Decision Making With Risk-Based Weather Warnings[1]

Di Mu[2]
University of Exeter Business School, Exeter, UK

Todd R. Kaplan
University of Exeter Business School, Exeter, UK
Department of Economics, University of Haifa, Haifa, Israel

Rutger Dankers
UK Met Office, Exeter, UK

2 *Corresponding author address*: Di Mu, University of Exeter Business School, Streatham Court, Rennes Drive, Exeter, EX4 4PU,UK

E-mail: dm347@exeter.ac.uk

Abstract

We study decisions under different weather warning systems that vary in format and/or information conveyed using a laboratory experiment. Participants have to decide between a safe but costly option (spending to protect from a storm) and a risky option (of not spending for protection). We ran three treatments based upon the severe weather warning system for the UK that the Met Office has been using since 2011 - a risk matrix to communicate the impact and likelihood of an event. In Treatment 1, participants received a colored table with a check in the box of the matrix that showed the likelihood and impact level of the warning. In Treatment 2, participants had the colored table and the color of the warning communicated but without a check in the exact box. In Treatment 3, participants only had the color of the warning communicated without seeing the associated table. Overall our work shows that while increasing the information with content of warnings is usually beneficial and increases the trust in the warning system, it must be done with caution since better decisions (judged by higher profits) are not always made with an increase of information.

1. INTRODUCTION

High impact weather events significantly affect our lives and have done so throughout history. In 2005, hurricane Katrina in U.S caused 75 billion dollars worth of damage and over 1,200 deaths.[3] In the final report of IPET Draft (Committee on New Orleans Regional Hurricane Protection Projects, 2009), the pre-event risk assessment of Katrina did not have a clear presentation of the risks. This may have made decision-making more difficult.

Early Warning Systems (EWS), combined with effective communication and emergency preparedness at national and local levels, have a huge potential in limiting the human and economic losses from natural disasters such as hurricane Katrina. Around the world, EWS have been developed for a wide range of hazards, most notably for extreme weather, floods and tsunamis, but also for other hazards like snow avalanches, wildfires, landslides, earthquakes and volcanic eruptions, and more complex hazards such as drought, food security and desertification (UNISDR, 2006). For hydro meteorological hazards, in particular, scientific and technological advances in recent decades have resulted in a marked improvement in the quality, timeliness, and lead time of hazard warnings. The adoption of ensemble prediction methods has resulted in a shift towards quantitative probabilistic forecasts, providing a range of possible outcomes and indicating the probability or chance of a particular weather or flood event happening.

Both likelihood of an event and potential impact play an important role in weather forecast systems (Doksaeter Sivle and Kolstø, 2016), but also provide additional challenges to communication and decision-making (Demeritt et al., 2013; Dale et al., 2014). Deciding whether to evacuate a large number of people or to issue a warning only to the most vulnerable needs to be risk-based, meaning the decision should be based not just on the likelihood of an event happening, but also on its consequences. A warning system can be enhanced by providing both these types of information (Casteel, 2016).

However, in many EWS currently in operation the focus is on forecasting the hazard, and impacts are not commonly assessed. Already in 2006 the UNISDR noted a lack of knowledge on risks and vulnerabilities, and the limited engagement of relevant social sciences, as an area of weakness for EWS (UNISDR, 2006; Basher, 2006). This also includes the integration of risk information into the hazard warning itself. A better understanding of the risks and potential impact could help improve decision making, which in turn could mitigate the damage. It could also result in warnings that are more relevant to users and easier to communicate to the public at risk.

Within the UK, the National Severe Weather Warning Service (NSWWS) adopted a risk-based approach in 2011. Before then, warnings were issued based on the chance of 'widespread disruption', which was loosely defined and required a subjective judgement on the expected disruption depending on the vulnerability within a county (Neal et al. 2014). The current NSWWS system improves on this by using a risk matrix, combining the likelihood of a particular severe weather event on the vertical axis, and its potential impact on the horizontal axis. Meteorologists use their expert knowledge of the area forecast to be affected during the weather event to assess the level of impact expected and therefore where to issue the 'tick' on the horizontal axis. For the

---

[3]   URL|www.nhc.noaa.gov/outreach/history/#katrina

vertical axis, output from various models are analyzed to assess the probability of the event occurring. These assessments are combined to determine the color of the warning that is issued.

The specific level of risk – the tick in the matrix – is predominantly used in communications with emergency managers, but is also available to the public. The headline message to the public, however, is based around the overall warning color.

Our paper's main objective is to evaluate this approach to early warnings using students making decisions in an experimental laboratory.

Understanding how users make decisions with uncertain information is of paramount importance and past research has focused on understanding and improving the decision making process with weather forecasts and warnings. Our experimental methodology has been deployed previously.   Joslyn (2007) found that when probability information is provided, participants more often make economically rational decisions. Roulston and Kaplan (2009) reported that by giving participants graphical information (uncertainty information of the forecast) they were able to make better decision than participants who were not provide with uncertainty information. Marimo et al. (2015) found that participants make decisions faster when uncertainty information is provided in a graph compared to a table. Abraham (2015) also showed that uncertainty information could affect participants' decisions. Previous research has also shown that providing more information will encourage people to make safer decisions (Ash et al. 2014, Wickens et al. 2000). These results demonstrated that information can not only affect participants' decisions but also their risk preferences. There is also a desire among users to receive more information in hope that it will lead them to make better decisions. For instance, when uncertainty information is not part of the initial forecast, 85% of the participants choose to have additional information to help them make decisions (Nadav-Greenberg and Joslyn 2009). However, Jacoby, Speller and Kohn (1974) found that when people had more information, actually, they were making worse decisions even though they more likely to feel satisfied and less confused. Sivle & Kolstø (2016) found that when people need to make a quick decision, the complexity of the warning must be reduced.   In extreme weather events, people need to make decisions in short notice, so the right amount of information will help the public as well as emergency manages and responders to understand the warning and make better decisions.

Participants' decisions are not only related to the information conveyed. Winett and Kagel (1984) found out when messages had the same information, the presentation format of the information affected participants' decisions.

A key element in most warning systems, including the NSWWS, is the warning color. The color itself can significantly affect participants' behaviour; Ryan (1991) showed the warning color red was associated with the highest level of hazard followed by orange, yellow and green. In Braun and Silver's (1995) study when a red warning color resulted in higher compliance rate than green.

There are therefore two main interrelated questions that we wish to answer in the present paper:

(1) To what extent does the amount of information that is included in the warning affect decisions?

(2) Does including information on likelihood and impact level lead to better decisions?

We try to answer these two questions with a lab-based decision making experiment. The findings of this paper have potential for application to communication of risk in other fields, as risk matrices are used not just in weather warnings, but more widely in project management and corporate planning, including by organisations such as the US Federal Highway Administration[4] and the US Federal Aviation Association[5].

We also try to see if there are behaviour biases in decision-making. Tversky and Kahneman, (1973) indicated the decisions participants made are based on the results they have in the earlier periods, and people may develop a bias. For instance, a person who has suffered serious damage from a previous storm will have a biased perception of the likelihood and impact level. We consider whether or not there is a "Cry Wolf Effect." Jared LeClerc and Susan Joslyn (2015) show that there is an effect in that subjects tend to trust the warning system more and there is a significant improvement in decision quality as the false alarm level decreases.

