and on mitigation choices (e.g. Demski et al., 2017; Ogunbode et al., 2019; Hazlett and Mildenberger, 2020). At most, EEA has been shown to be somewhat conducive to post-event adaptation (Lusk, 2017; Singh et al., 2019). This is also true for the communication and perception stage, where the evidence is mixed (Howe et al., 2019; Osaka

has been focusing on more readily accessible outcome variables such as changes in perceptions (Osaka and Bellamy, 2020; Spence et al., 2011; Weber, 2006) or behavioral intentions (Wong-Parodi and Rubin, 2022; Yang et al., 2015), elicited through surveys and hypothetical scenarios. Observing people's actions in consequential study environments is more involved and more costly, but likely has merits for two reasons. One, it brings the behavioral study of EEA closer to the outcome variable of interest. Two, consequential environments constitute a more stringent test for EEA because EEA is – by definition – backward-looking rather than decision-supporting, descriptive rather than predictive. EEA feedback on an event can therefore not affect the expected benefits that a well-informed decision-maker associates with different future courses of action. Put differently, a well-informed decision-maker could be seen as committing a costly error when changing behavior in response to EEA.[7] Once people invest the cognitive effort to avoid costly errors, then the impact of EEA could vanish in consequential environments. If it does not vanish, the impact on behavior would be strong evidence in favor of EEA, pointing beyond cognitive bias correction towards affective pathways.

The present paper employs an experimental approach in order to examine what we regard as the core question about the possible behavioral impacts of EEA, namely whether there is a distinct behavioral role for EEA above and beyond informing about climate change and its consequences. Put differently, the question is whether in a world of accurate, unambiguous, and trusted attributality, ex-post attribution of a materialized event to natural or anthropogenic causes has the ability to impact on future choices. To have the desired effect, the impact should be directed: When the event is attributed to an anthropogenic, rather than a natural, cause, the decision-maker will choose an action that *reduces* anthropogenic risk. We carefully examine this question in an abstract setting following three steps. First, we determine whether there is an *attributability effect*: Is behavior different in a world in which causal attribution is conducted compared to one in which it is not? Second, we investigate the *causal attribution effect*: If a past adverse event is attributed, do future actions depend on which of the causes, natural or anthropogenic, it is attributed to? In particular, does attribution to an anthropogenic cause induce people to reduce anthropogenic risk? Third, we establish the *demand for causal attribution*: Are people who experienced an adverse event willing to give up money

and Bellamy, 2020; Boudet et al., 2020; Ettinger et al., 2021).

[7]Unless the decision-maker changes between actions with the same expected benefits. In their authoritative survey on the psychology of climate change Swim et al. (2009) also note that when looking at past events, "[t]he distinction between natural and human-made causes may appear irrational in the face of consequential considerations".

to receive feedback that makes attribution possible?

To complete the three steps, we propose an experimental design that captures – in an abstract way – key aspects of human choice in bicausal systems.[8] Our participants play two rounds of a modified, two-player Chicken Game. Adverse events are zero-payoff outcomes that can arise as a result of an unfavorable draw from an urn (natural cause) or as a result of players' interactions (anthropogenic cause). Importantly, players are re-matched with a randomly selected stranger between rounds. The parameters of the Game are common knowledge and – importantly – identical in both rounds and across causes. Treatments vary the feedback environment, from full causal attributability after both rounds to attributability on demand after round 1 or 2. There, players can bid money for receiving the causal feedback. These three treatments generate the required variation to identify the three effects.

Our online experiment with 3,031 participants leads to four results. First, there is evidence for an attributability effect: The share of subjects switching action is significantly higher when the adverse event is attributable to an anthropogenic cause compared to when it is not (45% versus 32%, p=0.032). For a natural cause, the share of subjects switching action is lower when the adverse event is attributable than when it is not (24% versus 33%, p=0.075). Second, there is very strong evidence for a causal attribution effect: Attributing an adverse event to an anthropogenic cause makes a significantly higher share of subjects switch away from the 'high stress' action compared to a natural cause (45% versus 24%, p=0.004). Both effects are robust to controlling for a range of covariates such as age, gender, risk aversion, and education. Third, demand for causal attribution is highly heterogeneous, yet clearly present: The overwhelming majority (68%, p<0.001) express an interest in causal feedback and around half submit positive bids. Fourth, while demand for causal feedback depends on the experienced history, it does differ whether it has been elicited before or after the second round.

The results are significant for three reasons. One is methodological: The re-

[8]Widely present in socio-ecological (Erisman et al., 2015; Leichenko and O'Brien, 2008; Vélez et al., 2002; Hallegraeff, 2003) and technological contexts (Helbing, 2013; Atalay et al., 2011), bicausal systems are characterized by the property that both natural variability and human coordination failure can cause adverse system events, including system collapse. In the climate context, for example, the occurrence of a heat wave can be the result of anthropogenic (such as the burning of fossil fuel), but also of natural causes (El Nino-Southern Oscillation). In the fisheries context, sudden population collapses can be the result of excess harvesting (anthropogenic cause) or upwelling of anoxic waters through episodic changes in ocean currents (natural cause) (Grantham et al., 2004). Internet outages are also an example, being attributable to equipment breakdown (natural cause), but also to network congestion (anthropogenic cause) Feldmann et al. (2020); Labovitz et al. (1999). In all of these systems, the possibility that ex-post knowledge about the causes might change how humans interact with the system in the future raises important questions about whether to support causal attribution.

sults showcase the potential of bicausal experimental designs to help identify causal effects of attribution on behavioral outcomes. Such designs offer controlled and decontextualized settings that still retain a core property of many socio-ecological and socio-technological systems, namely the simultaneous presence of man-made and natural risk. Establishing the attributability and attribution effects within this design suggests that the bicausal paradigm could be a suitable workhorse for conducting behavioral research across a range of structurally similar environments. The second significance is topical: The results provide what we believe to be the first proof-of-principle of whether the practice of EEA can be behaviorally impactful, and one derived under rather demanding conditions: Several thousand "workers" on an online platform take decisions with monetary consequences in a bicausal system of low complexity, in which causal feedback is uninformative. This setting largely neuters possible cognitive benefits of EEA for system understanding, favors consequentialist rationales for action, and removes possible contextual framings associated with climate change. Yet, we find very strong evidence that causal attribution after an adverse event matters for future behavior. This evidence points to the affective channel and to possible contributions of EEA to behavioral change that lie beyond its scope for correcting cognitive biases. The idea that Attribution Science could be impactful for behavioral change therefore seems not without merit. Properly developed, evidence for its social value may well underpin future arguments for a continued provision of EEA in the climate context and beyond where behavioral responses could contribute to solving human-induced damages. The third significance is programmatic: Given the rapid rise of Attribution Science, the media interest in its outputs, the scale of the mitigation challenge, but also given the early stage of research on its possible societal impacts, our results but provide a first step in what could be a larger research effort towards establishing the impact and optimal design of EEA for behavioral impact. We conclude the paper with the sketch of a research agenda that not only examines underlying mechanisms, but charts a path to its external validity in the climate context by probing robustness to procedural, substantive, and structural changes to the bicausal design.

2 Experimental Design

The experimental design is intended to capture key features of bicausal systems: Two causes, natural and anthropogenic, can both lead to an adverse system event at any given time and, therefore, within the same treatment. In this, we depart from previous experiments that examine behavioral change in monocausal systems

(Blount, 1995; Buckenmaier and Dimant, 2021; Schwerter and Zimmermann, 2020) where only one cause can lead to an adverse system event at a given time, and the treatment manipulation consists of varying whether the cause is human or nature-generated. In our view, only a genuinely bicausal design can credibly capture situations in which players co-generate anthropogenic risk in the presence of natural risk, leading to cause uncertainty in the absence of attributability.

The three treatments share the same basic structure and protocol, but differ in aspects of feedback and timing. The common element is a stylized system in which natural and anthropogenic causes can both lead to adverse events. The system is presented schematically in Figure 1. Participants always play two rounds of what we call a Modified Chicken Game (MCG). One of the two rounds is later randomly selected for payout. In each round, players are randomly (re-)matched into pairs of two. Each pair plays a standard chicken game. The normal-form game is presented in Figure 2. If both players 1 and 2 choose action L, they both receive a payoff of x. If both players choose action H, they both receive a payoff of zero. When actions do not coincide, the player playing action L receives x, while the player playing action H receives $x + y$. Action L can be thought of as a low-stress action, while H represents a high-stress action.[9] Before they choose their action, players are informed about the probability that a co-player chooses action L or action H, based on the empirical distribution of play in the population of players.

At the same time, a ball is drawn from an urn. Players are informed about the known distribution of red and green balls. If a red ball is drawn, both players receive a payoff of zero, independent of the outcome of the chicken game. If a green ball is drawn, both players receive the payoff determined by the outcome of the chicken game. The MCG therefore contains two sources of risks, the randomly matched co-player's action and the urn draw, with easily comprehensible statistical properties.[10]

After each round, players receive feedback. The minimum feedback that is always given is event feedback: Players are informed about the payoff they receive if the round is chosen for final payment. The experimental treatments vary whether, additionally, causal feedback is given and players are made aware of result of the urn draw and of the co-player's action. The treatments vary whether causal feedback is completely withheld, is always given, or can be acquired at a cost.

[9] In the paper we refer to the low-stress action as L and the high-stress action as H for a more intuitive presentation. In the experiment, the actions were abstractly referred to as A and B, respectively.

[10] We verify comprehension of the experimental task by subjects in a quiz that precedes the experiment. See section 2.4 for more details.

