HEIDELBERG UNIVERSITY
DEPARTMENT OF ECONOMICS

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

The Behavioral Economics of Extreme Event Attribution

Florian Diekert

Timo Goeschl

Christian König-Kersting

AWI DISCUSSION PAPER SERIES NO. 741
January 2024

The Behavioral Economics of Extreme Event Attribution

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

January 4, 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 attracts 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).

Increasingly,[3] the possible usefulness of EEA for influencing the future mitigation behavior of the general public has become an important motivation for conducting EEA studies (Drake and Henderson, 2022; Oldenborgh et al., 2021; Ettinger et al., 2021; Osaka and Bellamy, 2020; Shepherd, 2019; Trenberth et al., 2015). The idea is, in short, that EEA should become behaviorally relevant. To the economist, the possibility of such behavioral relevance is conceptually not obvious: EEA is by nature backward-looking rather than decision-supporting. In other words, EEA sheds light on the nature of past events rather than providing new information about the nature of future events. It is not intended or designed to better predict adverse events. By pursuing a backward-looking causal perspective rather than a forward-looking probabilistic one means that the mechanisms that could possibly underpin

[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]The initial drive for EEA came from two related, but distinct questions: The feasibility of relying on causal evidence in climate-related litigation (Allen, 2003; Allen and Lord, 2004) and the feasibility of attributing (and possibly apportioning) losses and damages from climate change to anthropogenic factors (Mills, 2005; Diffenbaugh et al., 2017). Jézéquel et al. (2020) provides a short social history of EEA.

a behavioral relevance of EEA must lie outside standard consequentialist models of human decision-making. This is because without providing new information on future damages and probabilities, EEA cannot obviously affect the expected utilities associated with different courses of action. In addition to the conceptual uncertainties, empirical evidence has yet to document instances in which conducting EEA studies and making them available has been behaviorally impactful.[4] At most, EEA has been shown to be somewhat conducive to post-event adaptation (Lusk, 2017; Singh et al., 2019), but evidence that it supports mitigation behavior is still lacking.[5] This is, perhaps, unsurprising in light of the challenges of measurement and identification that an observational study on the behavioral impacts of EEA would face.

The present paper employs an experimental approach in order to examine what we regard as the core claim behind the behavioral relevance of EEA. This claim is that even though every extreme event has well understood ex-ante statistical properties that combine natural and anthropogenic factors (captured by climate science), ex-post attribution of a materialized event to natural or anthropogenic causes has the ability to impact on future choices. This impact is 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 claim through experimental manipulations in 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 to receive feedback that makes attribution possible? - To remain faithful to the practice of EEA in a climate context, all three steps must be conducted in an environment with well-understood stochastic properties, limiting consequentialist reasons for attribution.

The economic experiment designed to complete these three steps situates the test in an abstract setting of a bicausal system, thereby avoiding undesirable interference

[4]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).

[5]This is also true for the communication and perception stage, where the evidence is mixed (Howe et al., 2019; Osaka and Bellamy, 2020; Boudet et al., 2020; Ettinger et al., 2021).

with subjects' view on climate change and generalizing to the broader class of bicausal settings.[6] In that setting, individual players, after screening for game comprehension, choose between a 'low stress' and 'high stress' action in two rounds a modified Chicken Game, each time randomly matched with a new co-player. In this Game, adverse events are zero-payoff outcomes that can arise as a result of both participants choosing the 'high stress' action that potentially yields a higher payoff (anthropogenic cause) and as a result of an unfavorable draw from an urn (natural cause). The parameters of the Game are common knowledge and identical in both rounds and across causes: The probability of an unfavorable draw from the urn and the probability of encountering a co-player who chooses the high-stress action are the same and do not change from round to round. The subset of players of interest are those who experienced a zero-payoff event in round 1 that could have both a natural and anthropogenic case.[7] The behavioral outcome variable on which we focus is whether the player changes their action between round 1 and round 2. Treatments vary the feedback environment. One treatment features no attributability: Players only receive feedback after round 1 that a zero-payoff event has occurred, but the feedback is silent on the history of play. Another treatment features full causal attribution: Players receive feedback after round 1 that is explicit about the history of play. In a third treatment, players can submit bids after round 1 for receiving feedback that provides the history of play. Irrespective of the treatment and the event, the decision situation remains essentially the same in round 2 as in round 1.

Our hypotheses about the experimental outcomes among the subset of relevant subjects reflect a narrow consequentialist stance. This stance reflects that in our experimental setting, expected payoffs from engaging in essentially the same game for a second time are not altered by ex-post causal feedback. Such a setting is,

[6]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.

[7]This requires that subjects chose the 'high stress' action in round 1, thus exposing them to the possibility of both causes of adverse events.

first, not predicted to exhibit an *attributability effect*. This effect could be said to be present if we found that individuals switch action significantly more or less often in treatments in which causal attribution is provided compared to when it is not. Our hypothesis is that there are no significant differences in switching action. Second, a *causal attribution effect* could be said to be present if in treatments with causal attribution provided, subjects switch action significantly more often if the adverse event was attributed to an anthropogenic, rather than a natural, cause. Our hypothesis is, again, that there are no significant differences in switching action, irrespective of cause. Third, a *demand for causal attribution* could be said to exist if in treatments in which subjects can bid money for causal feedback, we find that a significant share do so. Our hypothesis is that no such bidding will occur.

The evidence from our online experiment with 3,031 participants leads to three main findings. 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 (42%) compared to when it is not (32%). For a natural cause, the share of subjects switching action is lower when the adverse event is attributable (24%) than when it is not (33%). Second, there is very strong evidence for a causal attribution effect: Given such attribution, an anthropogenic cause of an adverse event makes a significantly higher share of subjects (45%) switch away from the 'high stress' action compared to a natural cause (22%). 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 (75%) express an interest in causal feedback, with around half submitting a positive bid. These results challenge the narrow consequentialist stance of our hypotheses and, as we show, cannot be explained by confusion or lack of attention.

The results are significant for three reasons. First, by comparing the behavior of those who receive feedback on what caused an adverse event they experienced with those who do not the paper provides - to our knowledge - the first pieces of evidence that can speak to the question of the behavioral relevance of extreme event attribution. A rich literature has examined the behavioral relevance of experiencing adverse events (see footnote 4 above), but less so the possibly important heterogeneities among those who experienced an adverse event. Causal attributability of the adverse event emerges from our experiments as a possibly important heterogeneity. Secondly, by establishing divergent behavioral responses for natural and anthropogenic causes among those who experienced an adverse event, the paper adds nuance to our understanding of how people respond to the seemingly inconsequential past of environmental problems. Previously, a small body of research in environmental economics found evidence that causal awareness affects environmental valuation (Bulte et al., 2005; Shreedhar and Mourato, 2020; Hindsley and Ashton Morgan, 2020) and a small body of research in psychology that causal awareness affects people's reported responses to adverse environmental events (Siegrist and Sütterlin, 2014; Hoogendoorn et al., 2020; Liu and Du, 2022). A common element across this research is a reliance on hypothetical scenarios and survey evidence, often favoring context specificity over experimenter control. By deriving from an abstract economic experiment with real payoff consequences under tight experimental control, our evidence is a first proof of principle for a more general behavioral effect that merits further investigation across the wide range of bicausal domains. Finally, by demonstrating the presence of a core demand for causal feedback among our subjects, we provide evidence that there is a social value to EEA. Given the experimental design, this evidence is not easily reconciled with narrowly consequentialist uses of information. It is more in line with the notion of a utility of purely epistemic knowledge for human decision-makers stressed by behavioral economists (Loewenstein, 1994). Similar arguments may apply to EEA (Hulme, 2014). 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.