2. Experimental design/method

In total, 341 University of Exeter undergraduate students from various disciplines were recruited to participate in the experimental sessions. There were 145 male and 196 female students; 149 year one students, 105 year two students and 87 year three students. We used the recruitment software ORSEE (Greiner, 2015) and all sessions used software programmed in z-Tree (Fischbacher, 2007). The sessions were computer based in sizes of 30 and took place in the Finance and Economics Experimental Laboratory (FEELE). Each participant's payoff only depended upon his/her decisions and not that of other participants. The experiment consisted of 60 rounds of questions plus three test questions at the start and a questionnaire afterwards.

The use of students as participants in this type of lab experiment is widely accepted as standard (Harrison and List, 2004). As the variances in age, education and income are small, internal validity can be assured (any difference of results among treatments are caused by the difference of the treatments themselves rather than difference between the subjects allocated to each treatment). Non-students will have a larger variance of background knowledge and experience which would affect the internal validity. To assure the external validity (i.e., the results apply to decisions elsewhere), an important element is experimental realism. We do this by using a real payoff in cash with the decisions of the participants affecting their payoffs. Furthermore, a warning of severe weather is also something that may face in their real life and the experiment is framed as such. Thus, by doing so using students in an experimental study will not reduce experimental realism and not affect external validity (Druckman and Kam, 2009).

In each round of the experiment, participants were asked a question framed as deciding whether or not to move a product into a warehouse after they receiving a warning of coming a storm. Moving the product involves a fixed cost but no risk from the storm. Not moving has no fixed cost but risks damage from the storm. Participants were divided into three treatments according to the information given to them.

---

[4] California Department of Transportation, Federal Highway Administration, California Division. (2007). Systems Engineering Guidebook for ITS Version 2.0.

[5] Federal Aviation Administration. (2007). Introduction to Safety Management Systems (SMA) for Airport Operators (Advisory Circular), February 28, 2007. Washington, DC: U.S. Department of Transportation.
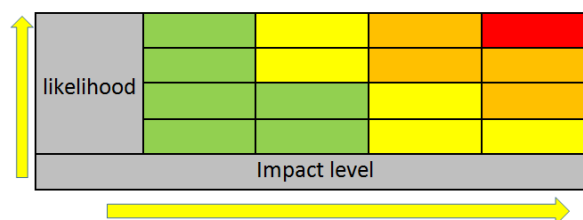
Treatment 1: Colored table with a check in the box that shows both what the likelihood and potential damage is.

Treatment 2: Colored table without a check in the exact box, but with the color of the warning communicated.
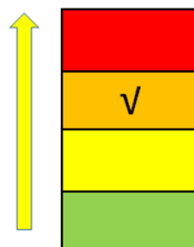
Treatment 3: The color of the warning communicated without the associated table.
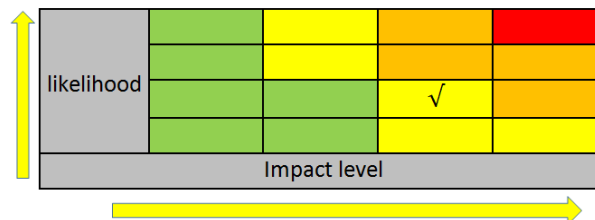
(Treatment 1)

(Treatment 2)

(Treatment 3)

Figure 1: The form of the warnings presented in Treatments 1, 2 and 3. In the matrix, likelihoods were 20%, 40%, 60%, and 80% (this information is not available to participants) and impact level corresponded from 1 to 4.

The warning in Treatment 1 shows impact level and likelihood level information. The warning in Treatment 2 shows only information about the color of the warning, but participants are visually shown how this color can vary according to likelihood and impact. Participants in Treatment 3 receive information about the color of the warning only.

The reasoning behind the three Treatments is as follows:

T1: Impact/likelihood information (although no precise probabilities): this is similar to the information that is available from the current Met Office warning system.

T3: Color only: this is the top-level information that is mostly communicated to the public; also most "traditional" warning systems (also for other areas beyond weather) use a simple color level approach.

T2: Participants are visually reminded that the warning is a combination of impact and likelihood, but only the color is communicated. This is an intermediate step that allows us to better explain some of the differences between T1 and T3.

For instance when we compare T1 and T2, we can find out whether including the impact/likelihood information leads to better decisions. When we compare T2 with T3 we learn whether awareness of uncertainty in impact/likelihood leads to better decisions. When we compare T1 with T3 we can test how information on impact and likelihood affect the decisions of the participants. This risk-based warning system that is tested in Treatment 3 is based upon the warning system for severe weather in the UK. This system targets different audiences; for example, the main warning message to the public is based on the color only. The impact/likelihood information (i.e., the position to the matrix) is also available to the public (on the Met Office website) but is used more in the communication to emergency managers and the responder community.

The information to the participants was kept to a minimum too allow them to answer a large number of questions within reasonable time. Clearly, an experimental setting like ours is always going to be different from decision-making in the real world; however also in the real world the amount of information received by the public through TV, radio and newspapers may well be limited to headlines only. The Met Office provides more detail information on weather warnings, for example on its website, however this information is only viewed by a part of the public, although admittedly professional users may well obtain more detailed information. Sivle (2018) indicated that the general public may not have enough background study to understand the more detailed information in weather warnings. In our study, we used university students from diverse backgrounds as experiment participants, so arguably our results should represent more the general public rather than professional decision makers.

Endowment for this period	2000.0
The fixed cost of moving your products to a safe place is	500.0
The amount of damage caused by level 1 (very low impact) storm	300.0
The amount of damage caused by level 2 (low impact) storm	600.0
The amount of damage caused by level 3 (medium impact) storm	900.0
The amount of damage caused by level 4 (high impact) storm	1200.0
What is your decision?	Do not move
	Move

(Treatment 1)

Endowment for this period   2000.0
The fixed cost of moving your products to a safe place is   500.0
The amount of damage caused by level 1 (very low impact) storm   925.0
The amount of damage caused by level 2 (low impact) storm   975.0
The amount of damage caused by level 3 (medium impact) storm   1025.0
The amount of damage caused by level 4 (high impact) storm   1075.0
The color of the warning   Red
What is your decision?   [Do not move] [Move]

Endowment for this period   2000.0
The fixed cost of moving your products to a safe place is   500.0
The amount of damage caused by level 1 (very low impact) storm   300.0
The amount of damage caused by level 2 (low impact) storm   600.0
The amount of damage caused by level 3 (medium impact) storm   900.0
The amount of damage caused by level 4 (high impact) storm   1200.0
What is your decision?   [Do not move] [Move]

(Treatment 3)

Figure 2: Payoff information shown to participants in Treatments 1, 2 and 3. (The payoff function and probability of each payoff type is consist though all treatments.)

The format of the warning is the same for every participant in the same treatment, but the impact level, likelihood and whether or not the storm actually happened are randomly generated by the computer. The probability of receiving a green, yellow, amber or red warning is was set at 10%, 30%, 30%, 30%, respectively. The reason that the green warning only has a 10% chance of beign chosen is because the green warning is the default status and we are more interested in how people behave while receiving an actual warning, that is, yellow, amber, or red. Once a color was selected, there was an equal probability of each box within the matrix with that color being selected. Participants were not told the probability of each warning or the likelihood of each box being selected and were not informed about the precise likelihood of each storm level occurring corresponding to the columns in the risk matrix (which were 20%, 40%, 60%, and 80%), but were told the cost of the damage associated with each storm level. So they needed to build their own judgement of the warning system to make decisions with different payoff functions.