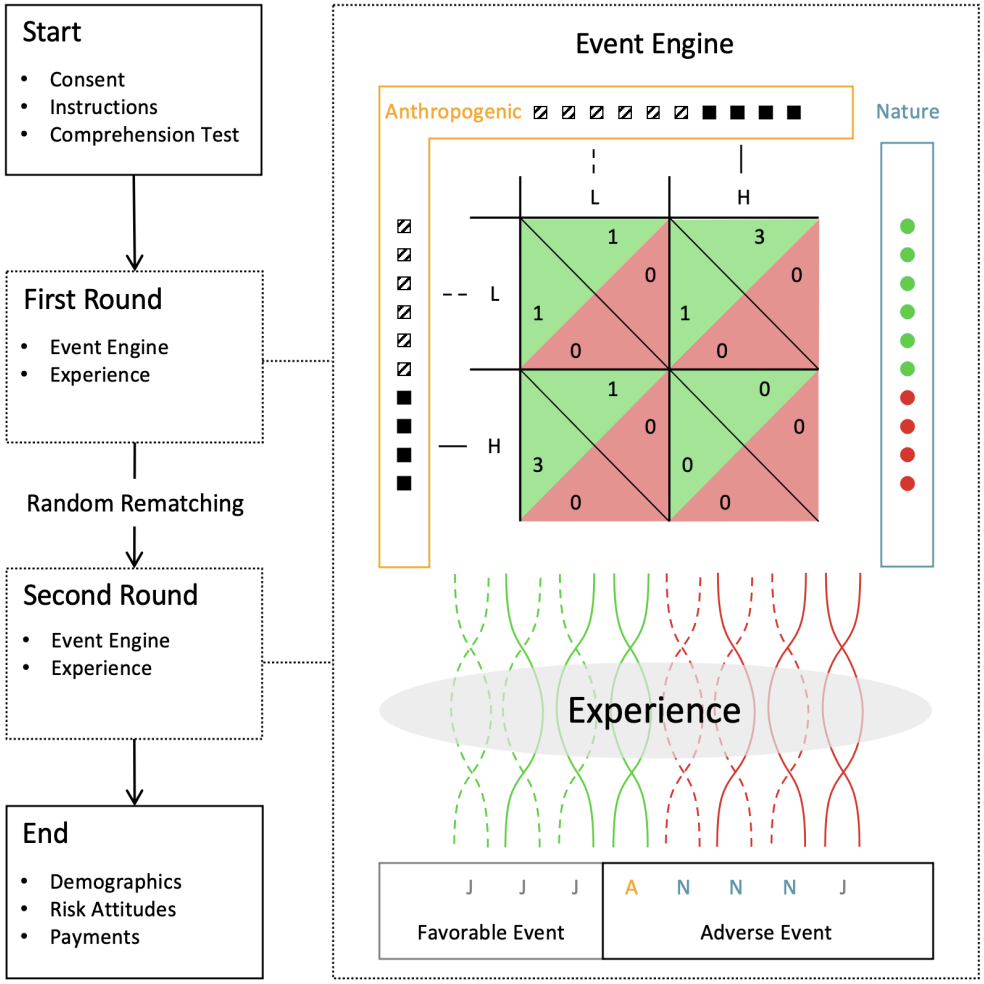

Figure 1: The left panel charts, from top to bottom, the progression of the experiment, from the consent form via the first and second round to experimental payments at the end. The right panel shows, for each round, the path from event determination to experience. From top to bottom, it shows how events are determined by the event engine by combining anthropogenic factors (action L – dashed, empirically 60% of actions; action H – solid, empirically 40% of actions) of two participants with natural factors, the draw from an urn of green (60% probability) and red (40%) balls. Eight combinations of anthropogenic and natural factors are possible, three leading to events with positive payouts (favorable) and five to events with zero payoffs (adverse). Among adverse events, one combination is uniquely attributable to anthropogenic cause (A), three are uniquely attributable to nature (N) and one is jointly attributable (J).

Three considerations are important with respect to the design and the treatments. One is that feedback is always truthful: Subjects are never deceived about what history of play led to the specific adverse event they experienced, if any. Second, the history of play is inconsequential in a number of ways: Round 2 does not depend on what happened in round 1; round 1 does not provide clues about round 2; and whether chance draws a red ball from the urn or a co-player that chose action H from the pool of players could be thought to make little difference to players who chose action H themselves, ostensibly in pursuit of a higher payoff. Third, subjects have agency: Those of interest in this study are those who deliberately exposed themselves to both natural and anthropogenic risk, and they have the ability to respond to experience by changing their action in round 2.

		Player 2	
		L	H
Player 1	L	x , x	x , $x+y$
	H	$x+y$, x	0 , 0

Figure 2: Normal form of the chicken game

2.1 Treatments

We conducted three treatments. These treatments varied whether causal feedback is available after a certain round and if so, whether it needs to be acquired at cost. This variation in the timing and availability of causal feedback provides the between-subjects variation required to establish the effects of attributability (comparison between causal feedback being available or unavailable) and attribution (comparison between different causes, given feedback) as well as the WTP for causal feedback.

The 'BASELINE' (BL) treatment features free and full feedback after both rounds. Participants complete round 1 of the experiment by choosing an action (L or H) and then responding to a non-incentivized belief elicitation about two items, their belief about their co-player's action (L or H) and about the draw from the urn (*red* or *green*). After completion, participants individually and automatically receive both event and causal feedback. The event feedback consists of the payoff outcome of the round (0, x, or $x+y$). The causal feedback consists of the co-player's action (L or H) and of the urn draw (*red* or *green*). Participants then move on to round 2, in which they are randomly re-matched with a perfect stranger. They again choose one of the two actions and respond to the two-item belief elicitation. After round 2, participant again receive event feedback and causal feedback. Participants

of treatment BL can therefore always attribute past outcomes to specific causes, whether a favorable or unfavorable realization of the natural cause (urn draw) or the realization of the human cause (strategic choice of co-player), or both.

The second treatment contains most of the elements of treatment BL, but restricts causal feedback in a particular way. After completing round 1, participants only receive event feedback (0, x, or $x + y$), but no causal feedback. The same is true after the final round 2. After round 2, they have the opportunity to acquire feedback at a cost. Their WTP for causal feedback is elicited through a two-step filter. In the first step, there is a yes-no question about their interest in causal feedback. If an interest is expressed, a BDM-mechanism elicits their WTP, bounded between \$0 and \$0.50. If the BDM-mechanism accepts their bid, participants pay and receive causal feedback for round 2. We refer to this treatment variation, in which causal feedback is withheld after round 1 and can only be acquired at cost after the final round 2, as the 'FINAL' (FI) treatment.

The third treatment is called the 'INTERIM' (IM) treatment because causal feedback can now be purchased in the interim between round 1 and round 2. It differs from treatment FI only in timing: The two-step WTP filter is now moved from the end of round 2 (treatment FI) to the end of round 1. After round 2, participants automatically receive both event and causal feedback. The alternative timing means that compared to treatment FI, those who receive causal feedback on round 1 following the BDM proceed to round 2 with feedback that is as rich as that in treatment BL, but a slightly lower income. Conversely, those who do not receive causal feedback proceed to round 2 with feedback that is as poor as that in treatment FI, but at a slightly higher income. Informationally, whether feedback is rich or poor makes no difference: Due to partner re-matching and independent urn draws, all probabilities and therefore all expected payoffs remain unchanged.

2.2 Histories

There are eight different histories in round 1 of each treatment. Let the choice of player i in round t be denoted by $C_{i,t} = \{\text{L}; \text{H}\}$ (parallel for player j). The realization of the urn draw (lottery) is denoted by $L_t = \{\text{red}; \text{green}\}$. The payoff of player i for round t is denoted by $\pi_{i,t}$. The probability that a red ball is drawn in the lottery is p (that is, $p = \Pr(L = \text{red})$) and the probability that a player chooses action H is q. We adopt the following convention to label the histories: "LLg" refers to the history where the player chose action L, the co-player chose action L, and the ball was green. "LLr" refers to the history where the player chose action L, the

co-player chose action L, and the ball was red, etc. Table 1 shows the histories, corresponding choices, urn draw outcomes, as well as payoffs and probabilities.

Table 1: The eight different histories of round 1

$\mathcal{H}$	$C_{i,1}$	$C_{j,1}$	L	$\pi_{i,1}$	$\Pr(\mathcal{H}{=}\mathrm{h}n)$
LLg	L	L	green	x	$(1-p)(1-q)^2$
LLr	L	L	red	0	$p(1-q)^2$
LHg	L	H	green	x	$(1-p)(1-q)q$
LHr	L	H	red	0	$p(1-q)q$
HLg	H	L	green	$x+y$	$(1-p)q(1-q)$
HLr	H	L	red	0	$pq(1-q)$
HHg	H	H	green	0	$(1-p)q^2$
HHr	H	H	red	0	pq^2

Table 1 identifies five histories that lead to a zero payoff event. In two of these, LLr and LHr, the event feedback is sufficient for a decision-maker to deduce that the ball drawn from the urn was red and that nature therefore led to the zero event. Causal feedback is therefore redundant for any player who chose action L. For any player who chose action H, the occurrence of a zero event does not settle the question of causality. In history HHr, natural and human risk jointly materialize to cause the adverse event. In history HLr it is uniquely nature, in history HHg uniquely the human co-player whose action caused the player to suffer a zero payoff.

From a strictly consequentialist point of view, it is sufficient that the instructions provide truthful statements about the stochastic properties of the two causes of zero payoffs. Looking ahead, however, behavioral differences between subjects with experiental biography HLr and biography HHg could possibly result from differences in the *perceived likelihood* of a zero payoff event stemming from the two sources of uncertainty. In order to attenuate this possible source of behavioral differences, we exploit that for a given q, the experimenter can calibrate p to obtain different distributions of our sample over histories in terms of ex-ante likelihood. The experimenter can therefore parametrize the urn such that it is equally likely for a player choosing H that the zero payoff event is uniquely caused by the co-player choosing H or by the ball being red. To generate a setting in which $p \approx q$, three pilots were conducted. The parametrization was based on the pilot ($n = 81$), which returned for $p = 0.4$ a share $q = 0.36$ of subjects choosing action H. The instructions accordingly inform participants that "The urn from which the ball is drawn contains 40 red balls and 60 green balls" and "Typically, co-players choose action

L in about 60 out of 100 cases and action H in about 40 out of 100 cases.".[11]

2.3 Beliefs and controls

In addition to the payoff-relevant choices, we elicit participants' perceptions about the cause of the outcome before providing feedback. In each of the two rounds, we ask participants whether they believe the color of the ball drawn to be green or red, and whether they believe their co-player's choice to be either L or H. To allow participants to express how confident they are in these beliefs, they use an unlabeled slider inputs ranging from red to green (L to H) to give their answer. The sliders do not show a starting position. We treat the responses as believed probabilities of the ball being green (the co-player choosing H). Furthermore, we collect demographic information on age, gender, and the level of education. To control for risk attitudes, we also ask the SOEP question on general willingness to take risks.