2 Experimental Design

The three experimental 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). Only one of the two rounds is later randomly selected to determine final study payments. 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.[8] Before they choose their action, players

[8] In the paper we refer to the low-stress action as L and the high-stress action as H for a more

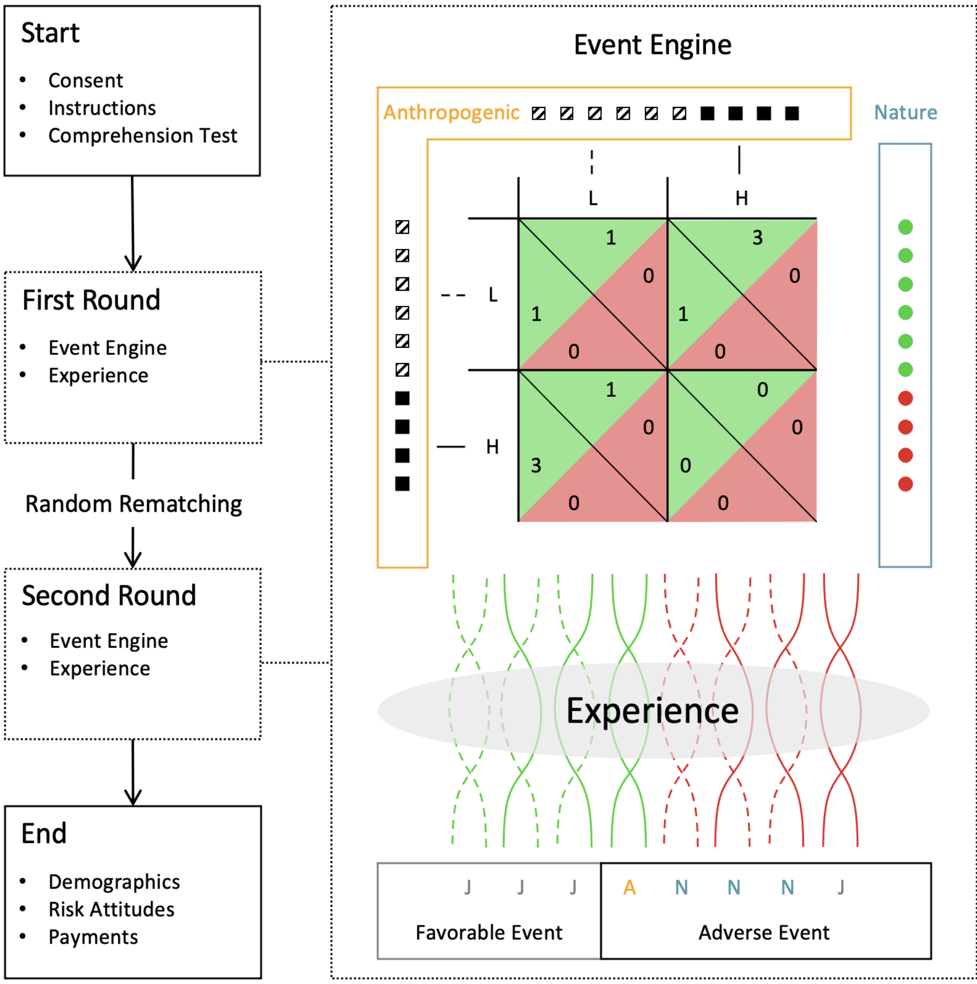


Figure 1: The left panel charts, from top to bottom, the progression of the experiment, from the consent form via the first and second round to experimental payments at the end. The right panel shows, for each round, the path from event determination to experience. From top to bottom, it shows how events 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).

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.[9]

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 additional feedback is given. We refer to causal feedback when 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.

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

intuitive presentation. In the experiment, the actions were abstractly referred to as A and B, respectively.

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

2.1 Treatments

We conducted three treatments that varied whether causal feedback is available after a certain round and if so, whether it needs to be acquired at cost.

Treatment 1 features the treatment with the richest feedback. 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 1 can therefore always attribute past outcomes to specific causes, whether a fortunate or unfortunate realization of the natural cause (urn draw) or the realization of the human cause (strategic choice of co-player), or both. We call this the 'BASELINE' (BL) treatment.

Treatment 2 contains most of the elements of treatment BL, but differs in the feedback made available. After completing round 1, participants only receive event feedback (0, x, or $x + y$), but no causal feedback. After round 2, participant receive event feedback and go through a two-step filter that elicits WTP for causal feedback regarding the last round. 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 for causal feedback, bounded between \$0 and \$0.50. We refer to this treatment variation, in which causal feedback is withheld after round 1 and can only be acquired at cost after round 2, the 'FINAL' (FI) treatment.

Treatment 3 contains all the elements of treatment FI and only differs 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 receiving event feedback on round 1, participants are asked a yes-no question about their interest in causal feedback. If an interest is expressed, a BDM-mechanism elicits their WTP for causal feedback, bounded between \$0 and \$0.50. After round 2, participants automatically receive both event and causal feedback. The alternative timing of the feedback opportunity in treatment 3 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 lower income. At the same time, given the design, participants with richer feedback do not have better information than those in treatment FI: Due to partner re-matching and independent urn draws, the probabilities remain unchanged. We refer to this as the 'INTERIM' (IM) treatment.

2.2 Histories

Because natural and strategic uncertainty are independent, there are eight different histories in round 1 of each treatment. We introduce the following notation: The choice of player i in round t is denoted by $C_{i,t} = \{\mathrm{L}; \mathrm{H}\}$ (parallel for player j). The realization of the urn draw (lottery) is denoted by $L_t = \{\mathrm{red}; \mathrm{green}\}$. The payoff of player i for round t is denoted by $\pi_{i,t}$. We denote the probability that a red ball is drawn in the lottery by p (that is, $p = \Pr(L = \mathrm{red})$) and the probability that a player chooses action H by q. This means, for example, that the history according to which a green ball is drawn from the urn and both players choose L, occurs with probability $(1-p)(1-q)^2$. For easier reference, we adopt the following convention to label the histories: "LLg" refers to the history where player one chose action L, player two chose action L, and the ball was green. "LLr" refers to the history where player one chose action L, player two chose action L, and the ball was red, etc. The histories, their corresponding choices, urn draw outcomes, as well as payoffs and probabilities are shown in Table 1.