At the beginning of each period (each period has one question to answer); participants received 2000 tokens, and then had to choose whether or not to move the product to a warehouse after they received a warning of a storm. The payoff function for the participants at each period is below. The decision that participants made and their payoff in the previous period do not affect the initial tokens after period.

$$\begin{cases} 2000(initial\ endowment) - moving\ cost\ (if\ choose\ to\ move) \\ 2000(initial\ endowment) - damage\ of\ the\ storm\ (if\ choose\ not\ move\ and\ the\ storm\ occurred) \\ 2000(initial\ endowment)\ (if\ choose\ not\ move\ and\ the\ storm\ did\ not\ occur) \end{cases}$$

There were four different payoff functions: **low damage**, **high damage**, **likelihood sensitive** and **impact sensitive**, determining the amount of damage associated with each level of storm. The damage of every type and level are showed in the Table 1 below.

Type\level	Level 1	Level 2	Level 3	Level 4	Moving cost
Low damage	300	600	900	1200	500
High damage	600	1000	1400	1800	500
Likelihood sensitive	925	975	1025	1075	500
Impact sensitive	300	400	1600	1700	500

Table 1: Damage of each impact level in each payoff function and the cost of moving.

When making decisions, in the impact sensitive payoff function, the difference in damage between impact level 2 and level 3 is larger than in any other payoff function, so participants will be more sensitive to the impact level.

In reality, people will be heterogeneously affected by weather events, and the level of impact may be different for different users. In real life, the cost of a decision will almost never be as clearly defined as in our experiment, and the impact depends on the level of exposure to the hazard, as well as the vulnerability of the people, area or assets being affected. The vulnerability in turn may depend on a range of factors, including the capacity to cope with the effects of the hazard. Participants can also be affected by other factors as well, for example the maximum value of the damage, the real payoff function can be different for each person and much more complicated. However, the choices made are a trade-off between being able to test the differences between payoff structures, and not introducing too much variability between payoffs, and not too many different payoff structures. In our experiment, exposure is implicitly assumed to be there, while the varying levels of impact in the different payoff functions reflect different sensitivities of the people to the weather, which may result in different responses.

Each participant faced 15 periods of each type of payoffs, and the order was random to control for any learning effect.

Endowment for this period	2000.0
The level of the storm	level 3 (medium impacct)
The storm	Did not occur
Your decision was	Not move
Your profit from this round will be	2000(endowment)
Your payoff for this period	2000.0

Instructions were provided to participants on paper (see the Appendix). After reading the instructions, each participant needed to finish three test questions (also see the Appendix). If they answered a test question wrong, participants were asked to stop and wait for explanation before they move on to the main part of the experiment. Participants were given the warning (Figure 1) and the payoff information (Figure 2) at the same time. After choosing whether to move or not move the product they are responsible for, they saw the results in a screen similar to Figure 3: the initial endowment, the level of the storm whether or not the storm occurred, their decision at this period, and the profit after each question.

At the end of the experiment participants were paid £5 plus £1 for every 2000 tokens they made in 5 randomly chosen periods. The average participant profit per period was 1525 tokens. This was slightly below what they were expected to make (1527 tokens) given their decisions. This contrasts with the 1561 tokens they would make if they chose the highest expected payoff decision every period in Treatment 1. Cash reward is standard in the field of experimental economics. The benchmark is the incentivized version with researchers concerned that lack of monetary incentives (or insufficient incentives) will affect results.

3. Data analysis

We define the **highest expected payoff decision (HEPD)** as the choice with higher expected monetary payoff given full information about the potential impact and likelihood include the impact level, likelihood level and likelihood of each storm level occurring corresponding to the columns in the risk matrix. The reason that we want to measure this variable as higher fraction of HEPD can represent a better warning system (judged by higher profits)

We define the **conditional highest expected payoff decision (conditional HEPD)** as the choice with higher expected monetary payoff given information received by the participant and likelihood of each storm level occurring corresponding to the columns in the risk matrix. This also assumes the ability for the participants to learn the parameters of the experiment.   For Treatments 1 and 2, they must learn that each likelihood level corresponding likelihood probability indicated for each box, 20% to 80%. For Treatment 2, they must learn that each box with a color was selected with an equal chance. For Treatment 3, they must learn the overall likelihood and impact that a warning indicates. So making the conditional HEPDs could be challenging. This variable can used to measure how participants understand/follow the warning.

We were also interested in how participants behaved when they did not made decisions that were the HEPD. To do so, we define **risk direction** as follows. If participants choose the conditional HEPD, then the risk direction is assigned 0. If the conditional HEPD is 'move' and the participants choose 'not move', then risk direction is assigned 1. If the conditional HEPD is 'not move' and participants choose 'move', then risk direction is assigned-1. Risk direction represents a participants' risk-aversion level. High risk direction means that participants favour a more risky option (not move)

over the conditional HEPD, while low risk direction means participants more often choose a risk free option (move) over the conditional HEPD.

We also tested the response time as well, but found that there is no significant difference in response time between different treatment groups, and there is no significant relation between response time and proportion of highest expected payoff decisions. This latter result contrasts with work of Kahneman (1979) and Rubinstein (2007), who found in many cases those that responded quicker chose differently than those who responded slower.

In this section, we analyze the responses of the participants. In order to reduce noise in the analysis, we use the expected profit based upon a participant's decision rather than actual payments received. This eliminates the noise of induced from whether or not the storm occurred.

Result 1: *The participants that received information on both the likelihood and impact levels not only had higher expected profits but made decisions with higher expected payoffs given their information than participants receiving only the warning color.*

To analyze the difference between different treatments, we compare the expected profit per period. One-way ANOVA shows the difference between treatments is significant ($F_{(2, 20457)} = 61.04$, $p < 0.001$) in expected profit per period. Participants in Treatment 1 had significant (significant is defined at 5% level) higher expected payoffs than participants in Treatment 2 ($t = -10.52$, $p < 0.001$) and 3($t = -7.49$, 0.001). However, there is no significant difference between Treatment 2 and Treatment 3($t = 1.45$, $p = 0.146$).

	Expected profit per period	Standard deviation	Percent of conditional HEPDs	Standard deviation
Treatment 1	1541	147	82.9%	37.6
Treatment 2	1515	159	78.8%	40.8
Treatment 3	1520	160	80.1%	39.9

Table 2: Expected profit per period and percent of conditional HEPDs in Treatment 1, 2 and 3.

A similar relationship holds on the fraction of conditional HEPDs.　Participants in Treatment 1 are more likely to make the HEPDs than the participants in Treatment 2 ($t = -6.54$, $p < 0.001$) and Treatment 3($t = -3.76$, $p < 0.001$). The difference between Treatment 2 and 3 is not significant ($t = 1.80$, $p = 0.072$). When we look at the fraction of HEPDs, the differences between Treatment 1 and the other two treatments (Treatment 2 ($t = -11.28$, $p < 0.001$), Treatment 3 ($t = -8.25$, $p < 0.001$)) are significant.　The difference between treatments 2 and 3 is not significant ($t = 1.34$, $p = 0.181$).