2.4 Procedures

The experiment was conducted online on Amazon Mechanical Turk (AMT). We used o-Tree (Chen et al., 2016) to program and conduct the experiment. Figure 3 illustrates the flow of the experiment. After reading the study description on Amazon Mechanical Turk and giving their consent to participate in the study, participants first had to pass a qualification check consisting of the general instructions on payoff calculation, duration, and the different stages of the experiment and the design-specific instructions. The qualification check serves to filter out potential non-human (bot) participants, inattentive or confused participants. Participants then played two rounds of the MCG as described above. In between, we elicited their beliefs about the urn drawn and the co-player's decision. We also elicited their interest in and willingness to pay for causal feedback where appropriate. Finally, participants filled in a short questionnaire before learning about their final payoff.

Across conditions FI, IM, and BL, we recruited a total of 3,031 participants. A table with detailed information on the demographic characteristics of the sample, split by treatment, is shown in Appendix, Table A-1. Data was collected between April 16, 2021 and May 26, 2021. One of the two rounds was randomly selected for payout. Participants received \$0.80 for participating, could earn up to an additional \$3.00 in the game (that is, we set the payoffs to x=1 USD and y=2 USD). In condition FI and IM, participants received a bonus of \$0.50 after completing the

[11]The approximation also succeeds ex post: 1009 out of 3031 subjects, i.e. over 33%, choose action H.

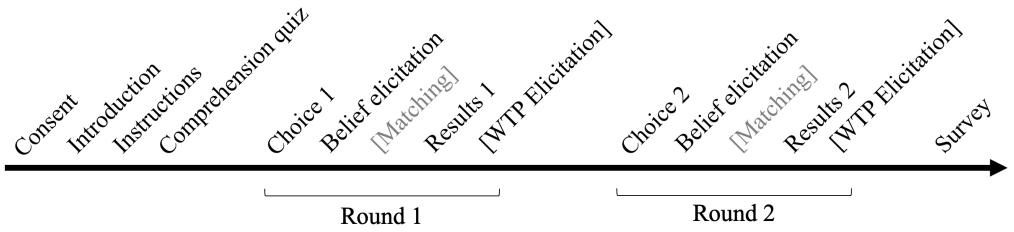

Figure 3: Stages of the experiment

comprehension test (this compensates for the causal feedback acquisition). One of the rounds was randomly selected to be payout-relevant. Each participant was then paid accordingly through AMT. Final payments are comparatively high for short online studies that only take 10 to 15 minutes to complete. No deception was used.

3 Hypotheses

Our primary outcome variable of interest is behavioral change among participants. This indicator of change in choice is denoted by Y_i=0 if $C_{i,1}$=$C_{i,2}$, i.e. the player chose the same action in both rounds, and Y_i=1 if the action chosen in the second round is the opposite of the action implemented in the first round ($C_{i,1}\neq C_{i,2}$). Our secondary outcome variable is individual willingness to pay. For individual i expressing a willingness to pay for causal feedback on round t, this is a variable $b_{i,t}$.

The set of histories is denoted by $\mathcal{H}$ and the outcome of the urn-draw (lottery) in round t is given by L_t. Finally, the variable F denotes whether a participant received causal feedback, where $F = 1$ means that a participant received causal feedback (either because they were in the BL treatment or because they were in the IM treatment and successfully purchased feedback) and $F = 0$ when a participant did not receive causal feedback.

3.1 The attributability effect

Recall that the strategic choice situation of the participants has the form of a chicken game. If all participants are fully rational players and believe that all other participants are also fully rational, standard game theory gives a clear prediction about the outcome of the experiment. The chicken game has a Nash equilibrium in mixed strategies.[12] Participants are informed about average play in this game form

[12]If players were risk neutral, we should observe – in the stage game – a mixed strategy of $p(L) = \frac{1}{3}$ and $p(H) = \frac{2}{3}$. Clearly, from the data on average play, subjects are risk-averse, since the action L is taken with much higher frequency than in the mixed-Nash.

and are given the same parametrization. Participants are also informed about the probability of an unfavorable (red) draw from the urn. Moreover, this probability is independent of the strategic choice and co-player re-assignment between rounds is randomized. Both the information about average play and about the lottery remain the same from round 1 to round 2.

The attributability effect would be present if we found that for the same history of play, the frequency of behavioral change is significantly different when the cause of the adverse event is disclosed compared to when it is not. In the context of the three treatments, this question can be rephrased as asking whether, compared to subjects that only receive event feedback after an adverse event (treatment FI and treatment IM with no feedback), do subjects that receive causal feedback (treatment BL and treatment IM with feedback) behave differently in the next round.

Consistent with our stance, we predict observing no change in *average* behavior. A narrowly consequentialist decision-maker that experienced a bicausal adverse event[13] would be expecting a uniquely anthropogenic and natural cause with equal probability – and would be expecting the same probabilities in the following round. The realization of the specific cause in round 1 is immaterial to such a decision-maker's considerations. This agnostic prediction is captured in Hypothesis 1.

Hypothesis 1 *Given the same history, causal feedback after an adverse event does not affect average behavior.*

Hypothesis 1 is threatened by non-consequentialist behavior among participants. For example, feedback that the adverse event was due to anthropogenic action could trigger feelings of regret (Loomes and Sugden, 1982), guilt (Battigalli and Dufwenberg, 2007), or responsibility (Swim et al., 2009). On the other hand, without attributability, affective responses to the event feedback are unconstrained by causal feedback. In such a situation, causal beliefs – and affective responses to causal beliefs – are the likely, but also unpredictable drivers of behavior, as the literature on attribution bias has demonstrated (Hestermann and Le Yaouanq, 2021).

To test hypothesis 1, we investigate the change in choice Y_i and examine whether causal feedback affects behavior. A total of five histories generate an adverse event with $\pi_{i,1} = 0$. In two of these histories, {LLr,LHr}, the event feedback invariably includes causal feedback: Only the urn draw could have been the cause of the adverse event if player i chose $C_{i,1}{=}L$. As a result, treatments with causal feedback do not substantively differ in these two histories. In the remaining three histories, namely

[13]Recall that the focus of the analysis must be on those subjects that chose the high-stress action in round 1.

{HHg,HLr,HHr}, the event feedback does not coincide with the causal feedback: Both the co-player action or the urn draw could have been the cause of the adverse event if player i chose $C_{i,1}$=H. We therefore test Hypothesis 1 by comparing behavior across the three histories in which the treatment induces differences in causal feedback. This choice of histories means that we consider only those participants with action H in the first round (see Table 1). As a consequence, a value of Y_i=1 uniquely means $C_{i,2}$=L.

$$\mathrm{E}[Y|\mathcal{H}{=}\mathrm{HHg} \wedge F{=}1] = \mathrm{E}[Y|\mathcal{H}{=}\mathrm{HHg} \wedge F{=}0] \tag{1a}$$

$$\mathrm{E}[Y|\mathcal{H}{=}\mathrm{HLr} \wedge F{=}1] = \mathrm{E}[Y|\mathcal{H}{=}\mathrm{HLr} \wedge F{=}0] \tag{1b}$$

$$\mathrm{E}[Y|\mathcal{H}{=}\mathrm{HHr} \wedge T{=}\mathrm{BL}] = \mathrm{E}[Y|\mathcal{H}{=}\mathrm{HHr} \wedge F{=}1] \tag{1c}$$

These tests set up three group comparisons that examine whether there are behavioral differences depending on the specific feedback received. Note that given the same history, the test does not require the average actions in round 2 to match those in round 1. This reflects the fact that there could be reasons for behavioral change between round 1 and 2 that are orthogonal to causality, such as diversification, experimentation, or boredom. It only requires that the average rate of behavioral change does not differ across the statistically equivalent causes.

One could argue that for participants that do not receive causal feedback it is irrelevant whether they experience history HHg,HLr,HHr, as they have the same information. Hence, we test, in addition to (1a)-(1c), the corresponding comparisons (2a)-(2c)

$$\mathrm{E}[Y|\mathcal{H}{=}\mathrm{HHg} \wedge F{=}1] = \mathrm{E}[Y|\mathcal{H}{\in}\{\mathrm{HLr},\ \mathrm{HHg},\ \mathrm{HHr}\} \wedge F{=}0] \tag{2a}$$

$$\mathrm{E}[Y|\mathcal{H}{=}\mathrm{HLr} \wedge F{=}1] = \mathrm{E}[Y|\mathcal{H}{\in}\{\mathrm{HLr},\ \mathrm{HHg},\ \mathrm{HHr}\} \wedge F{=}0] \tag{2b}$$

$$\mathrm{E}[Y|\mathcal{H}{=}\mathrm{HHr} \wedge T{=}\mathrm{BL}] = \mathrm{E}[Y|\mathcal{H}{\in}\{\mathrm{HLr},\ \mathrm{HHg},\ \mathrm{HHr}\} \wedge F{=}0] \tag{2c}$$

3.2 The causal attribution effect

Recall the strategic choices available to the participants. Given these choices, the causal attribution effect would be present if we found that in those treatments in which subjects receive causal feedback, the frequency of behavioral change is

significantly different when the cause of the adverse event is natural as opposed to anthropogenic. In the context of the three treatments, this question can be rephrased as asking whether in treatments BL, the average change in choice after experiencing history HHg (anthropogenic cause) differs from the average change in choice after experiencing history HLr (natural cause). In line with a narrowly consequentialist stance, the prediction for the causal attribution effect is that there is no such difference.