Table 1: The eight different histories of round 1

H	$C_{i,1}$	$C_{j,1}$	L	$\pi_{i,1}$	$\Pr(H{=}\mathrm{h}n)$
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

As Table 1 makes clear, there are five histories that lead to a zero payoff event. In two of these, namely 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 and 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.".[10]

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.

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

2.4 Procedures

The experiment was conducted online on Amazon Mechanical Turk. 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.

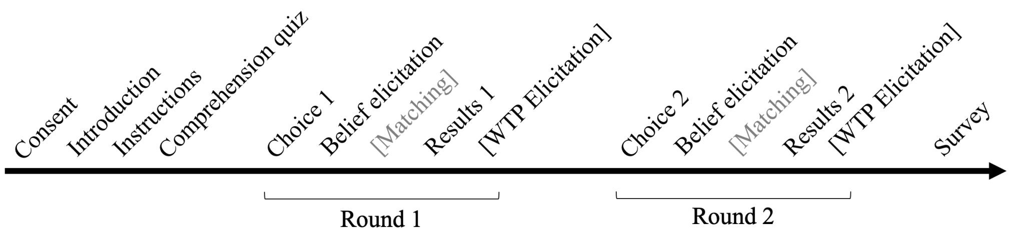


Figure 3: Stages of the experiment

Across conditions FI, IM, and BL, we recruited a total of 3,031 participants. 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 we set the payoffs to x=1 USD and y=2 USD). In condition FI and IM, participants received received a bonus of \$0.50 after completing the 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

3.1 Overview

We have three objectives in the pursuit of the three treatments. The first is to determine the effect of making an adverse event attributable on subsequent choice. This is the *attributability effect.* The second is to determine the effect of having an adverse event attributed to a natural or anthropogenic cause on subsequent choice. This is the *causal attribution effect.* The third objective is to determine the willingness to pay for causal feedback after experiencing an adverse event, if any.

Before operationalizing these three objectives in terms of the experimental data and testing the corresponding hypotheses, the outcome variables of interest need to be defined. For the first two effects, we require an outcome variable that captures behavioral impact. This is an indicator of change in choice in the following way: Let 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 ($C_{i,1}$≠$C_{i,2}$). For the third objective, we require an outcome variable that captures individual willingness to pay. For individual i expressing a willingness to pay for causal feedback on round t, this is a variable $B_{i,t}$.

In addition, recall the variable definitions used in Table 1: The set of histories is denoted by H and the outcome of the urn-draw (lottery) in round t is given by L_t. The treatment condition is BL for "baseline" (treatment BL), FI for "final" (treatment FI), or IM for "interim" (treatment IM). 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 succesfully purchased feedeback) and $F = 0$ when a participant did not receive causal feedback.

3.2 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.[11] Participants are informed about average play in this game form

[11]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.

In line with our stance, we predict that no change in *average* behavior should be observed. A narrowly consequentialist decision-maker that experienced a bicausal adverse event[12] 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.*

To test hypothesis 1, we investigate the change in choice Y_i to test whether causal feedback affects behavior. A total of five histories, namely {LLr,LHr,HHg,HLr,HHr}, generate an adverse event with $\pi_{i,1} = 0$. In two of the three 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, treatment with causal feedback do not substantively differ in these two histories. In three of the five histories, namely {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. As a result, we test Hypothesis 1 by comparing behavior across in the three histories in which the treatment induces differences in causal feedback. Note that by conditioning on history HHg, HLr, or HHr, we consider only those participants with action H in

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

the first round (see Table 1). As a consequence, a value of Y_i=1 uniquely means $C_{i,2}$=L.

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

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

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

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 that 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 may 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)

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

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

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

3.3 The causal attribution effect

Recall the strategic choices available of 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.*

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

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

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.

3.4 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.

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.*

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. We code an expression of no interest $I_i = 0$ as a zero bid $b_i = 0$. 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:

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

Hypothesis 3 could fail due to a possible curiosity effect: 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. If such a curiosity effect is present in our data, then it should not matter whether causal feedback can be acquired after an adverse event in round 1 or in round 2. 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.*

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

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

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

The subsequent non-parametric and the multivariate analysis focuses on participants who chose action H (the 'high-stress' action) and experience a zero payoff event in round 1. These participants then either switched their choice to L in round 2, or they did not. We treat the decision to reduce stress on the system from H to L as our binary dependent variable. The participants in this 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 our baseline for the analysis of attributability. We include feedback of a natural cause, an anthropogenic cause, and joint causes as indicators in the analysis.

Participants' actions and the event engine resulted in 614 relevant experimental biographies. The focus of the analysis is, first, how attributability after round 1 affects behavior in round 2 (Result 1) and, second, how attribution to a natural cause affects behavior in round 2 differently from attribution to an anthropogenic cause (Result 2). Finally, we consider the demand for causal attribution (Result 3).

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.*

We first focus on participants in conditions FI ($1,031$ observations) and IM ($1,002$ observations) 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 observations) 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 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.

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). Control variables come from an exit questionnaire administered after round 2.

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

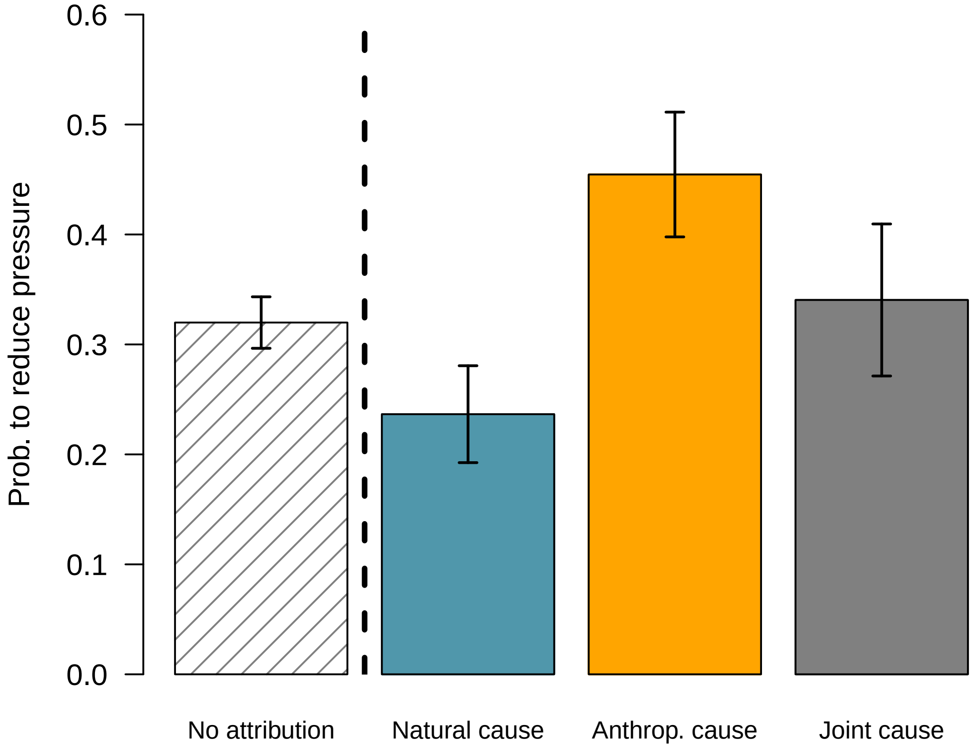


Figure 4: Awareness of the natural or anthropogenic cause of an adverse event has a behavioral impact. The bar plot shows the shares of participants reducing anthropogenic stress in round 2 after an adverse event in round 1. The share when the cause is unknown is about one third (32%, hatched, left). 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).