Result 2: *Higher warning levels increased the percentage of both highest and conditional HEPDs, but the expected profit still decreased.*

It is important to find out how participants behaved under different situations. When presented with a yellow warning participants' expected profit per period was significantly higher ($t = 38.04$, $p < 0.001$) with amber warnings, while the fraction of conditional HEPD was significantly lower ($t = -33.69$, $p < 0.001$).　Participants had the highest expected profit when the warning color was green (1719)

and the lowest expected profit (1467) when the color was red. To a large extent, this was due to the fact that the expected damage when not moving is higher at higher warning levels.

Participants are more likely to make conditional HEPDs when the warning color is red (94.8%) than with other colors. Participants made the lowest fraction of conditional HEPD (62.8%) when the warning color was yellow. Also when looking at the HEPDs, participants were more likely to make HEPDs with red warnings (94.8%) than other colors. The lowest fraction of conditional HEPD (61.7%) was obtained when the warning color was yellow.

These results show that participants behaved differently based upon the level (color) of the warning they received. With green warnings, participants made more profit than in any other situation, not surprising given that the expected damage is lowest. With warnings ranging from yellow to red, the fraction of conditional HEPD increased from 62.8% to 94.8%. This shows that if expected damage increases, participants will behave more according to the HEPDs. On the other hand, participants made less profit as the expected damage increased.

	Expected profit per period	Percent of conditional HEPDs	Percent of HEPDs
Green warning	1719	79.6%	79.3%
Yellow warning	1566	62.8%	61.7%
Amber warning	1474	85.5%	80.0%
Red warning	1467	94.8%	94.8%

Table 3: Expected profit per period, percent of conditional HEPDs and percent of HEPDs in different warning level.

We now examine the differences between treatments for each color warning.

Differences between treatments for green warnings

For green warning, one way ANOVA shows the differences between treatments are significant (F (2, 2201) = 7.87, p<0.001) for the fraction of conditional HEPDs and for the fraction of HEPDs (F (2, 2201) = 9.27, p<0.001). Participants in Treatment 1 made significantly more conditional HEPDs than participants in Treatment 2 (t=2.43, p=0.015) and Treatment 3 (t=3.59, p<0.001). Similar to the overall case there are no significant differences between Treatment 2 and 3 in fraction of conditional HEPDs. Participants in Treatment 1 also made significantly more HEPDs than participants in Treatment 2 (t=2.73, p<0.006) and Treatment 3 (t=3.75, p<0.001), similar to the overall case there was no significant difference between Treatments 2 and 3 in the fraction of conditional HEPDs.

Differences between treatments for yellow warnings

For yellow warning, one way ANOVA shows the differences among all three treatments are significant with respect to the fraction of conditional HEPDs (F (2, 6221) = 46.54, p<0.001) as well as the fraction of HEPDs (F (2, 20116) = 57.43, p<0.001). Participants in Treatment 1 made more conditional HEPDs than participants in Treatment 2 (t= 9.61, p<0.001) and Treatment 3 (t= 4.76,

p<0.001), participants in Treatment 3 significantly made more conditional HEPDs than participants in Treatment 2 (t=3.42, p<0.001). A similar pattern was observed with the HEPDs.

Differences between treatments for amber warnings

For amber warnings, one-way ANOVA shows the differences between treatments are significant with respect to the fraction of conditional HEPDs (F (2, 6026) = 10.15, p<0.001) as well as the fraction of HEPDs (F (2, 6026) = 11.64, p<0.001). These results also show that differences in participants' behaviour between treatments decreased (F value decreased) as the warning color went up from yellow to amber. Participants in Treatment 1 made significantly less often to choice for conditional HEPDs than participants in Treatment 2 (t= -3.83, p<0.001) and Treatment 3 (t= -3.77, p<0.001). Participants in Treatment 2 and 3 had no significant differences in the fraction of conditional HEPDs. However as participants in Treatment 1 had full information about likelihood and impact level, they made significantly more HEPDs than participants in Treatment 2 (t= 4.71, p<0.001) and Treatment 3 (t= 2.95, p=0.003) while there was no significant difference between Treatment 2 and 3. HEPDs is under the assumption that HEPDs if they all have full information about likelihood and impact level, conditional HEPDs are the highest payoff decision based on the information they have. Participants in Treatment 1 had higher fraction of highest payoff decisions and lower fraction of conditional highest payoff decisions means overall they more often to choose HEPDs as they have more information, but worse understanding of the warning than the other two groups.

Differences between treatments for red warnings

When the participants are issued with a red warning, the HEPDs is always move (risk free decision), so the fraction of conditional HEPD and HEPDs are the always same. One way ANOVA showed the differences between treatments are significant for the fraction of HEPDs and conditional HEPDs (F (2, 6000) = 14.99, p<0.001). Participants in Treatment 1 made significantly more conditional HEPDs than participants in Treatment 2 (t= 4.98, p<0.001) and Treatment 3 (t= 4.18, p<0.001). Participants in Treatments 2 and 3 do not show a significant difference in the fraction of HEPDs.

These results show that participants in each treatment respond differently to a given warning. Participants in Treatment 1 had the highest fraction of HEPDs cross all four warning colors; however they had the lowest fraction of conditional HEPDs when for amber warnings. There is another interesting point: when issued with a red warning, participants in Treatment 1 and 2 had exactly same information (there is only one box in the warning matrix which is red). However, the data shows that participants in Treatment 1 made significantly more HEPDs than participants in Treatment 2, while there were no differences between participants in Treatment 2 and 3. In other words, people follow the warning more by providing more information about the impact level and likelihood level.

Result 3. Participants' behaviour varied depending upon their payoff functions.

Participants behaved differently under different types of payoff functions. One way ANOVA shows the differences between types of payoffs are significant in expected payoffs per period (F (3, 20456) = 119.02, p<0.001) and in the fraction of conditional HEPDs (F (3, 20456) = 105.26, p<0.001). With a low damage payoff function (the damage is lower for each impact level comparing with high damage payoff), the participants' expected payoff per period is significantly higher (t=16.91, p<0.001) than

with a high damage payoffs, while the fraction of conditional HEPD is significantly lower (t=-7.54, p<0.001).

Comparing all four types of payoff functions, participants were more likely to make conditional HEPDs with impact sensitive payoffs (85.3%), and had the lowest fraction of conditional HEPD (73.4%) when they were answering likelihood sensitive payoff questions.

	Expected profit per period	Fraction of conditional HEPDs	Fraction of HEPDs
Low damage	1552	79.1%	76.5%
High damage	1500	84.9%	82.1%
Likelihood sensitive	1514	73.4%	72.4%
Impact sensitive	1540	85.3%	83.6%

Table 4：Expected profit per period, percent of conditional HEPDs and percent of HEPDs for different payoff functions.

We see that payoff type has a strong effect on the both fraction of conditional HEPD and earnings per period (Table 4). By increasing the impact level (damage) of the storm and moving cost unchanged, the expected profit per period decreases; however, with higher damages participants are more likely to choose the conditional HEPD. Results also indicate that participants more often behave risk neutral with impact sensitive payoff compared to likelihood sensitive payoff, potentially proving participants are more sensitive to the impact level rather than the overall likelihood. Participants chose to move significantly more often with the impact sensitive payoff than with the other payoff functions.