Hypothesis 2 *The average change in behavior after experiencing an adverse event is the same for subjects irrespective of whether the event occurred on account of an anthropogenic or a natural cause.*

Like Hypothesis 1, Hypothesis 2 is also threatened by non-consequentialist behavior. For example, feedback that the adverse event was caused by anthropogenic causes might be perceived as indicating that the player has situational control, leading to behavioral change (Van der Linden, 2015).[14] At the same time, it is important to recall that a player who finds herself in our sample must have chosen the 'high stress' action to begin with. Affective reactions towards that action, such as guilt or unease about a perceived norm violation (after all, the safe and equitable 'low stress' was available), would therefore not have to depend on either the adverse event materializing or on it being due to natural or anthropogenic causes.

To test whether the data align with Hypothesis 2, we conduct the following test on the experimental outcomes in treatments BL and IM:

$$\mathrm{E}[Y|\mathcal{H}{=}\mathrm{HHg} \wedge F{=}1] = \mathrm{E}[Y|\mathcal{H}{=}\mathrm{HLr} \wedge F{=}1] \tag{3}$$

Note, again, that Hypothesis 2 does not postulate a general absence of behavioral change following an adverse event. Instead, it simply proposes that these changes in behavior are statistically indistinguishable across the two causes of adverse events.

3.3 The willingness to pay for causal feedback

The third research question focuses on the willingness to pay for causal feedback. In the context of the three treatments, this question can be rephrased as asking whether subjects that only receive event feedback after an adverse event place positive bids for being shown the urn draw and the co-player's action.

[14] We are grateful to an anonymous referee for suggesting this link to the locus-of-control literature Rotter (1966).

The opportunity to place bids for causal feedback arises after round 2 in treatment FI and after round 1 in treatment IM, irrespective of histories. Causal feedback is strictly backward looking and contains no information for future rounds, given that round 1 and round 2 are statistically indistinguishable. The value of information of causal feedback is thus zero both in treatment FI and treatment IM.

In line with a narrowly consequentialist stance, Hypothesis 3 postulates that strictly backward-looking causal feedback will not attract significant demand.

Hypothesis 3 *Backward-looking causal feedback after an adverse event attracts an average bid of zero.*

Hypothesis 3 could fail due to a possible curiosity effect (Loewenstein, 1994): Kim et al. (2013) conduct a small-scale laboratory experiment that provides tentative evidence that decision-irrelevant information can attract willingness to pay. If subjects attach a utility to satisfying their curiosity about the chain of events that led to the adverse event, then this could be a reason for positive bids.

To test Hypothesis 3, we observe the expressions of interest in causal feedback $I_i \in \{0; 1\}$ and, conditional on interest, the bid b_i. As before, we test Hypothesis 3 by comparing behavior in and across treatments FI and BL in the three histories {HHg,HLr,HHr} in which the treatment induces differences in causal feedback. For Hypothesis 3 to hold, we expect:

$$\mathrm{E}[b|\mathcal{H}\in\{\text{HHg, HLr, HHr}\} \wedge T{=}\text{FI}] \;=\; 0 \tag{4a}$$

$$\mathrm{E}[b|\mathcal{H}\in\{\text{HHg, HLr, HHr}\} \wedge T{=}\text{IM}] \;=\; 0 \tag{4b}$$

Should Hypothesis 3 fail, pointing to a possible curiosity effect, then the question arises about when the feedback opportunity arises. In our experiment, causal feedback can be acquired after an adverse event in round 1 (in treatment IM) or in round 2 (in treatment FI). This leads to Hypothesis 4.

Hypothesis 4 *Backward-looking causal feedback after an adverse event attracts the same average bid after round 1 and after round 2.*

Hypothesis 4 could be threatened if demand for causal feedback is driven by different motives at different points in the experiment. The curiosity effects that could lead to Hypothesis 3 failing might be present throughout, but would not threaten Hypothesis 4 unless it varied across rounds. One reason for a variation might be a "need for closure" (Webster and Kruglanski, 1994), which would favor acquiring

feedback after round 2 rather than after round 1. Alternatively, participants could erroneously believe that the feedback adds information to what they already know about the system and thereby enable a better informed decision in round 2. This would favor acquiring feedback after round 1 rather than after round 2, after which no decision follows.

We test Hypothesis 4 by comparing mean bids in treatment FI and treatment IM.

$$\mathrm{E}[b|\mathcal{H}\in\{\mathrm{HHg},\ \mathrm{HLr},\ \mathrm{HHr}\} \wedge T{=}\mathrm{FI}] = \mathrm{E}[b|\mathcal{H}\in\{\mathrm{HHg},\ \mathrm{HLr},\ \mathrm{HHr}\} \wedge T{=}\mathrm{IM}] \quad (4)$$

If confirmed, this means that while curiosity might be a driver of WTP, the interest in causal attribution is entirely backward looking. If we find, on the other hand, that the average bid after round 1 is significantly higher than the bid after round 2, then this could be evidence for a forward looking interest in causal attribution that goes beyond curiosity.

4 Results

As the hypotheses makes clear, the subsample of interest for the analysis are participants who chose action H (the 'high-stress' action) and experience a zero payoff event in round 1. The outcome variable of interest is therefore whether a participant decided to reduce stress on the system, from H in round 1 to L in round 2. These participants have experienced the zero payoff event in round 1 being either uniquely caused by the strategy combination in the group (anthropogenic cause), or by it being uniquely caused by a red ball draw (natural cause), or because both occurred simultaneously (joint cause). However, only participants in condition BL and successful bidders in condition IM receive this feedback. Participants in condition FI serve as the baseline for the analysis of attributability. Participants' actions and the event engine resulted in 614 relevant experimental biographies.

4.1 The attributability effect

We first focus on participants in conditions FI (1,031 participants, of which 352 chose high stress) and IM (1,002 participants, of which 322 chose high stress) to whom causal feedback was not made available after round 1. We find that in the absence of attributability, 32% of these participants reduced anthropogenic stress from high (H) to low (L) after experiencing an adverse event (see Figure 4, hatched

bar, left). These participants could not know whether the urn draw (natural cause) or the coincidence of high stress actions (anthropogenic cause) led to the outcome. A substantial amount of behavioral changes therefore happens without the ability to attribute causality. This amount serves as a baseline of behavioral change for subsequent comparisons.

Both in condition BL (998 participants, of which 335 chose high stress) and IM there are participants who received causal feedback after round 1. In condition BL, all participants received feedback on the cause of the adverse event in round 1 before proceeding to round 2. The same is true in condition IM for the 227 participants whose bid for causal feedback was accepted. Of the participants who received feedback that the adverse event could be attributed to both players choosing the high-stress action (anthropogenic cause), 45% reduced anthropogenic stress on the system in round 2 (see Figure 4, yellow bar, center-right). Comparing the propensity to reduce stress without causal attribution and after attribution to an anthropogenic cause, we find that the difference is statistically significant, both when we condition on the history in the case without attributability (Equation 1a) or when we pool across histories (Equation 2a, two-sided tests of proportions: 30% vs. 45%, χ^2=4.38, p=0.036 in the former case and 32% vs. 45%, χ^2=4.62, p=0.032 in the latter case).

Of the participants who received feedback in conditions BL and IM that the adverse event could be attributed to the urn draw (natural cause), 24% switched from high to low stress in round 2 (see Figure 4, blue bar, center-left). Comparing the propensity to reduce anthropogenic stress on the system without causal attribution and after attribution to a natural cause, we find a difference which is significant, but less clearly so. In two-sided tests of proportions, we find 35% vs. 24%, χ^2=3.17, p=0.075 when we condition on the history HLr and 32% vs. 24%, χ^2=2.09, p=0.148 when we pool across HHg, HLr, and HHr. Relative to the baseline of no attributability, attribution of the adverse event to a natural cause therefore tends to increase future anthropogenic stress, but the effect is less clear cut than the decrease in stress following attribution to an anthropogenic cause.

Of the participants who received feedback that the adverse event could be attributed to both the natural and the anthropogenic cause, 34% reduced anthropogenic stress on the system in round 2 (see Figure 4, grey bar, right). Comparing the propensity to reduce stress without causal attribution and after joint attribution to both causes, we find that the difference is statistically insignificant (two-sided tests of proportions, 29% vs 34%, χ^2=0.18, p=0.670 when testing (1c) and 32% vs 34%, χ^2=0.01, p=0.905 when testing (2c)). Attribution of the adverse event to joint causation leads to future anthropogenic stress that is statistically indistinguishable

from the baseline without causal attribution.

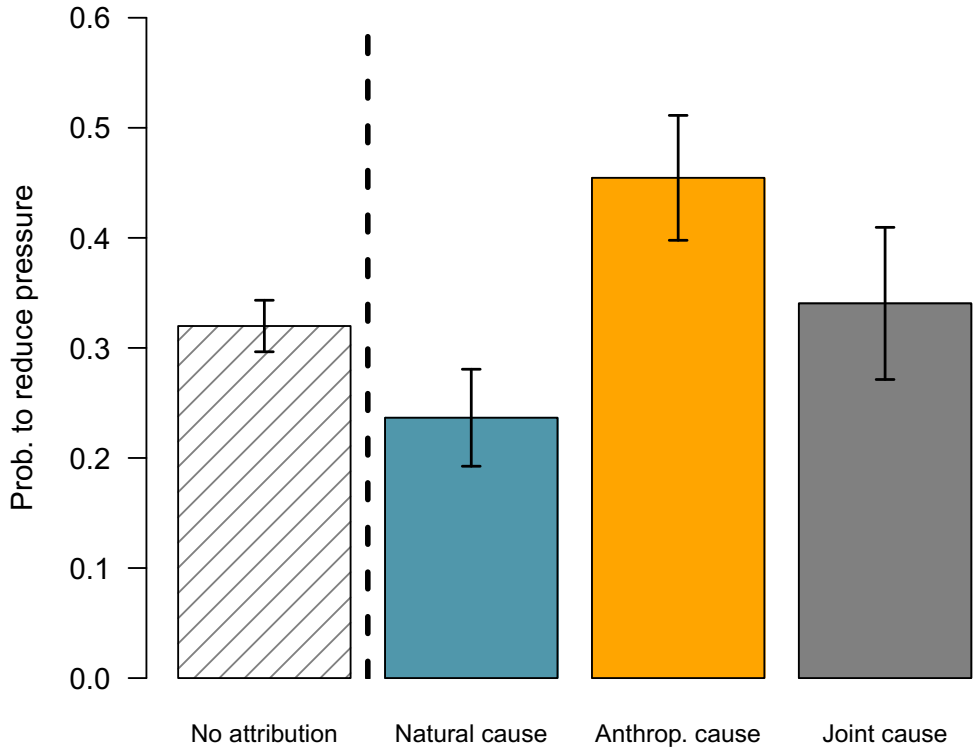

Figure 4: Awareness of the natural or anthropogenic cause of an adverse event has a behavioral impact. The bar plot shows the shares of participants reducing anthropogenic stress in round 2 after an adverse event in round 1. The share when the cause is unknown is about one third (32%, hatched, left). When the cause is known to have been natural, the share is significantly lower at about one quarter (24%, blue, mid-left). When the cause is known to have been anthropogenic, the share is significantly higher at about one half (45%, yellow, mid-right). When the cause was both natural and anthropogenic, the share is about the same as when the cause is unknown (34%, grey, right). Error bars, mean ± s.e.m. (see Supplementary Information for statistical analysis and further results).