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.

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

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

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.

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 accout of the anthropogenic, rather than a natural, cause.*

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 $\mathrm{E}[Y|H{=}\mathrm{HHg} \wedge F{=}1]{=}0.45$ and the latter by $\mathrm{E}[Y|H{=}\mathrm{HLr} \wedge F{=}1]{=}0.24$. The difference between the two propensities is highly significant ($p{=}0.004$, two-sided test of proportions).

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

To determine whether there is a 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 $\mathrm{E}[b|H{\in}\{\mathrm{HHg,HLr,HHr}\} \wedge T{=}IM]{=}\$0.07$. For round 2 feedback, we find an average bid of $\mathrm{E}[b|H{\in}\{\mathrm{HHg,HLr,HHr}\} \wedge T{=}FI]{=}\$0.09$. 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: $p < 0.0001$; round 2: $p < 0.0001$, two-sided t-tests). Hypothesis 3 can therefore be rejected.

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 measurement of the value of causal attribution to participants arises out of treatment conditions IM and FI. The value of attributability is measured in two ways: A binary expression of interest in receiving causal feedback; and the value of the bid for such feedback of those who expressed an interest.

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. In terms of 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.

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 69% 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 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 (the baseline category). 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 how much participants were willing to pay for attribution (conditional on wanting to know what caused the outcome). Looking at the intensive margin, we do not find any differences between player's experienced history in the FI treatment, but we do see some differences in the IM treatment. Interestingly, players that exerted high stress on the system are less willing to pay for learning what caused the outcome than those in the baseline category.

Overall, players in the IM treatment are more likely to want to know and willing to pay slightly more to learn what caused the outcome. Finally, player specific co-variates also 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.

5 Conclusion

This paper set out to test the claim that backward-looking causal attribution of adverse events to natural or anthropogenic causes can be behaviorally impactful. This claim supports the notion that extreme event attribution in the climate context could influence the future mitigation behavior of the general public. It is a claim that does not easily align with narrowly consequential models of decision-making among members of the public and that so far has little empirical support. Given the challenges of a research design that could test the claim on the basis of observational data, this paper adopts an experimental approach that allows the researcher to test the claim in an abstract and decontextualized setting that also offers considerable experimenter control over causes, behavioral relevance, and feedback.

On the basis of a novel design for generating adverse events in bicausal systems and of an online experiment with over 3,000 participants, the experimental al evidence demonstrates three important behavioral consequences of extreme event attribution. First, there is evidence for an Attributability Effect: Compared to a situation in which an adverse event cannot be attributed, attributing a past adverse event to a natural or anthropogenic cause induces behavioral differences. For adverse events attributed to an anthropogenic cause, subjects switch away from the “high-stress” action considerably more frequently than when the event is not attributed. For adverse events attributed to a natural cause, subjects switch away from the “high-stress” action considerably less frequently than when the event is not attributed, even though the effect is less strong for the natural cause. Attributing the adverse event to joint causation has no impact. Causal attribution appears to

Table 3: OLS 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^{***}	-0.036^{**}	-0.165^{***}	-0.016
	(0.043)	(0.018)	(0.047)	(0.020)
Adverse	-0.009	0.026^{**}	0.009	0.004
	(0.032)	(0.012)	(0.040)	(0.015)
Action H × Adverse	0.120^{**}	-0.051^{**}	0.167^{***}	0.027
	(0.058)	(0.023)	(0.061)	(0.025)
Age	0.001	0.0005	0.001	0.001^{*}
	(0.001)	(0.0004)	(0.001)	(0.0005)
Female	0.041	-0.011	0.074^{**}	0.025^{**}
	(0.027)	(0.010)	(0.029)	(0.011)
Risk tolerance	0.017^{***}	0.023^{***}	0.026^{***}	0.026^{***}
	(0.005)	(0.002)	(0.005)	(0.002)
Education	0.018	0.014^{***}	0.025^{**}	0.022^{***}
	(0.011)	(0.004)	(0.012)	(0.005)
Comprehension	0.002	0.031^{***}	0.015	0.035^{***}
	(0.013)	(0.005)	(0.014)	(0.005)
Speed	-0.0004	-0.001^{***}	0.0003	-0.0005^{***}
	(0.0003)	(0.0001)	(0.0003)	(0.0001)
Constant	0.609^{***}	-0.116^{***}	0.328^{***}	-0.218^{***}
	(0.082)	(0.032)	(0.083)	(0.033)
Observations	1,002	773	1,031	709
Adjusted R^2	0.022	0.367	0.042	0.322

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.
$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

make a behavioral difference that also affects future levels of anthropogenic stress in the system, decreasing it after an anthropogenic cause and increasing it after a natural cause. This evidence substantiates the presence of a critical link between the practice of causal event attribution and behavior. The result on joint causation shows, however, that attribution to an anthropogenic cause needs to be sufficiently unambiguous for it to lead to the desired reduction in stress on the system.

Second, there is evidence for a Causal Attribution Effect: Which cause an adverse event is attributed to matters for future behavior. Behavior following a causal attribution to an anthropogenic cause is characterized by significantly less frequent choice of a "high-stress" action compares to behavior following a causal attribution to a natural cause. The Causal Attribution Effect substantiates an important building block of the claim of the behavioral relevance of EEA: Once causal attribution is conducted and leads to unambiguous results, the cause is important for the direction of behavioral change.

Third, event attribution is in demand: Despite being backward-looking rather than decision-supporting, the majority of participants request causal feedback and about one third are willing to pay for it. Such demand provides an economic justification for resources being spent on EEA. It also makes it plausible that a significant share of affected people will incur the search cost involved in looking for causal feedback, if available. The conjecture that attribution science can be behaviorally impactful, socially valuable, and in demand therefore rests on promising experimental foundations.