Participants in different treatments groups behaved differently in each individual question type. To find out how do different treatments affect participants under different payoff function, we looked at participants expected payoff based on their decisions, expected payoff if they only made the highest expected decisions based on the information they had, and the fraction of HEPDs and percent of HEPDs.

Low damage	Participants'	HEPD strategy	Percent of	Percent of HEPDs

	decision		conditional HEPDs	
Treatment 1	1566	1593	78.9%	78.9%
Treatment 2	1539	1582	78.4%	73.7%
Treatment 3	1551	1586	80.6%	77.2%
High damage				
Treatment 1	1512	1545	84.8%	84.8%
Treatment 2	1494	1520	85.6%	80.6%
Treatment 3	1491	1522	84.1%	80.0%
Likelihood sensitive				
Treatment 1	1538	1566	82.0%	82.0%
Treatment 2	1500	1525	69.3%	67.7%
Treatment 3	1495	1533	65.2%	63.6%
Impact sensitive				
Treatment 1	1549	1587	85.7%	85.7%
Treatment 2	1528	1580	81.8%	80.3%
Treatment 3	1542	1578	90.6%	85.7%

Table 5：Expected profit per period based on participants' decisions, HEPD strategy based on the information they have in each treatment for each payoff functions, percent of conditional HEPDs and percent of HEPDs in each treatment for each payoff function.

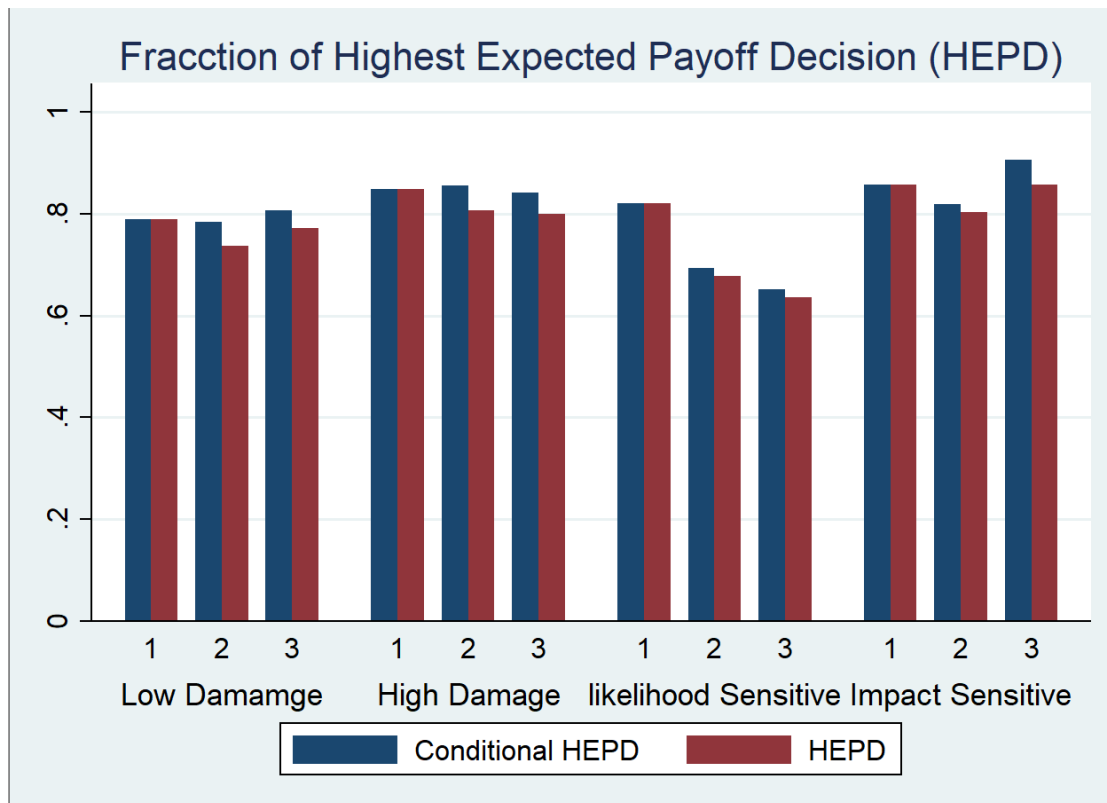highest expected payoff decision (HEPD)

Figure 4: Percent of conditional HEPDs and percent of HEPDs in each treatment for each payoff functions.

Low damage payoffs

For low damage payoffs, one way ANOVA showed the differences between treatments are significant ($F_{(2, 5111)} = 16.61$, $p<0.001$) in expected profit per period. Participants in Treatment 1 had significantly higher expected profit per period compared to participants in Treatment 2 ($t=2.76$, $p=0.006$) and Treatment 3 ($t=5.75$, $p<0.001$). However, there is no significant difference between Treatments 1 and 3 in HEPDs. Similar to the overall result there is no significant difference between Treatments 2 and 3 in profit per period. However, unlike the overall result there is no significant difference between treatments in the fraction of conditional HEPDs, one way ANOVA shows the differences between treatments are not significant ($F_{(2, 5111)} = 1.08$, $p=0.3408$). When we look at the fraction HEPD, one-way ANOVA shows that in this case the difference between treatments is significant ($F_{(2, 5111)} = 7.59$, $p<0.001$). Similar to the expectation profit, participants in Treatment 1 significantly more often chose the HEPDs compared to participants in Treatment 2 ($t=3.86$, $p<0.001$), but there is no significant difference between Treatment 1 and 3.

High damage payoffs

When participants were presented with a high damage payoff function, differences between treatments were significant in expected profit per period, as shown by one-way ANOVA ($F_{(2, 5111)} = 9.77$, $p<0.0001$). Participants in Treatment 1 had higher expected profit than participants in Treatment 2 ($t=3.71$, $p<0.001$) and Treatment 3 ($t=3.74$, $p<0.001$), Even though participants in Treatment 1 had higher payoff per period, the fraction of conditional HEPDs is not higher than the other groups. When we look at the fraction HEPD, one way ANOVA showed the differences between

treatments are significant (F (2, 5111) = 8.33, p<0.001). Participants in Treatment 1 chose the HEPDs significantly more often than participants in Treatment 2 (t=3.48, p<0.001) and Treatment 3 (t=3.40, p<0.001)

Likelihood sensitive payoffs

With likelihood-sensitive payoffs, participants reacted differently compared to participants with other payoffs. One way ANOVA shows the differences between treatments are significant (F (2, 5111) = 67.98, p<0.001) in the fraction of conditional HEPDs. To analyze the effect of treatments, we used a regression model to estimate the fraction of conditional HEPD. These results show the differences c between different treatments (group 2 coefficient=12.7%, t=9.19, p<0.001, group 3 coefficient =16.9%, t=10.39, p<0.001) are significant higher than in the general case (across all payoff functions) (group 2 coefficient=4.1%, t=6.54, p<0.001, group 3 coefficient =2.76, t=3.76, p<0.001). Looking at the fraction of HEPDs, one way ANOVA shows the differences between treatments are significant (F (2, 20457) = 74.55, p<0.001). The difference between Treatment 1 and Treatments 2 and 3 increased compare with conditional highest expected payoff decisions (group 2, t=9.79, p<0.001, group 3, t=-10.75, p<0.001) and the difference between Treatment 2 and 3 has no significantly change.