Table 2 reports on additional econometric analysis of the experimental data to examine whether the findings are robust to an alternative empirical methodology. The results from a probit model are presented in three specifications. We run a basic specification (column 1) plus extended specifications that include control variables for age, female gender, self-reported propensity to take risks, and educational background (column 2), and within-experiment non-choice behavior such as the number of attempts required to pass the comprehension test and the time required to read the instructions (column 3).

The analysis supports the conclusions drawn from non-parametric testing. In specification [1], we see that, relative to the baseline without attributability, at-

tributing an adverse event to natural cause quantitatively decreases the propensity to reduce anthropogenic stress while attributing it to an anthropogenic cause increases it. A joint cause has no effect. Statistically, only attribution to an anthropogenic cause has a significant effect on outcomes. Adding demographic controls in specification [2] does not change the conclusion, but highlights the influence of risk preferences: Participants that self-report higher risk tolerance are less likely to reduce anthropogenic stress on the system. The results are also robust to adding individual-level experimental non-choice data in specification [3] such as the performance in the task comprehension test and the speed of reading the instructions. Quantitatively and statistically, the effects of natural, anthropogenic, and joint causes remain stable across specifications.

Result 1 (Attributability Effect) *There is evidence for an attributability effect: Causal feedback after an adverse event affects behavior, particularly when participants receive feedback about a human cause.*

One implication of the attributability effect is that attributability to anthropogenic causes contributes to preventing anthropogenic adverse events. Comparing the stress imposed on the system by participants ignorant about the anthropogenic cause of a past adverse event and the stress imposed on the system by players that can attribute the event to an anthropogenic cause, we find that the probability of a future anthropogenic adverse event declines from 29.4% in the former group to 17.9% in the latter. In our experiment, attribution to anthropogenic causes therefore helps prevent one third of future man-made adverse events through behavioral change.

4.2 The causal attribution effect

To test for the causal attribution effect, we compare the propensity to reduce anthropogenic stress after attribution to a natural cause and attribution to an anthropogenic cause. The former is given by $\mathrm{E}[Y|\mathcal{H}{=}\mathrm{HHg} \wedge F{=}1]{=}0.45$ and the latter by $\mathrm{E}[Y|\mathcal{H}{=}\mathrm{HLr} \wedge F{=}1]{=}0.24$. The difference between the two propensities is highly significant ($p{=}0.004$, two-sided test of proportions).

Result 2 (Causal Attribution Effect) *There is evidence for a causal attribution effect: The change in behavior after experiencing an adverse event is greater for subjects whose event happened on account of an anthropogenic, rather than a natural, cause.*

Table 2: Probit regression to explain reduction of anthropogenic stress

	Dependent variable:		
	Reduced stress		
	(1)	(2)	(3)
Nat.cause	−0.249	−0.221	−0.183
	(0.157)	(0.159)	(0.161)
Anthro.cause	0.354**	0.391**	0.402**
	(0.157)	(0.160)	(0.161)
Joint cause	0.057	0.068	0.052
	(0.200)	(0.202)	(0.203)
Age		0.003	0.004
		(0.004)	(0.005)
Female		0.197*	0.184*
		(0.109)	(0.110)
Risk tolerance		−0.057***	−0.063***
		(0.021)	(0.021)
Education		0.030	0.027
		(0.041)	(0.041)
Comprehension			0.103*
			(0.055)
Speed			−0.001
			(0.001)
Constant	−0.468***	−0.474*	−0.738**
	(0.065)	(0.259)	(0.313)
Observations	614	614	614
Log Likelihood	−382.911	−375.378	−373.053
Akaike Inf. Crit.	773.822	766.756	766.105

*Note: The baseline is the propensity to reduce anthropogenic stress when causal attribution is not possible. The sample consists of participants who have experienced adverse events that could have either a natural or anthropogenic cause. Risk tolerance is self reported. Education is the highest level of school or college completed. Comprehension is the number of attempts required to pass the comprehension test. Speed is the time spent reading the experimental instructions. *p<0.1; **p<0.05; ***p<0.01*

4.3 The willingness to pay for causal feedback

To study the demand for causal feedback, we examine the bids submitted after round 1 in treatment IM and after round 2 in treatment FI. For round 1 feedback, we find an average bid of E[$b|\mathcal{H}\in$\{HHg, HLr, HHr\} $\wedge$ T=IM]=\$0.07. For round 2 feedback, we find an average bid of E[$b|\mathcal{H}\in$\{HHg, HLr, HHr\} $\wedge$ T=FI]=\$0.09. These bids are non-hypothetical and can therefore be taken as first-order estimates of participants' willingness-to-pay. Both estimates are significantly different from zero (round 1: $p < 0.001$; round 2: $p < 0.001$, one–sided t-tests[15]). Hypothesis 3 can therefore be rejected.

Result 3 (Demand for causal feedback) *There is evidence for a demand for causal attribution: Feedback after an adverse event attracts a positive average bid.*

To test for differences in the demand for causal feedback after round 1 and round 2, we compare the bids submitted after round 1 in treatment IM (\$0.07) and after round 2 in treatment FI (\$0.09) to each other rather than to zero (as for Hypothesis 3). We find that the difference between the average bids after the two rounds is not significant (p=0.24, two-sided t-tests). Hypothesis 4 can therefore not be rejected. Based on the average bid, subjects do not seem to attach an additional value to causal feedback after round 1 that can be used as an input into decision-making for round 2 compared to causal feedback after round 2 that cannot be used as an input into future decisions.

Result 4 (Differences in demand for causal feedback) *Demand for backward-looking causal feedback does not differ by round: The causal feedback after an adverse event attracts the same average bid after round 1 as after round 2.*

A slightly more nuanced picture arises at the extensive margin, that is the share of participants who express an interest in receiving causal feedback, irrespective of willingness-to-pay. In the IM treatment, 77% of participants express an interest in causal feedback after round 1, against 59% of the participants in the FI treatment, i.e. after round 2. This difference is significant (p <0.001, two-sided test of proportions). This is indicative of a higher interest in causal feedback when it can influence a future decision. The interest is highly price-elastic, however. In both

[15]To test whether $b > 0$ while $b \in [0, 1]$ (hence whether b is different from its boundary value) we employ a two step procedure suitable for large samples. In step 1, the outcome variable is regressed on a constant using OLS to obtain the residuals from the mean. In step 2, we employ a one-sided test to determine whether the mean is different from zero (degrees-of-freedom for the test whether $b > 0$: 153 in the IM treatment, and 130 in the FI treatment).

treatments, about half of those that want to receive full causal feedback are also willing to pay a positive amount for it (IM: 51%, FI: 54%, p=0.632, two-sided test of proportions). Restricting the analysis to those with a strictly positive willingness to pay, willingness-to-pay is \$0.14 in the IM treatment and \$0.17 in the FI treatment. This difference is also not significant (p=0.31).

In Table 3, we explore the covariates of the demand for attribution with help of a linear probability model. The first two columns show the results for the IM treatment, where participants were asked between round 1 and round 2, and last two columns show the results for the FI treatment, where participants were asked at the end of the game.

Columns (1) and (3) show whether participants were interested in knowing the color of the ball and the co-player's action. In line with rationality, we find that those participants that chose action H but did not experience an adverse event are less interested in learning the color of the ball or the action of the co-player – these participants can deduce from the outcome that the ball must have been green and their co-player must have played L. In contrast, those that played H and experienced an adverse event are as likely to have a positive demand for causal attribution than those that played L and did not experience an adverse event. Both types of players can neither deduce the co-player's action nor the color of the ball. Participants that chose action L but experienced an adverse event can deduce that the ball must have been red, but they cannot deduce the co-player's action. These participants are as likely to demand causal attribution as the two aforementioned types of players. These effects hold irrespective of whether causal attribution was offered in between round 1 and 2, or at the end of the game.

Columns (2) and (4) show regression results from a linear model explaining players' willingness to pay for attribution (conditional on wanting to know what caused the outcome). In the FI treatment (column 4), we do not find any differences between player's experienced history, but we do see some differences in the IM treatment (column 2). Interestingly, players that exerted high stress and experienced an adverse event are willing to pay least for causal attribution, while players that did not exert high stress but still experienced an adverse event are willing to pay most for causal attribution. Moreover, player specific covariates play a role, with education and comprehension being positively associated with the willingness to pay for causal attribution. Gender plays a role in the FI treatment, but not in the IM treatment and age plays no role in either treatment.