To date, EEA is mostly conducted in the context of climate change. Our experiment deliberately abstracts from a concrete setting in order to avoid undesirable framing effects. This has important implications for relating the experimental results to the specifics of Attribution Science in the climate context. One is that the experimental environment differs in two important respects from climate change: One, the cause of the adverse event was disclosed with certainty. In the climate change context, such certainty in event attribution is a limit case. At the same time, it is a limit case that has already occurred and will occur with increasing frequency as climate change progress.[13] The other implication is that the experiment considers a limit case of two players 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. This could mean that the incentives for behavioral change are struc-

[13]See footnote 2 for the relevant studies.

turally weaker, if not negligible, in the climate context.[14] 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. 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.

We find that the majority of participants demand causal feedback. Such feedback is demanded even after round 2, despite not being actionable. Some part of the demand for causal attributability must therefore originate for reasons unrelated to consequentialist models of decision-making. At the same time, interest in causal feedback is higher after round 1. Another part of the demand for attributability could therefore originate from participants anticipating that causal feedback is relevant for future action, even if the feedback is not decision-supporting.

Additional research is needed to understand the behavioral impacts of ambiguous attribution: How clear does attribution to anthropogenic causes need to be trigger a behavioral response? This could be answered by extending the present experimental paradigm. Future research will also have to address the question of impact persistence with respect to behavioral change. The underlying mechanism driving the behavioral impact of and demand for attributability merits further investigation. Cognition-based theories of event saliency and recency compete with theories of emotional affect and guilt for causing adverse events. 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).

[14] 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.

References

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

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

Arias, P., Bellouin, N., Coppola, E., Jones, R., Krinner, G., Marotzke, J., Naik, V., Palmer, M., Plattner, G.-K., Rogelj, J., et al. (2021). Climate change 2021: The physical science basis. contribution of working group14 i to the sixth assessment report of the intergovernmental panel on climate change; technical summary.

Atalay, E., Hortacsu, A., Roberts, J., and Syverson, C. (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13):5199–5202.

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

Bulte, E., Gerking, S., List, J. A., and De Zeeuw, A. (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.

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

Demski, C., Capstick, S., Pidgeon, N., Sposato, R. G., and Spence, A. (2017). Experience of extreme weather affects climate change mitigation and adaptation responses. *Climatic Change*, 140:149–164.

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

Diffenbaugh, N. S., Singh, D., Mankin, J. S., Horton, D. E., Swain, D. L., Touma, D., Charland, A., Liu, Y., Haugen, M., Tsiang, M., et al. (2017). Quantifying the influence of global warming on unprecedented extreme climate events. *Proceedings of the National Academy of Sciences*, 114(19):4881–4886.

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

Erisman, J. W., Brasseur, G., Ciais, P., van Eekeren, N., and Theis, T. L. (2015). Global change: Put people at the centre of global risk management. *Nature News*, 519(7542):151.

Ettinger, J., Walton, P., Painter, J., Osaka, S., and Otto, F. E. (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, O. and Shenhav, A. (2019). Resolving uncertainty in a social world. *Nature human behaviour*, 3(5):426–435.

Feldmann, A., Gasser, O., Lichtblau, F., Pujol, E., Poese, I., Dietzel, C., Wagner, D., Wichtlhuber, M., Tapiador, J., Vallina-Rodriguez, N., et al. (2020). The lockdown effect: Implications of the covid-19 pandemic on internet traffic. In *Proceedings of the ACM Internet Measurement Conference*, pages 1–18.

Grantham, B. A., Chan, F., Nielsen, K. J., Fox, D. S., Barth, J. A., Huyer, A., Lubchenco, J., and Menge, B. A. (2004). Upwelling-driven nearshore hypoxia signals ecosystem and oceanographic changes in the northeast pacific. *Nature*, 429(6993):749–754.

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

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

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

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

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

Hoogendoorn, G., Sütterlin, B., and Siegrist, M. (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, P. D., Marlon, J. R., Mildenberger, M., and Shield, B. S. (2019). How will climate change shape climate opinion? *Environmental Research Letters*, 14(11):113001.

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

Imada, Y., Shiogama, H., Takahashi, C., Watanabe, M., Mori, M., Kamae, Y., and Maeda, S. (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.

James, R. A., Jones, R. G., Boyd, E., Young, H. R., Otto, F. E. L., Huggel, C., and Fuglestvedt, J. S. (2019). *Attribution: How Is It Relevant for Loss and Damage Policy and Practice?*, pages 113–154. Springer International Publishing, Cham.

Jézéquel, A., Dépoues, V., Guillemot, H., Rajaud, A., Trolliet, M., Vrac, M., Vanderlinden, J.-P., and Yiou, P. (2020). Singular extreme events and their attribution to climate change: A climate service–centered analysis. *Weather, Climate, and Society*, 12(1):89–101.

Kew, S., Pinto, I., Alves, L., Santos, D., Libonati, R., Liberato, M., Philip, S., Zachariah, M., Barnes, C., Vahlberg, M., et al. (2023). Strong influence of climate change in uncharacteristic early spring heat in south america. *Grantham Research Institute, Imperial College*.

Kim, J.-Y., Lee, H., Min, I., et al. (2013). The economics of curiosity. *Korean Economic Review*, 29:23–50.

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

Labovitz, C., Ahuja, A., and Jahanian, F. (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)*, pages 278–285. IEEE.

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

Leichenko, R. and O'Brien, K. (2008). *Environmental change and globalization: double exposures*. Oxford University Press.

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

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

Lohmann, P. M. and Kontoleon, A. (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.

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

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

Mills, E. (2005). Insurance in a climate of change. *Science*, 309(5737):1040–1044.

Ogunbode, C. A., Demski, C., Capstick, S. B., and Sposato, R. G. (2019). Attribution matters: Revisiting the link between extreme weather experience and climate change mitigation responses. *Global Environmental Change*, 54:31–39.

Oldenborgh, G. J., Wiel, K., Kew, S., Philip, S., Otto, F., Vautard, R., King, A., Lott, F., Arrighi, J., Singh, R., et al. (2021). Pathways and pitfalls in extreme event attribution. *Climatic Change*, 166(1):1–27.

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

Otto, F. E. (2017). Attribution of weather and climate events. *Annual Review of Environment and Resources*, 42:627–646.

Pfrommer, T., Goeschl, T., Proelss, A., Carrier, M., Lenhard, J., Martin, H., Niemeier, U., and Schmidt, H. (2019). Establishing causation in climate litigation: admissibility and reliability. *Climatic Change*, 152(1):67–84.

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

Shepherd, T. G. (2019). Storyline approach to the construction of regional climate change information. *Proceedings of the Royal Society A*, 475(2225):20190013.

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

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

Singh, R. K., Arrighi, J., Coughlan de Perez, E., Warrick, O., Suarez, P., Koelle, B., Jjemba, E., van Aalst, M. K., Roberts, D. C., Pörtner, H. O., 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.