Impact sensitive payoffs

With impact sensitive questions, one way ANOVA shows the differences between treatments are significant (F (2, 20457) = 22.43, p<0.001) in the fraction of conditional HEPDs. However unlike all the other question types and the general case participants in Treatment 3 made more profit and had a higher fraction of conditional HEPDs than the other treatments. Participants in Treatment 3 significantly more often chose HEPDs compared to participants in Treatment 1 (t=3.50, p<0.001), and participants in Treatment 1 significantly more often chose HEPDs than participants in Treatment 2 (t=-3.70, P<0.001). One way ANOVA shows these differences in the fraction of HEPDs between treatments are significant (F (2, 5111) = 13.29, p<0.001). There was no significant difference between participants in Treatments 1 and 3 (t=-0.02, p<0.984), but the difference between participants in Treatment 1 and 2 is still significant (t=4.68, p<0.001). One way ANOVA shows the differences in the fraction of expected profit between treatments are also significant between treatments (F (2, 20457) = 5.58, p<0.0038).

Based on these results, participants in Treatment 1 had at least one of the highest fraction in HEPD across all four payoff functions. When participants were presented with likelihood sensitive questions participants in Treatment 1 did not show significant change compare with other payoff function, however participants in Treatment 2 and 3 had significant lower fraction in HEPDs and conditional HEPDs. Participants in Treatment 3 had the highest fraction in conditional HEPD and one of the highest fractions in HEPD with impact sensitive payoff decisions.

Result 4 *Participants' risk directions can be affected by the payoff function and the accuracy of the warning.*

It is important to know how participants' behave biased. On average, participants more often chose the risk-free decision than the risky decision (risk direction= -0.052, standard deviation=0.437). In other words, they had a slight preference for making a decision to move the product over the conditional HEPD. This also holds for each individual treatment. Risk direction did not vary between

treatments: one way ANOVA showed the differences in risk direction between treatments is not significant (F (2, 20457) = 1.83, p=0.16).

	Average risk direction	Standard deviation
Average	-0.0516	0.4364
Treatment 1	-0.0553	0.4101
Treatment 2	-0.0540	0.4574
Treatment 3	-0.0406	0.4441

Table 6： Average risk direction in each treatment.

For the analysis, we assume that amber and red warnings give a signal of 'move' (risk free), green and yellow warnings give a signal of 'not move' (risky). We classify the fraction of amber and red warning when HEPD is not move as rate of **false alarm**, as the warning shows a signal of move when the HEPD is not move; and we classify the fraction of green and yellow warning as **missed event**, as the warning shows a signal of not move when the HEPD is move. We use these two factors to evaluate warning systems in different treatments. We define **Error Level** as the fraction of false alarm and missed events of the total warning issued.

By comparing low damage and high damage payoffs functions, we examine whether the fraction of error level relates not only to the probability of the event but the impact level as well.

Firstly we want to evaluate the warning matrix, as we defined green and yellow is sigh of safe, amber and red warning is sigh of risky. So the optimal warning system should present green and yellow when the HEPD is 'not move', amber and red warning when the HEPD is 'move'.

likelihood	1.66%	6%	7.5%	30%
	1.66%	6%	7.5%	7.5%
	1.66%	1.66%	6%	7.5%

	1.66%	1.66%	6%	6%
	Impact level			

Table 7. The probability distribution for each element of the matrix

Based on Table 7 we can calculate the probability of false alarm (the probability of Amber or red warning if the HEPDs is not move over the probability of the HEPDs is not move) and miss event (the probability of green or yellow warning if the HEPDs is move over the probability of the HEPDs is move).

	Probability of false alarm	Probability of miss event	Percent of conditional HEPDs	Risk direction
Low damage	15.8%	0%	79.1%	-0.0760
High damage	0%	23.1%	84.9%	0.0606
Likelihood sensitive	23.3%	22.6%	73.4%	-0.1673
Impact sensitive	0%	9.1%	85.3%	-0.0236

Table 8: When the HEPDs were move/not move, percent of Green, Yellow / Amber, red warning and their corresponding percent of conditional HEPDs and risk directions.

In the previous discussion, we saw that when participants answered with high damage payoffs they chose the HEPD significantly more often than when they were answering with low damage payoffs (t=7.72, p<0.001).   Error level has significant effect on conditional HEPDs, one way ANOVA showed the differences between treatments are significant negative (t=-14.37, p<0.001, F (1, 20458) = 206.64, p<0.001) and significant effect on the risk direction, one way ANOVA showed the differences between treatments are significant negative (t=-14.37, p<0.001, F (1, 20458) = 288.88, p<0.001).

Table 8 shows, participants significantly less chose conditional HEPDs (t=-15.41,p<0.001,F (1, 20458) = 234.36, p<0.001) and significant more often to make risk-free decisions while they did not choose conditional HEPDs (t=-22.16,p<0.001,F (1, 20458) = 491.19, p<0.001) with likelihood sensitive payoff function.

With impact sensitive payoffs, as participants were sensitive to the impact level in these questions, so error level is high on the likelihood level but low on the impact level. For example, when the impact level is 3 and the likelihood level is 40%, the warning color is yellow but the HEPD is to move due to the high damage associated with this impact level (see Table 1). One way ANOVA shows participants significantly more often chose conditional HEPDs (t=9.65,p<0.001,F (1, 20458) = 93.06, p<0.001) and significantly less often chose to make the risk-free decisions while they were not the conditional HEPDs (t=5.30, p<0.001,F (1, 20458) =28.05, p<0.001) with likelihood sensitive payoff function.

To analyze the effect of false alarms and missed events, we build a regression model. One way ANOVA showed the fraction of false alarms and missed events have significant effect in fraction of conditional HEPDs (F (1, 20458) = 206.64, p<0.001). Both the fraction of false alarms (t=-17.53, p<0.001) and missed events (t=-2.28, p=0.023) have a significant negative effect on the fraction of conditional HEPDs, in other works the higher error level the less conditional HEPD participants will make. One way ANOVA shows the fraction of false alarm and missed events have a significant effect on the participants' risk direction (F (1, 20458) = 288.88, p<0.001). The fraction of false alarms (t=-25.33, p<0.001) has a significant negative effect on the risk direction level, in other works the false alarm level the more risk-free decision will be made by the participants. The fraction of missed events (t=1.92, p=0.055) has a positive effect on the risk direction but is not significant.

If we look deeper in each treatment, we can find out that both false alarm and missed events have an effect on the risk direction in each treatment. Similar to the overall case, false alarms have a negative effect on the risk direction in each treatment. In Treatment 1, the missed events had a positive effect on the risk direction (t=8.56, p<0.001). This means the higher the level of missed event participants more likely to choose risky decisions. Missed events have a significant negative effect in both Treatments 2 (t=-2.45, p=0.005) and 3 (t=3.16, p<0.001).

Our results show how false alarms and missed events affect participants' behaviour in the fraction of conditional HEPDs and risk direction. The higher the error level the lower the fraction of conditional HEPDs. This may be caused by the level of trust that participants have for this warning system.

Trainor et al. (2015) showed that participants felt there were less false alarms than the actual number of false alarms, and in our experiment the false alarm rate is too low (maximum 23.3%) which dampens any cry wolf effect and we find no significant evidence for one.