Table 3: Regressions to explain demand for attribution

	Dependent variable:			
	IM Treatment		FI Treatment	
	wants_to_know	wtp	wants_to_know	wtp
	(1)	(2)	(3)	(4)
Action H	-0.121^{***}	-0.042	-0.165^{***}	-0.030
	(0.043)	(0.028)	(0.047)	(0.033)
Adverse	-0.009	0.035^{*}	0.009	0.006
	(0.032)	(0.019)	(0.040)	(0.025)
Action H × Adverse	0.120^{**}	-0.079^{**}	0.167^{***}	0.055
	(0.058)	(0.037)	(0.061)	(0.041)
Age	0.001	0.001	0.001	0.001^{*}
	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.041	-0.005	0.074^{**}	0.043^{**}
	(0.027)	(0.017)	(0.029)	(0.019)
Risk tolerance	0.017^{***}	0.035^{***}	0.026^{***}	0.042^{***}
	(0.005)	(0.003)	(0.005)	(0.004)
Education	0.018	0.025^{***}	0.025^{**}	0.036^{***}
	(0.011)	(0.008)	(0.012)	(0.008)
Comprehension	0.002	0.046^{***}	0.015	0.058^{***}
	(0.013)	(0.008)	(0.014)	(0.008)
Speed	-0.0004	-0.001^{***}	0.0003	-0.001^{***}
	(0.0003)	(0.0002)	(0.0003)	(0.0002)
Constant	0.609^{***}	-0.329^{***}	0.328^{***}	-0.527^{***}
	(0.082)	(0.053)	(0.083)	(0.057)
Observations	1,002	773	1,031	709
Adjusted R^2	0.022		0.042	
Wald Test (df = 9)		363.623^{***}		286.707^{***}

Note: The baseline is the demand of participant that chose action L and did not experience an adverse event. The sample consists of all participants in the IM and the FI treatments. Risk tolerance is self reported. Education is the highest level of school or college completed. Comprehension is the number of attempts required to pass the comprehension test. Speed is the time spent reading the experimental instructions. Columns (1) and (3) present results from OLS regressions and columns (2) and (4) results from Tobit regressions. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

5 Discussion

The experimental evidence unearthed in the previous section demonstrates the important behavioral consequences of extreme event attribution. While these results offer first evidence that EEA can have behavioral impacts and that these impacts largely align with the hopes around EEA, they also have to be qualified in light of the specific setting in which they were derived.

The experimental setting differs in at least three important respects from climate change: One, the cause of the adverse event was disclosed with certainty. This certainty has three dimensions: One is model certainty, in that the bicausal system of the experiment was accurately specified and exhaustively described. Clearly, this is not the case in climate science. Even though the climate system is reasonably well understood, there are still considerable model uncertainties (Berger et al., 2017). The second dimension is attributive certainty, in that system events in the experiment have either a pure natural cause, pure anthropogenic cause, or balanced double cause. In the climate change context, such certainty in event attribution is a limit case: It is only consistent with those cases in the EEA literature in which the presence or absence of an anthropogenic cause of an extreme event can be identified with near-certainty. Such limit cases have already occurred and will occur with increasing frequency as climate change progress.[16] The majority of EEA statements, however, will come in the shape of changed likelihoods. For example, the likelihood of the July 2023 heatwave in China increased being from 1 in 250 year event without to a 1 in 5 years event with anthropogenic climate change (Zachariah et al., 2023). This means that a July heatwave in China still has a small, but non-vanishing chance of having occurred naturally. In our experiment, subjects responded to an extreme event when both natural and anthropogenic causes were present in equal measure in the same way as to the same event not being attributable. The third dimension of certainty is trusted attribution: Participants of the experiment have no reason to doubt the veracity of the causal feedback provided. That such an absence of attribution skepticism is not a given in the climate context has been extensively studied by social scientists (Dunlap, 2013).

A second respect in which the experimental environment differs clearly from climate change is that it considers a limit case of two players. This implies a setting in which individual behavioral change has the potential to significantly contribute to stress reduction in the system. Climate change, by contrast, is a social dilemma in which individual behavioral change has a negligible impact on risk (Diederich et al.,

[16] See footnote 2 for the relevant studies.

2016; Nordhaus, 1993). This could mean that the incentives for behavioral change are structurally weaker, if not negligible, in the climate context.[17] On the other hand, the incentive structure that climate change presents to individuals provides very limited traction for individual climate action based on narrow consequentialist reasoning, yet a significant share of individuals engages in voluntary climate action (Diederich and Goeschl, 2014). The motives that drive individual climate action may therefore well be amenable to Attributability and Causal Attribution Effects that are themselves difficult to reconcile with such reasoning.

A third dimension in which the experimental environment differs from climate change is that adverse events took the form of zero-payoff outcomes for participants. Such outcomes are clearly not comparable to material and immaterial losses that human populations suffer when adverse climate events strike in the form of extreme heat waves or flooding. At the same time, suffering a zero pay-off event is in fact costly to participants in the experiment who incur the opportunity cost of time spent on an experiment that yielded no reward. We interpret our findings therefore as likely constituting a lower bound, given loss aversion.

Given the strength of the effects in what is a demanding setting for EEA, we believe that the results call for more research on the potential of EEA to foster pro-environmental behavior both in a climate context and in other settings of man-made environmental change. We outline a research agenda in the next section. We also believe that the results support the policy idea of continuing public investment in the infrastructure that supports EEA. Such a policy appears to have merit on two grounds: It looks like to constitute a climate policy measure conducive to climate action by the general population. And EEA addresses a genuine demand among the population for receiving feedback on the causes of extreme climate events that people are beginning to experience with increasing frequency.

6 Conclusion

This paper set out to provide a rigorous proof-of-principle of whether backward-looking causal attribution of adverse events to natural or anthropogenic causes can be behaviorally impactful. The rigor comes from assessing the impacts of extreme event attribution in a demanding setting: More than 3,000 comprehension-checked "workers" on an online platform take decisions with monetary consequences in a

[17] It also means that failure to detect by means of observational data a behavioral impact of EEA in a climate context cannot be taken to suggest that causal attributability is not behaviorally impactful in principle.

bicausal system of low complexity, in which causal feedback is uninformative. This setting largely neuters possible cognitive benefits of EEA for system understanding, favors consequentialist rationales for action, and removes possible contextual framings associated with climate change. Yet, we find strong evidence that causal attribution after an adverse event matters for future behavior. Moreover, the behavioral change induced by feedback about anthropogenic causes leads to reduced future anthropogenic contributions to adverse events. Participants give up money in order to acquire causal feedback when it is not freely available. These results affirm the notion that EEA could be more than just a better way of communicating climate science, but a way to influence the mitigation behavior of the general public.

The proof of principle provided by our experimental evidence provides an impetus for future research in several directions. Here, we sketch a road map to guide such a research agenda. Apart from the obvious tasks of replication and validation, for example in the direction of effect persistence, one direction on this map leads towards learning more about the underlying mechanisms that drive the attributability, attribution, and feedback demand effects observed in the present experiment. Several candidate mechanisms could be responsible. Subjects could be processing feedback between rounds in ways that systematically differ from standard assumptions, such a non-Bayesian updating (Epstein et al., 2010) or parsimonious heuristics (Ba et al., 2022). Subjects might also be responding to anticipated regret, possibly amplified by social interaction (Lauharatanahirun et al., 2012; Cooper and Rege, 2011). As emphasized by psychological research, notions of responsibility or guilt (Swim et al., 2009) could be important. Theories of 'sense-making' could provide an explanation for human desire to resolve the uncertainty about the history of the game (Loewenstein, 1994; FeldmanHall and Shenhav, 2019). A literature in behavioral decision-making has been exploring differences between learning from description and learning from experience (Lejarraga et al., 2016; Lejarraga and Müller-Trede, 2017). Our results raise the possibility that the so-called 'learning-experience gap' is mediated by the cause of the experience. Each of these candidate mechanism gives rise to testable hypotheses within suitably designed follow-up experiments. Treatments in which subjects would be assigned an action in the first round rather than choosing one, for example, could help understand the role of responsibility in driving the effects uncovered here. If subjects also responded differently to zero-payoff events even when they had no active role in choosing their first-round action, for example, this would mean that responsibility or guilt are likely less important drivers.

A second direction on the road map leads towards an unpacking of the attribu-

tion certainty in the present experiment. A better understanding of how uncertainty about the cause affects the attributability, attribution, and feedback demand effects uncovered here is required. For example, this includes testing of how large the weight of the evidence has to be for participants to exhibit the attribution effect. This should include the possibility of motivated beliefs arising through the right channels. One possibility would be to add an error rate to the causal feedback, providing participants with a degree of freedom to attribute the event to a natural cause. One hypothesis could be that subjects will exploit any positive error rate in order to channel their motivated beliefs. On the other hand, it may not have an impact as long as the error rate is small. To address attribution skepticism, future experimental designs could add a third player as the conveyor of feedback and vary that player's real or perceived interest in certain outcomes of the game.

A third direction for future research is to examine the structural features of the game, most prominently the group size. This could be done by employing suitably modified versions of game forms such as the collective risk dilemma (Milinski et al., 2008) and then investigating how the attributability effect, the attribution effect, and demand for causal feedback change as group size increases.

Taken together, these steps contribute to closing the gap between the present experiment, which provides a proof-of-principle, and the application context of climate change. Ultimately, and in an ideal world, natural field experiments that selectively treat populations with different EEA products could be conducted, perhaps even involving the observation of experienced losses among affected populations. In the meantime, we anticipate more corroborating evidence to emerge from the pursuit of the research agenda sketched.

Data availability statement

Instructions, data, and analysis files are freely and openly available online at `https://osf.io/y58mg/?view_only=de73fd8da6354fb39b7c77d736bc1304`.

References

Allen, Myles. 2003. Liability for climate change. *Nature* 421 (6926): 891–892.

Allen, Myles R, and Richard Lord. 2004. The blame game. *Nature* 432 (7017): 551–552.

Arias, P.A., N. Bellouin, E. Coppola, R.G. Jones, G. Krinner, J. Marotzke, V. Naik,

et al. 2021. *Technical Summary.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 33–144.

Atalay, Enghin, Ali Hortacsu, James Roberts, and Chad Syverson. 2011. Network structure of production. *Proceedings of the National Academy of Sciences* 108 (13): 5199–5202.

Ba, Cuimin, J Aislinn Bohren, and Alex Imas. 2022. Over-and underreaction to information. *Available at SSRN 4274617* .

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. Guilt in games. *American Economic Review* 97 (2): 170–176.