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

Stuart-Smith, R. F., Otto, F. E., Saad, A. I., Lisi, G., Minnerop, P., Lauta, K. C., van Zwieten, K., and Wetzer, T. (2021). Filling the evidentiary gap in climate litigation. *Nature Climate Change*, 11(8):651–655.

Trenberth, K. E., Fasullo, J. T., and Shepherd, T. G. (2015). Attribution of climate extreme events. *Nature Climate Change*, 5(8):725–730.

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

Walsh, J. E., Thoman, R. L., Bhatt, U. S., Bieniek, P. A., Brettschneider, B., Brubaker, M., Danielson, S., Lader, R., Fetterer, F., Holderied, K., et al. (2018). The high latitude marine heat wave of 2016 and its impacts on alaska. *Bull. Am. Meteorol. Soc*, 99(1):S39–S43.

Appendix – not for publication

Instructions, data, and analysis files are freely and openly available online at `https://osf.io/y58mg/?view_only=de73fd8da6354fb39b7c77d736bc1304`. For convenience, we include the instructions here.

Instructions

Legend: [page references], [treatment differences], [user interface elements], [check elements], [conditionals]

[mTurk title]
Research in Decision Making

[mTurk description]
Participate in a game and a short survey. Please note that the task is to be completed within 10-15 minutes as you are matched with a co-player.

[mTurk preview, on separate screen]
Please read this carefully before clicking 'accept'.
This HIT is an academic research study on decision making.

Research goal:
In this study, we are interested in decision making under uncertainty. You will be matched with co-players and you will be asked to take a decision.

Duration and reward:
The entire study will take about 10 minutes. Your payment consists of a fixed reward of $0.80 via Amazon Mechanical Turk for successful completion and a bonus that depends on your and the co-players' decisions as well as on chance. You are also asked to complete a short and anonymous survey.

Please note that the task should be completed without delay.

Confidentiality:
All data we collect is treated confidentially and will only be used for our research purpose. Your name will not be linked to the results in any way.

Requirements:
To participate, you need to be located in the United States of America. You may not have participated in this study before. **After reading the instructions, there will be a short comprehension quiz. For successful completion of the HIT you need to pass this test.** There are no other formal requirements for participation.

Voluntary participation:
Participation in this study is voluntary. If you do not want to take part in the study, please do not accept the HIT. If you want to participate, please be sure you can commit to completing the HIT before accepting it - if you discontinue participation, you will not receive any bonus.

Contact:
If you have any questions regarding this study, please contact {experimenter_name}: {experimenter_email}.
[accept HIT]

1

[Introduction]

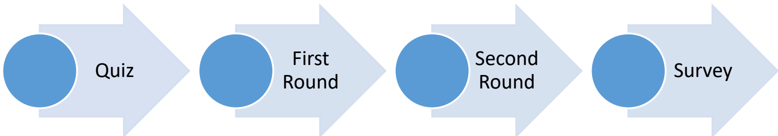
Introduction
Thank you for participating.

[BL treatment]
If you read these rules carefully and choose wisely, you can earn a bonus of up to US$ 3.00 by participating in a game that involves other participants.

[IM and FI treatments]
If you read these rules carefully and choose wisely, you start with a bonus of US$0.50 and can earn an additional bonus of up to US$ 3.00 by participating in a game that involves other participants.

Completing this task will take about 10 minutes and it is important that you pay close attention during this time so that you do not spoil the task for you and the other participants.

After we have explained the game, there comes the task. The task has four stages.



First, you take a small **quiz** about the game.

Then you are matched with another participant and the **first round** of the game is played.

Thereafter, you are matched with a different participant and the **second round** of the game is played.

Finally, you are asked to fill out a short **survey**.

Note that only one of the two rounds will be selected for payment at random. As it is unknown which of the two rounds counts, it is important to pay equally close attention to both.

Please solve the captcha below to continue with the study.
[captcha]

[next]

2

[Instructions 1]

Rules

You will be a player in a game. Here are the rules. Read them carefully, they will be part of the quiz.

There is the player, a co-player, and a virtual urn that contains 100 balls, some red, some green.

For each player, three factors together determine the payoff:

(1) The draw of a ball from the urn,
(2) The player's own action, and
(3) their co-player's action.

The player and the co-player take one of two actions, either **A** or **B**. The urn from which the ball is drawn contains 60 **green** balls and 40 **red** balls.

If a **red** ball is drawn, the actions of the player and the co-player do not matter for the outcome. The player receives nothing (US$ 0). So does the co-player (US$ 0).

If a **green** ball is drawn, then the actions of the player and the co-player matter for the outcome. There are four possibilities:

- A **green** ball is drawn and the player's **action is A** and the **co-player's action is A**: The player receives US$ 1. So does the co-player (US$ 1).
- A **green** ball is drawn and the player's **action is A** and the **co-player's action is B**: The player receives US$ 1. The co-player receives US$ 3.
- A **green** ball is drawn and the player's **action is B** and the **co-player's action is A**: The player receives US$ 3. The co-player receives US$ 1.
- A **green** ball is drawn and the player's **action is B** and the **co-player's action is B**: The player receives nothing (US$ 0). So does the co-player (US$ 0).

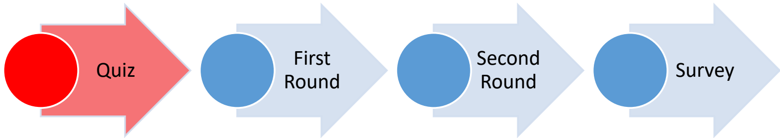
Typically, co-players choose action **A** in about 60 out of 100 cases and action **B** in about 40 out of 100 cases.

[next]

3

[comprehension 1]

Quiz (1/2)



Welcome to the quiz.

Please answer the following questions. **Make sure to select the correct answers. You will only be able to continue this study if you need at most three tries to correctly answer all seven questions of the quiz!**

Question 1:
Which of the following is correct? In the first round and the second round, my co-player is
- the same participant in both rounds.
- a different participant in each round.

[Option 2 is correct]

Question 2:
Which of the following is correct? On average, co-players
- choose A more often than B.
- choose B more often than A.
- choose A and B equally often.

[Option 1 is correct]

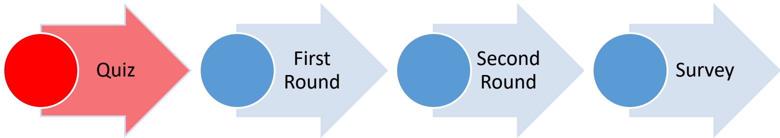
Question 3:
Remember that only one of the two rounds counts for your payment, with equal chance. What does this mean?
- The outcome of round 1 is less important than the outcome of round 2.
- The outcomes of both rounds are equally important.
- The outcome of round 2 is less important than the outcome of round 1.

[option 2 is correct]

[check answers] [if correct / corrected: next]

[comprehension 2]

Quiz (2/2)



Make sure to select the correct answers. You will only be able to continue this study if you need at most three tries to correctly answer all seven questions of the quiz!