Result 5 *When participants received both warning color and risk information (impact and likelihood level), they followed both the warning color and risk information contained in the warning, but followed the color more.*

To analyze how warning color and information affect the participants' behaviour, we built a probit regression model to show the relation of dependent variable, whether or not participants choose the risk free decision (not move, y=-1) or the risky decision (move, y=1), and independent variables, warning color and information. As only participants in Treatment 1 had the full information and the warning color, so we only looked at the data of these participants.

We define two new variables ***is*** (information suggestion) and ***cs*** (color suggestion). When the conditional HEPD is not move ***is***=0, when the conditional HEPD is move ***is***=1. If the warning color is green or yellow then ***cs***=0, and if the warning color is amber and red ***cs***=1. The probability predictor of participants' behaviour is shown in Table 9.

$$\Pr\big(y = 1, (not\ move)\big) = \phi(\beta_0 + \beta_1 cs + \beta_2 is)$$

y	Coefficient	Std err	z	p>z	95% confident interval	
is	0.8037	0.0273	29.39	0.000	0.7502	0.8573

cs	1.0127	0.0276	36.68	0.000	0.9586	1.0668
cons	-0.3988	0.0156	-25.49	0.000	-0.4295	-0.3681

Table 9： Summary of how information suggestion and color suggestion affect participants' decisions.

The regression results show that both variables are significant and influence the decision but the coefficient for **cs** is significantly higher that of **is**.

4. Discussion and conclusion

In this paper, we used experimental methodology to compare participants' behaviour under different warning formats that varied based upon the information sent and the presentation format. We discovered that giving the impact level and likelihood level helps participants make better decisions (as indicated by the fraction of conditional HEPDs and HEPDs). However, when participants cared more about impact than likelihood because their damages rises steeply at higher impact levels, those given only the warning color (Treatment 3) more often chose the option that had higher expected payoffs conditional on their information. These results suggest that an effective warning system should not just have one presentation format, but should vary the format based upon the needs of those receiving it.

Broken down by warning colors, participants more often chose the conditional HEPDs when they had amber or red warnings than when they had green or yellow warnings. Participants that were given the likelihood and impact levels (Treatment 1) had a higher fraction of conditional HEPDs and higher profit in all warnings except amber and a higher fraction of HEPDs for all warnings. In other words, participants more often chose the HEPDs when the warning was more serious and the potential damage was high.

It is interesting that when the warning color was red, participants in Treatments 1 and 2 had exactly the same information (as there is only one red box in the matrix), however, the participants in Treatment 1 had a significantly higher fraction of HEPDs. Thus, more information caused more participants to follow the warning. The reason could be that, while they had the same information participants in Treatment 2 were less likely to follow the warning because they had less information about the meaning of the warning. Thus, more information could increase the trust in the warning system.

Participants made different decisions under different types of payoff function (low damage, high damage, likelihood sensitive and impact sensitive). With high damage payoffs, participants more often chose conditional HEPDs than when they had low damage payoffs (independent of treatment). This is more evidence that participants more often to choose the HEPDs when the potential damage was high.

With likelihood sensitive payoffs where there is little difference in damages between the impact levels. Participants were more sensitive to the likelihood level than the impact level. Under such payoffs, participants in Treatment 1 more often chose HEPDs and conditional HEPDs than participants in the other two treatments, while participants in Treatment 2 more often chose HEPDs and conditional HEPDs than participants in Treatment 3. The reason for this is that the design of the

warnings is not the most suitable for those with likelihood sensitive payoffs. Ideally, for such payoffs, the warning colors in the matrix should be stripes with the red in the top row and the green in the bottom row.

With impact sensitive payoffs where damages increase steeply at higher impact levels, participants in Treatment 3 had the highest fraction of conditional HEPDs and HEPDs of all three treatments. This result shows that when participants cared more about impact than likelihood just giving the warning color will help them to make more HEPDs. This also shows that giving people more information will not always help people make better decisions. In this case, the color was almost sufficient to make the HEPD. The one difference from the ideal warning is that the level 3 storm with a likelihood of 40% should be an amber warning rather than a yellow warning with parameters of the current experiment.

On average, participants more often choose the safe option when they are not choosing the conditional HEPDs. We find out that the error level significantly affects participants' risk directions and fraction of conditional HEPDs. Amber and red warnings when the HEPD is 'not move') and missed events (i.e., green and yellow warnings when the HEPD is 'move'. The results show that the higher the fraction of false alarms and missed events, the more participants chose the safe option over the highest expected payoff one.

For participants in Treatment 1, false alarms had a negative effect on the risk direction (as with overall effect on all three treatment groups)and missed events had a positive effect on a participants' risk direction (in contrast to the overall effect on all three treatment groups). This could be caused by participants using both the warning color and the information contained, so it is easier for participants in Treatment 1 to find out when it is false alarm or missed event. Our probit regression model shows that when participants had both information (impact level and likelihood level) and warning color, participants made use of both but followed the warning color more than the information.

Overall, our work shows that while increasing the information content of warnings is usually beneficial, it must be done with caution since better decisions (judged by higher profit) are not always made with an increase of information. There are many directions of future research. We can continue to run similar experiments with different parameters for the payoffs in order to better pinpoint shifts of behavior. Also, in our experiments, the weather warning perfectly matched the actual probabilities and level of the storm. While we did this for simplicity in our initial experiment, actual storms may occur at a higher or lower level rather than at a specific level or not at all. In addition, there could be ambiguity about the likelihood of the storm given the chaotic nature of the weather. Hence, giving warnings that have these additional realistic errors could be useful in determining how people behave and adjust to mistakes.

References

Abraham, S., Bartlett, R., Standage, M., Black, A., Charlton-Perez, A. and McCloy, R. (2015). Do location-specific forecasts pose a new challenge for communicating uncertainty?. Meteorological Applications, 22(3), pp.554-562.

Anon, (2017). The recent earthquake and tsunami in Japan: implications for East Asia. [online] Available at: (http://siteresources.worldbank.org/INTEAPHALFYEARLYUPDATE/Resources/550192-1300567391916/EAP_Update_March2011_japan.pdf) [Accessed 13 Aug. 2017].

Basher, R. (2006). Global early warning systems for natural hazards: systematic and people-centred. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 364(1845), pp.2167-2182.

Braun, C. and Silver, N. (1995). Interaction of signal word and colour on warning labels: differences in perceived hazard and behavioural compliance. Ergonomics, 38(11), pp.2207-2220.

Casteel, M. (2016). Communicating Increased Risk: An Empirical Investigation of the National Weather Service's Impact-Based Warnings. Weather, Climate, and Society, 8(3), pp.219-232.

Dale, M., Wicks, J., Mylne, K., Pappenberger, F., Laeger, S. and Taylor, S. (2012). Probabilistic flood forecasting and decision-making: an innovative risk-based approach. Natural Hazards, 70(1), pp.159-172.

Dels.nas.edu. (2018). The New Orleans Hurricane Protection System: Assessing Pre-Katrina Vulnerability and Improving Mitigation and Preparedness (2009) : Division on Earth and Life Studies. [online] Available at: http://dels.nas.edu/Report/Orleans-Hurricane-Protection-System/12647 [Accessed 3 Apr. 2018].

Demeritt, D., Nobert, S., Cloke, H. and Pappenberger, F. (2012). The European Flood Alert System and the communication, perception, and use of ensemble predictions for operational flood risk management. Hydrological Processes, 27(1), pp.147-157.