Berger, Loïc, Johannes Emmerling, and Massimo Tavoni. 2017. Managing catastrophic climate risks under model uncertainty aversion. *Management Science* 63 (3): 749–765.

Blount, Sally. 1995. When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes* 63 (2): 131 – 144.

Boudet, Hilary, Leanne Giordono, Chad Zanocco, Hannah Satein, and Hannah Whitley. 2020. Event attribution and partisanship shape local discussion of climate change after extreme weather. *Nature Climate Change* 10 (1): 69–76.

Brun, Wibecke. 1992. Cognitive components in risk perception: Natural versus manmade risks. *Journal of Behavioral Decision Making* 5 (2): 117–132.

Buckenmaier, Johannes, and Eugen Dimant. 2021. The experience is (not) everything: Sequential outcomes and social decision-making. *Economics Letters* 205: 109916.

Bulte, Erwin, Shelby Gerking, John A List, and Aart De Zeeuw. 2005. The effect of varying the causes of environmental problems on stated wtp values: evidence from a field study. *Journal of environmental economics and management* 49 (2): 330–342.

Carlson, Joshua M, Madeline Voltz, John Foley, Lisa Gentry, and Lin Fang. 2022. Changing how you look at climate change: Attention bias modification increases attention to climate change. *Climatic Change* 175 (3): 16.

Chen, Daniel L, Martin Schonger, and Chris Wickens. 2016. otree: An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* 9: 88–97.

Cooper, David J, and Mari Rege. 2011. Misery loves company: Social regret and social interaction effects in choices under risk and uncertainty. *Games and Economic Behavior* 73 (1): 91–110.

Demski, Christina, Stuart Capstick, Nick Pidgeon, Robert Gennaro Sposato, and Alexa Spence. 2017. Experience of extreme weather affects climate change mitigation and adaptation responses. *Climatic Change* 140: 149–164.

Deryugina, Tatyana. 2013. How do people update? the effects of local weather fluctuations on beliefs about global warming. *Climatic change* 118: 397–416.

Diederich, Johannes, and Timo Goeschl. 2014. Willingness to pay for voluntary climate action and its determinants: Field-experimental evidence. *Environmental and Resource Economics* 57 (3): 405–429.

Diederich, Johannes, Timo Goeschl, and Israel Waichman. 2016. Group size and the (in) efficiency of pure public good provision. *European Economic Review* 85: 272–287.

Drake, Henri F, and Geoffrey Henderson. 2022. A defense of usable climate mitigation science: how science can contribute to social movements. *Climatic Change* 172 (1-2): 10.

Dunlap, Riley E. 2013. Climate change skepticism and denial: An introduction. *American behavioral scientist* 57 (6): 691–698.

Epstein, Larry G, Jawwad Noor, and Alvaro Sandroni. 2010. Non-bayesian learning. *The BE Journal of Theoretical Economics* 10 (1): 00001022021935170416 23.

Erisman, Jan Willem, Guy Brasseur, Philippe Ciais, Nick van Eekeren, and Thomas L Theis. 2015. Global change: Put people at the centre of global risk management. *Nature News* 519 (7542): 151.

Ettinger, Joshua, Peter Walton, James Painter, Susan A Flocke, and Friederike EL Otto. 2023. Extreme weather events as teachable moments: Catalyzing climate change learning and action through conversation. *Environmental Communication* 17 (7): 828–843.

Ettinger, Joshua, Peter Walton, James Painter, Shannon Osaka, and Friederike EL Otto. 2021. "what's up with the weather?" public engagement with extreme event attribution in the united kingdom. *Weather, Climate, and Society* 13 (2): 341–352.

FeldmanHall, Oriel, and Amitai Shenhav. 2019. Resolving uncertainty in a social world. *Nature human behaviour* 3 (5): 426–435.

Feldmann, Anja, Oliver Gasser, Franziska Lichtblau, Enric Pujol, Ingmar Poese, Christoph Dietzel, Daniel Wagner, et al. 2020. The lockdown effect: Implications of the covid-19 pandemic on internet traffic. In *Proceedings of the ACM Internet Measurement Conference*. 1–18.

Gifford, Robert. 2011. The dragons of inaction: psychological barriers that limit climate change mitigation and adaptation. *American psychologist* 66 (4): 290.

Grantham, Brian A, Francis Chan, Karina J Nielsen, David S Fox, John A Barth, Adriana Huyer, Jane Lubchenco, et al. 2004. Upwelling-driven nearshore hypoxia signals ecosystem and oceanographic changes in the northeast pacific. *Nature* 429 (6993): 749–754.

Hallegraeff, GM. 2003. Harmful algal blooms: a global overview. *Manual on harmful marine microalgae* 33: 1–22.

Hazlett, Chad, and Matto Mildenberger. 2020. Wildfire exposure increases pro-environment voting within democratic but not republican areas. *American Political Science Review* 114 (4): 1359–1365.

Helbing, Dirk. 2013. Globally networked risks and how to respond. *Nature* 497 (7447): 51–59.

Herring, Stephanie C, Nikolaos Christidis, Andrew Hoell, Martin P Hoerling, and Peter A Stott. 2021. Explaining extreme events of 2019 from a climate perspective. *Bulletin of the American Meteorological Society* 102 (1): S1–S116.

Hestermann, Nina, and Yves Le Yaouanq. 2021. Experimentation with self-serving attribution biases. *American Economic Journal: Microeconomics* 13 (3): 198–237.

Hindsley, Paul R, and O Ashton Morgan. 2020. The role of cultural worldviews in willingness to pay for environmental policy. *Environmental and Resource Economics* : 1–27.

Hoogendoorn, Gea, Bernadette Sütterlin, and Michael Siegrist. 2020. The climate change beliefs fallacy: The influence of climate change beliefs on the perceived consequences of climate change. *Journal of Risk Research* 23 (12): 1577–1589.

Howe, Peter D, Jennifer R Marlon, Matto Mildenberger, and Brittany S Shield. 2019. How will climate change shape climate opinion? *Environmental Research Letters* 14 (11): 113001.

Hulme, Mike. 2014. Attributing weather extremes to 'climate change' a review. *Progress in Physical Geography* 38 (4): 499–511.

Imada, Yukiko, Hideo Shiogama, Chiharu Takahashi, Masahiro Watanabe, Masato Mori, Youichi Kamae, and Shuhei Maeda. 2018. 19. climate change increased the likelihood of the 2016 heat extremes in asia. *Bulletin of the American Meteorological Society* 99 (1): S97–S101.

IPBES. 2019. *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.* Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

James, Rachel A., Richard G. Jones, Emily Boyd, Hannah R. Young, Friederike E. L. Otto, Christian Huggel, and Jan S. Fuglestvedt. 2019. *Attribution: How Is It Relevant for Loss and Damage Policy and Practice?* Cham: Springer International Publishing, 113–154.

Jang, S Mo. 2013. Framing responsibility in climate change discourse: Ethnocentric attribution bias, perceived causes, and policy attitudes. *Journal of Environmental Psychology* 36: 27–36.

Jézéquel, Aglaé, Vivian Dépoues, Hélène Guillemot, Amélie Rajaud, Mélodie Trolliet, Mathieu Vrac, Jean-Paul Vanderlinden, et al. 2020. Singular extreme events and their attribution to climate change: A climate service–centered analysis. *Weather, Climate, and Society* 12 (1): 89–101.

Kew, Sarah, Izidine Pinto, Lincoln Alves, Djacinto Santos, Renata Libonati, MLR Liberato, S Philip, et al. 2023. Strong influence of climate change in uncharacteristic early spring heat in south america. *Grantham Research Institute, Imperial College* .

Kim, Jeong-Yoo, Haeree Lee, Insik Min, et al. 2013. The economics of curiosity. *Korean Economic Review* 29: 23–50.

Knutson, Thomas R, Jonghun Kam, Fanrong Zeng, and Andrew T Wittenberg. 2018. Cmip5 model-based assessment of anthropogenic influence on record global warmth during 2016. *Bulletin of the American Meteorological Society* 99 (1): S11–S15.

Kriss, Peter H, George Loewenstein, Xianghong Wang, and Roberto A Weber. 2011. Behind the veil of ignorance: Self-serving bias in climate change negotiations. *Judgment and Decision Making* 6 (7): 602–615.

Labovitz, Craig, Abha Ahuja, and Farnam Jahanian. 1999. Experimental study of internet stability and backbone failures. In *Digest of Papers. Twenty-Ninth Annual International Symposium on Fault-Tolerant Computing (Cat. No. 99CB36352)*. IEEE, 278–285.

Larcom, Shaun, Po-Wen She, and Terry van Gevelt. 2019. The uk summer heatwave of 2018 and public concern over energy security. *Nature Climate Change* 9 (5): 370–373.

Lauharatanahirun, Nina, George I Christopoulos, and Brooks King-Casas. 2012. Neural computations underlying social risk sensitivity. *Frontiers in human neuroscience* 6: 213.

Leichenko, Robin, and Karen O'Brien. 2008. *Environmental change and globalization: double exposures*. Oxford University Press.

Lejarraga, Tomás, and Johannes Müller-Trede. 2017. When experience meets description: How dyads integrate experiential and descriptive information in risky decisions. *Management Science* 63 (6): 1953–1971.

Lejarraga, Tomás, Jan K Woike, and Ralph Hertwig. 2016. Description and experience: How experimental investors learn about booms and busts affects their financial risk taking. *Cognition* 157: 365–383.

Liu, Peng, and Yong Du. 2022. Blame attribution asymmetry in human–automation cooperation. *Risk Analysis* 42 (8): 1769–1783.

Loewenstein, George. 1994. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin* 116 (1): 75.

Lohmann, Paul M, and Andreas Kontoleon. 2023. Do flood and heatwave experiences shape climate opinion? causal evidence from flooding and heatwaves in england and wales. *Environmental and Resource Economics* 86 (1-2): 263–304.

Loomes, Graham, and Robert Sugden. 1982. Regret theory: An alternative theory of rational choice under uncertainty. *The economic journal* 92 (368): 805–824.

Loy, Laura S, and Alexa Spence. 2020. Reducing, and bridging, the psychological distance of climate change. *Journal of Environmental Psychology* 67: 101388.