Question 4:
What is your payout if your action is A, your co-player's action is B, and the ball is red?
- US$ 0
- US$ 1
- US$ 3

[Option 1 is correct]

Question 5:
What is your payout if your action is A, your co-player's action is B, and the ball is green?
- US$ 0
- US$ 1
- US$ 3

[Option 2 is correct]

Question 6:
What is your payout if your action is B, your co-player's action is B, and the ball is green?
- US$ 0
- US$ 1
- US$ 3

[Option 1 is correct]

Question 7:
What is your payout if your action is B, your co-player's action is A, and the ball is green?
- US$ 0
- US$ 1
- US$ 3

[Option 3 is correct]

[check answers] [if correct / corrected: next]
[show payoff reminder]

[Quiz dropout, only shown if more than 3]

Feedback

Unfortunately, you did not pass the quiz. You needed {number of tries} to answer all seven questions correctly. Passing the quiz within at most three tries is a requirement for this study as explained in the HIT description.

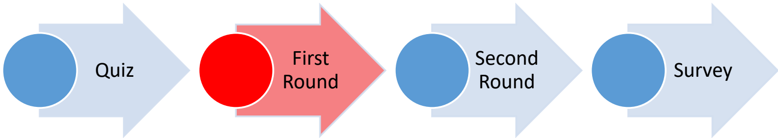
To avoid the HIT being rejected, we suggest to return the HIT on Mechanical Turk.

Thank you for your participation.

If you have any questions regarding this study, please write an email to {experimenter_name}: {experimenter_email}.

[decision 1]

First Round



Welcome to the first round of the game.

You are now a player in this game, where you can earn money if this round is selected for payout.

You are matched with another participant, your co-player for this round. Your co-player will choose an action, A or B, and a ball, red or green, will be drawn.

Which action do you choose?

- A
- B

[instructions reminder box]

Remember: Your payoff is jointly determined by

- whether a red or a green ball is drawn from the urn, which contains 60 green and 40 red balls,
- whether your action is A or B, and
- whether your co-player's action is A or B.

Typically, co-players choose action A in about 60 out of 100 cases and B in about 40 out of 100 cases.

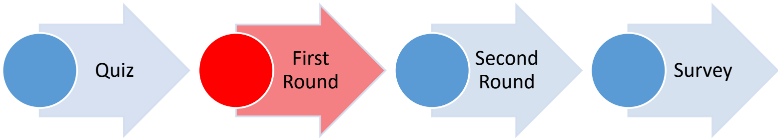
If a green ball is drawn, and

- your action is A, and your co-player chooses A: You receive US$ 1. Your co-player receives US$ 1.
- your action is A, and your co-player chooses B: You receive US$ 1. Your co-player receives US$ 3.
- your action is B, and your co-player chooses A: You receive US$ 3. Your co-player receives US$ 1.
- your action is B, and your co-player chooses B: You receive nothing (US$ 0). Your co-player receives nothing (US$ 0).

If a red ball is drawn, you and your co-player both receive nothing (US$ 0).

[next]

[belief ball color 1, on separate screen]
First round: Questions

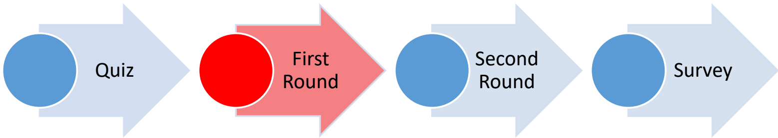


What is your gut feeling - is the color of the ball that was just drawn:

green or **red**?

[slider, 0-100]
[next]

[belief other's decision 1, on separate screen]
First round: Questions



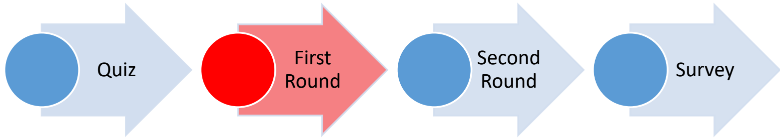
What is your gut feeling - did your co-player just choose:

A or **B**?

[slider, 0-100]
[next]

[Waiting room]

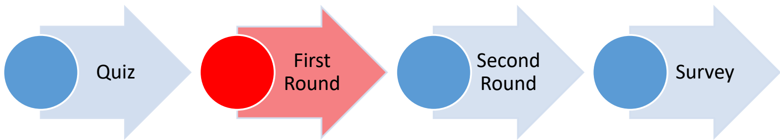
First round



Please wait while we match your action to your co-player's action.

[results 1]

First round: Results



Your action was **{A/B}**

[BL: Event + causal information]
Your co-player's action was **{A/B}**.
The ball drawn from the urn was **{green/red}**.
As a result, you earn **{payoff}** if this round is selected for payment.

[FI and IM: No causal information]
Based on

- the urn draw,
- your own action, and
- your co-player's action,

you earn **{payoff}** if this round is selected for payment.

[IM treatment]
Before you proceed to the second round, you can find out which action your co-player took and the color of the ball that was drawn.

Would you like to know the color of the ball drawn and the action your co-player took?
[two Buttons: Yes / No]

[if YES]:
Please set the slider below between $0.00 and $0.50 to state how much you are willing to pay at most for finding out. The computer will randomly determine the price of the information. If your willingness to pay is lower than the price, you will not buy the information. If it is higher, you buy the information. In this case its price will be deducted from your bonus.

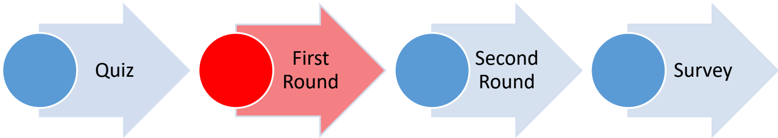
Your interests are best served by setting the slider to your true valuation of the information.

How many cents at most are you willing to pay to find out which action your co-player took and the color of the ball that was drawn?

[slider, 0-50]
[next]

[Causes after BDM] only shown in [IM treatment]

First Round



[If info not bought]
The amount you were willing to pay at most was less than the randomly determined price of {randomly drawn price}. You do not find out which action your co-player took and the color of the ball that was drawn. No money will be deducted from your bonus.

As a result, you earn **{payoff}** if this round is selected for payment.

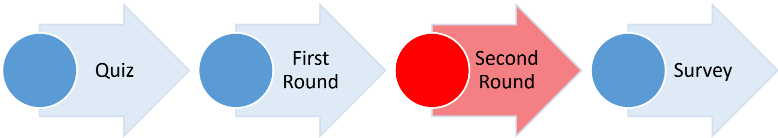
[If info bought]
The amount you were willing to pay at most was more than the randomly determined price of {randomly drawn price}. You find out which action your co-player took and the color of the ball that was drawn. The price of {randomly drawn price} will be deducted from your bonus.

Your action was **{A/B}**
Your co-player's action was **{A/B}**.
The ball drawn from the urn was **{green/red}**.
As a result, you earn **{payoff}** if this round is selected for payment.

[next]

[decision 2; new screen]

Second Round



Welcome to the second round of the game, where you can earn money if this round is selected for payout.

You are matched with a different participant now, your co-player for this round. Your co-player will choose an action, A or B, and a ball, red or green, will be drawn.

Which action do you choose?
- A
- B

[instructions reminder box]

Remember: Your payoff is jointly determined by
- whether a red or a green ball is drawn from the urn, which contains 60 green and 40 red balls,
- whether your action is A or B, and
- whether your co-player's action is A or B.

Typically, co-players choose action A in about 60 out of 100 cases and B in about 40 out of 100 cases.

If a green ball is drawn, and
- your action is A, and your co-player chooses A: You receive US$ 1. Your co-player receives US$ 1.
- your action is A, and your co-player chooses B: You receive US$ 1. Your co-player receives US$ 3.
- your action is B, and your co-player chooses A: You receive US$ 3. Your co-player receives US$ 1.
- your action is B, and your co-player chooses B: You receive nothing (US$ 0). Your co-player receives nothing (US$ 0).

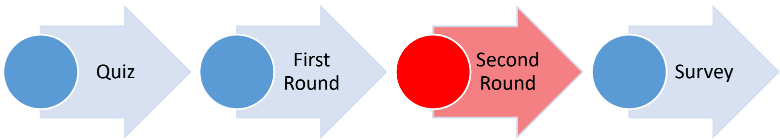
If a red ball is drawn, you and your co-player both receive nothing (US$ 0).

[next]

13

[belief ball color 2, new screen]

Second round: Questions



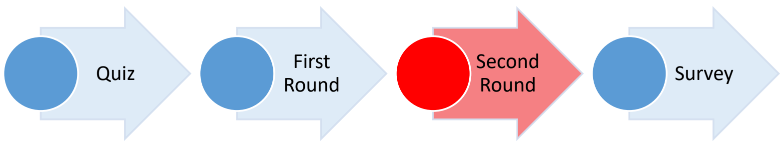
What is your gut feeling - is the color of that ball that was just drawn:

green or **red**?

[slider, 0-100]
[next]

[belief other's choice 2, on separate screen]

Second round: Questions



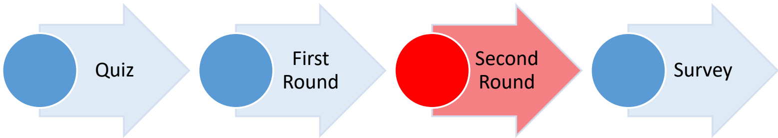
What is your gut feeling - did your co-player just choose:

A or **B**?

[slider, 0-100]
[next]

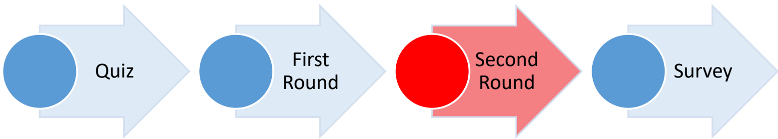
[Waiting room]

Second round



Please wait while we match your action to your co-player's action.

[results 2]



Results Round 2

Your action was **{A/B}**

[BL + MI treatment]
Your co-player's action was **{A/B}**.
The ball drawn from the urn was **{green/red}**.
As a result, you earn **{payoff}** if this round is selected for payment.

[FI: No causal information]
Based on

- the urn draw,
- your own action, and
- your co-player's action,

you earn **{payoff}** if this round is selected for payment.

[FI treatment]
Before you proceed, you can find out which action your co-player took and the color of the ball that was drawn.

Would you like to know the color of the ball drawn and the action your co-player took?
[two Buttons: Yes / No]

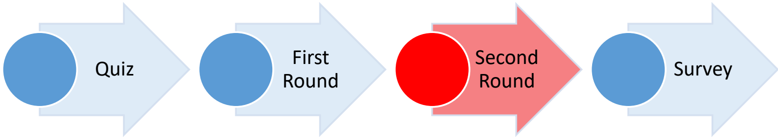
[if YES]:
Please set the slider below between $0.00 and $0.50 to state how much you are willing to pay at most for finding out. The computer will randomly determine the price of the information. If your willingness to pay is lower than the price, you will not buy the information. If it is higher, you buy the information. In this case its price will be deducted from your bonus.

Your interests are best served by setting the slider to your true valuation of the information.

How many cents at most are you willing to pay to find out which action your co-player took and the color of the ball that was drawn?

[slider, 0-50]
[next]

[Causes after BDM] only shown in [FI treatment]



Second Round
[If info not bought]
The amount you were willing to pay at most was less than the randomly determined price of {randomly drawn price}. You do not find out which action your co-player took and the color of the ball that was drawn. No money will be deducted from your bonus.

As a result, you earn **{payoff}** if this round is selected for payment.

[If info bought]
The amount you were willing to pay at most was more than the randomly determined price of {randomly drawn price}. You find out which action your co-player took and the color of the ball that was drawn. The price of {randomly drawn price} will be deducted from your bonus.

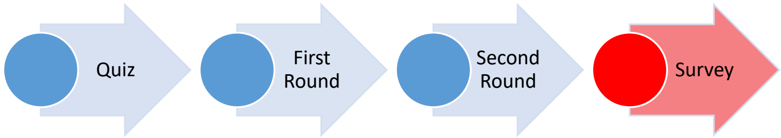
Your action was **{A/B}**
Your co-player's action was **{A/B}**.
The ball drawn from the urn was **{green/red}**.
As a result, you earn **{payoff}** if this round is selected for payment.

[next]

18

[demographics]

Survey



Welcome to the survey. We ask you to answer a few questions before you complete the experiment.

What is your age?
[number input]

What is your gender?
- Male
- Female
- Other
- I prefer not to tell

What is the highest level of school you have completed or the highest degree you have received?
- Less than high school degree
- High School degree or equivalent (e.g. GED)
- some college, but no degree
- Associate degree
- Bachelor degree
- Graduate degree

If you had at least some college education, please tell us your major:
[free text input]

How do you see yourself: Are you in general a person who takes risk (10) or do you try to avoid risks (0)? Please self-grade your choice (0-10).
[slider, 0, 10]

Do you have any comments about this study?
[free text input]

[next]

19

[last page]

Completed

Thank you for your participation, your answers were transmitted.

Round **{1/2}** was randomly selected for payout, where you earned **{payoff}**.

[If info bought]
You paid {randomly drawn price} for finding out the action of your co-player and the urn draw.

In total, you receive: **{payoff plus fixed reward + $0.50 -** {randomly drawn price}**}**. [only in IM / FI]

If you have any questions regarding this study, please write an email to {experimenter_name}: {experimenter_email}.

[finish study]