Lusk, Greg. 2017. The social utility of event attribution: liability, adaptation, and justice-based loss and damage. *Climatic Change* 143 (1): 201–212.

Malmendier, Ulrike. 2021. Exposure, experience, and expertise: Why personal histories matter in economics. *Journal of the European Economic Association* 19 (6): 2857–2894.

Marjanac, Sophie, Lindene Patton, and James Thornton. 2017. Acts of god, human influence and litigation. *Nature Geoscience* 10 (9): 616–619.

McDonald, Rachel I, Hui Yi Chai, and Ben R Newell. 2015. Personal experience and the 'psychological distance'of climate change: An integrative review. *Journal of environmental psychology* 44: 109–118.

Milinski, Manfred, Ralf D Sommerfeld, Hans-Jürgen Krambeck, Floyd A Reed, and Jochem Marotzke. 2008. The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences* 105 (7): 2291–2294.

Murtagh, Niamh, Birgitta Gatersleben, and David Uzzell. 2012. Self-identity threat and resistance to change: Evidence from regular travel behaviour. *Journal of Environmental Psychology* 32 (4): 318–326.

Nordhaus, William D. 1993. Reflections on the economics of climate change. *Journal of economic Perspectives* 7 (4): 11–25.

Ogunbode, Charles A, Christina Demski, Stuart B Capstick, and Robert G Sposato. 2019. Attribution matters: Revisiting the link between extreme weather experience and climate change mitigation responses. *Global Environmental Change* 54: 31–39.

Oldenborgh, Geert Jan, Karin Wiel, Sarah Kew, Sjoukje Philip, Friederike Otto, Robert Vautard, Andrew King, et al. 2021. Pathways and pitfalls in extreme event attribution. *Climatic Change* 166 (1): 1–27.

Osaka, Shannon, and Rob Bellamy. 2020. Natural variability or climate change? stakeholder and citizen perceptions of extreme event attribution. *Global Environmental Change* 62: 102070.

Otto, Friederike EL. 2017. Attribution of weather and climate events. *Annual Review of Environment and Resources* 42: 627–646.

Pfrommer, Tobias, Timo Goeschl, Alexander Proelss, Martin Carrier, Johannes Lenhard, Henrike Martin, Ulrike Niemeier, et al. 2019. Establishing causation in climate litigation: admissibility and reliability. *Climatic Change* 152 (1): 67–84.

Reser, Joseph P, and Janet K Swim. 2011. Adapting to and coping with the threat and impacts of climate change. *American Psychologist* 66 (4): 277.

Rotter, Julian B. 1966. Generalized expectancies of internal versus external control of reinforcements. *Psychological Monographs: General and Applied* 80: 1–28.

Sarabia-Sanchez, Francisco J, and Carla Rodriguez-Sanchez. 2016. The role of credibility and negative feelings in comparative perceptual bias related to environmental hazards. *Journal of environmental psychology* 47: 1–13.

Schiermeier, Quirin. 2018. Climate as culprit. *Nature* 560 (7716): 20–22.

Schwerter, Frederik, and Florian Zimmermann. 2020. Determinants of trust: The role of personal experiences. *Games and Economic Behavior* 122: 413–425.

Shepherd, Theodore G, Emily Boyd, Raphael A Calel, Sandra C Chapman, Suraje Dessai, Ioana M Dima-West, Hayley J Fowler, et al. 2018. Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Climatic change* 151: 555–571.

Shreedhar, Ganga, and Susana Mourato. 2020. Linking human destruction of nature to covid-19 increases support for wildlife conservation policies. *Environmental and Resource Economics* 76 (4): 963–999.

Siegrist, Michael, and Bernadette Sütterlin. 2014. Human and nature-caused hazards: The affect heuristic causes biased decisions. *Risk Analysis* 34 (8): 1482–1494.

Singh, Roop K, Julie Arrighi, Erin Coughlan de Perez, Olivia Warrick, Pablo Suarez, Bettina Koelle, Eddie Jjemba, et al. 2019. International conference on climate risk management, inputs for the intergovernmental panel on climate change's sixth assessment report. *Climate and development* 11 (8): 655–658.

Slovic, Paul, Baruch Fischhoff, and Sarah Lichtenstein. 1986. The psychometric study of risk perception. *Risk evaluation and management* : 3–24.

Spence, Alexa, Wouter Poortinga, Catherine Butler, and Nicholas Frank Pidgeon. 2011. Perceptions of climate change and willingness to save energy related to flood experience. *Nature climate change* 1 (1): 46.

Steg, Linda. 2023. Psychology of climate change. *Annual Review of Psychology* 74: 391–421.

Stott, Peter A, Dáithí A Stone, and Myles R Allen. 2004. Human contribution to the european heatwave of 2003. *Nature* 432 (7017): 610–614.

Stuart-Smith, Rupert F, Friederike EL Otto, Aisha I Saad, Gaia Lisi, Petra Minnerop, Kristian Cedervall Lauta, Kristin van Zwieten, et al. 2021. Filling the evidentiary gap in climate litigation. *Nature Climate Change* 11 (8): 651–655.

Swim, Janet, Susan Clayton, Thomas Doherty, Robert Gifford, George Howard, Joseph Reser, Paul Stern, et al. 2009. Psychology and global climate change: Addressing a multi-faceted phenomenon and set of challenges. a report by the american psychological association's task force on the interface between psychology and global climate change. *American Psychological Association, Washington* 66: 241–250.

Van der Linden, Sander. 2015. The social-psychological determinants of climate change risk perceptions: Towards a comprehensive model. *Journal of Environmental Psychology* 41: 112–124.

Vélez, Ricardo, et al. 2002. Causes of forest fires in the mediterranean basin. In *Risk management and sustainable forestry. EFI Proceedings*, volume 42. 35–42.

Walsh, John E, Richard L Thoman, Uma S Bhatt, Peter A Bieniek, Brian Brettschneider, Michael Brubaker, Seth Danielson, et al. 2018. The high latitude marine heat wave of 2016 and its impacts on alaska. *Bull. Am. Meteorol. Soc* 99 (1): S39–S43.

Weber, Elke U. 2006. Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). *Climatic change* 77 (1-2): 103–120.

Webster, Donna M, and Arie W. Kruglanski. 1994. Individual differences in need for cognitive closure. *Journal of personality and social psychology* 67 (6): 1049.

Weiner, Bernard. 1995. *Judgments of responsibility: A foundation for a theory of social conduct*. guilford Press.

Wong-Parodi, Gabrielle, and Nina Berlin Rubin. 2022. Exploring how climate change subjective attribution, personal experience with extremes, concern, and subjective knowledge relate to pro-environmental attitudes and behavioral intentions in the united states. *Journal of Environmental Psychology* 79: 101728.

Yang, Z Janet, Mihye Seo, Laura N Rickard, and Teresa M Harrison. 2015. Information sufficiency and attribution of responsibility: Predicting support for climate change policy and pro-environmental behavior. *Journal of Risk Research* 18 (6): 727–746.

Zachariah, M, S Philip, I Pinto, M Vahlberg, R Singh, F Otto, C Barnes, et al. 2023. Extreme heat in north america, europe and china in july 2023 made much more likely by climate change. *Grantham Institute for Climate Change: London, UK* .

Zhao, Jiaying, and Yu Luo. 2021. A framework to address cognitive biases of climate change. *Neuron* 109 (22): 3548–3551.

Appendix

Table A-1: Participant characteristics by treatment (BL, IM, and FI). The last column gives the smallest p-value of the three pairwise comparisons for sample differences (adjusted for multiple hypotheses testing).

variable	BL	IM	FI	min p-val
Age	39.09	38.98	38.56	0.83
Female	0.45	0.43	0.45	0.94
Education	3.61	3.59	3.49	0.1
Risk tolerance	5.68	5.6	5.47	0.26
Comprehension	3.25	3.25	3.28	0.96
Speed	64.57	63.29	66.54	0.52
N	998	1002	1031	

Note: Risk tolerance is self reported willingness to take risks (between 0 and 10). Education is the highest level of school or college completed (between 0="less than high school degree" and 5="graduate degree". Comprehension is the number of attempts required to pass the comprehension test. Speed is the time spent reading the experimental instructions (in seconds).

Table A-2: Participant characteristics by treatment, for players exerting low stress and experiencing the adverse event

variable	BL	IM	FI	min p-val
Age	39.16	38.98	37.77	0.4
Female	0.45	0.42	0.46	0.7
Education	3.68	3.73	3.62	0.62
Risk tolerance	5.82	5.48	5.62	0.51
Comprehension	3.4	3.47	3.38	0.84
Speed	58.7	55.83	59.68	0.84
N	250	284	297	

Table A-3: Participant characteristics by treatment, for players exerting low stress and not experiencing the adverse event

variable	BL	IM	FI	min p-val
Age	39.25	38.8	38.91	0.9
Female	0.41	0.43	0.43	0.94
Education	3.74	3.63	3.6	0.26
Risk tolerance	5.62	5.59	5.55	0.9
Comprehension	3.38	3.26	3.42	0.16
Speed	58.23	59.12	62.89	0.64
N	413	396	382	

Table A-4: Participant characteristics by treatment, for players exerting high stress and experiencing the adverse event

variable	BL	IM	FI	min p-val
Age	39.4	39.1	39.5	0.93
Female	0.53	0.46	0.48	0.53
Education	3.42	3.42	3.27	0.47
Risk tolerance	5.69	5.55	5.1	0.07
Comprehension	2.92	3.03	3.09	0.21
Speed	81.23	72.19	78.92	0.33
N	194	199	221	

Table A-5: Participant characteristics by treatment, for players exerting high stress and not experiencing the adverse event

variable	BL	IM	FI	min p-val
Age	38.09	39.35	37.76	0.77
Female	0.46	0.45	0.43	0.82
Education	3.35	3.4	3.27	0.78
Risk tolerance	5.6	5.95	5.48	0.48
Comprehension	3.01	3.05	2.98	0.79
Speed	70.66	79.54	71.87	0.57
N	141	123	131	