Doksaeter Sivle, A. and Kolstø, S. (2016). Use of online weather information in everyday decision-making by laypeople and implications for communication of weather information. Meteorological Applications, 23(4), pp.650-662.

Druckman, J. and Kam, C. (2009). Students as Experimental Participants: A Defense of the 'Narrow Data Base'. SSRN Electronic Journal.

Federal Aviation Administration. (2007). Introduction to Safety Management Systems (SMA) for Airport Operators (Advisory Circular), February 28, 2007. Washington, DC: U.S. Department of Transportation.

Fhwa.dot.gov. (2018). California Division | Federal Highway Administration | Federal Highway Administration. [online] Available at: https://www.fhwa.dot.gov/cadiv/ [Accessed 3 Apr. 2018].

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 10(2), pp.171-178.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. Journal of the Economic Science Association, 1(1), pp.114-125.

Harrison, G. and List, J. (2004). Field Experiments. SSRN Electronic Journal.

Jacoby, J., Speller, D. and Berning, C. (1974). Brand Choice Behavior as a Function of Information Load: Replication and Extension. Journal of Consumer Research, 1(1), p.33.

Joslyn, S., Pak, K., Jones, D., Pyles, J. and Hunt, E. (2007). The Effect of Probabilistic Information on Threshold Forecasts. Weather and Forecasting, 22(4), pp.804-812.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), p.263.

LeClerc, J. and Joslyn, S. (2015). The Cry Wolf Effect and Weather-Related Decision Making. Risk Analysis, 35(3), pp.385-395.

Marimo, P., Kaplan, T., Mylne, K. and Sharpe, M. (2015). Communication of Uncertainty in Temperature Forecasts. Weather and Forecasting, 30(1), pp.5-22.

Nadav-Greenberg, L. and Joslyn, S. (2009). Uncertainty Forecasts Improve Decision Making Among Nonexperts. Journal of Cognitive Engineering and Decision Making, 3(3), pp.209-227.

Neal, R., Boyle, P., Grahame, N., Mylne, K. and Sharpe, M. (2013). Ensemble based first guess support towards a risk-based severe weather warning service. Meteorological Applications, 21(3), pp.563-577.

Nhc.noaa.gov. (2017). Hurricanes in History. [online] Available at: http://www.nhc.noaa.gov/outreach/history/#katrina [Accessed 13 Aug. 2017].

Orsee.org. (2017). *ORSEE*. [online] Available at: http://www.orsee.org/web/ [Accessed 18 Aug. 2017].

Roulston, M. and Kaplan, T. (2009). A laboratory-based study of understanding of uncertainty in 5-day site-specific temperature forecasts. Meteorological Applications, 16(2), pp.237-244.

Rubinstein, A. (2007). Instinctive and Cognitive Reasoning: A Study of Response Times. The Economic Journal, 117(523), pp.1243-1259.

Ryan, J.P. (1991). Design of warning labels and instructions. New York: Van Nostrand Reinhold.UNISDR. Global Survey of Early Warning Systems: An Assessment of Capacities, Gaps and Opportunities Towards Building a Comprehensive Global Early Warning System for All Natural Hazards. United Nations Inter-Agency Secretariat of the International Strategy for Disaster Reduction (UNISDR), 2006. URL http://www.unisdr.org/we/inform/publications/3612.

Sivle, A. (2018). Oh no, it's raining! A study of how information in online weather reports is interpreted, integrated, and used in everyday decision-making by laypeople. [online] Bora.uib.no. Available at: http://bora.uib.no/handle/1956/15226?show=full [Accessed 6 Mar. 2018].

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5(2), pp.207-232.

Wickens, C., Gempler, K. and Morphew, M. (2000). Workload and Reliability of Predictor Displays in Aircraft Traffic Avoidance. Transportation Human Factors, 2(2), pp.99-126.

Winett, R. and Kagel, J. (1984). Effects of Information Presentation Format on Resource Use in Field Studies. Journal of Consumer Research, 11(2), p.655.

Ztree.uzh.ch. (2017). *UZH -.* [online] Available at: http://www.ztree.uzh.ch/en.html [Accessed 18 Aug. 2017].
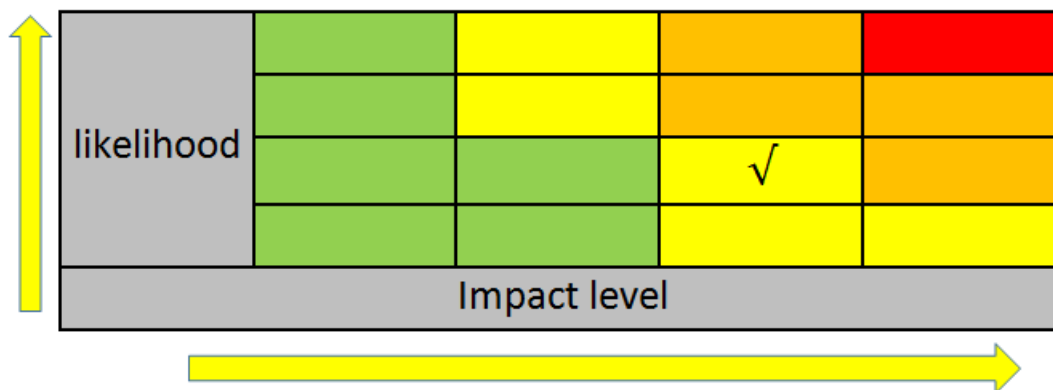
APPENDIX: Instruction

Instructions for Treatment 1

You will participate in an experiment involving extreme weather warnings. If you follow the instruction and make good decisions based on the warnings, your payoff may be higher. Your

earnings will depend on your decisions only, the other participants decisions will not affect your payoff. All the decisions that you make during this experiment will remain anonymous and will be recorded by the computer. You need to make decisions at the computer that you are seated at. Your earnings will be paid to you in cash, after you finish this experiment.

This experiment consists of 60 round. In each round you will see a graphic of a storm warning, similar to the one below. You are to choose whether or not to move your product to protect it from a potential storm. Moving involves a fixed cost of moving but no risk from the storm. Not moving has not fixed cost of moving but risks damage from the storm. In this experiment we successfully predicted the impact level and the likelihood of the warning with 100% certainty[6]. The vertical axis represents the likelihood that a storm going to happen, the horizontal axis represents the impact level of the potential storm.

You will also see information about the amount of the money that you have at the beginning of this round, the fixed cost of moving all products to a safe place, and the amount of the loss that may be caused by each level of the storm (all in tokens). For example

Endowment for this period	2000.0
The fixed cost of moving your products to a safe place is	500.0
The amount of damage caused by level 1 (very low impact) storm	300.0
The amount of damage caused by level 2 (low impact) storm	600.0
The amount of damage caused by level 3 (medium impact) storm	900.0
The amount of damage caused by level 4 (high impact) storm	1200.0
What is your decision?	Do not move
	Move

Your task is to study the graphic and information of the storm warning you then should choose whether or not to move all your products to a safe place.

At the beginning of each round, you need to look at the storm warning and the payoff information and then choose to move or not move all your products. If you choose to move your products, your

the reason we add this sentence in the instruction is because during the pilot experiment, participants were confused about the warning system, so we add this sentence to make sure they understand the experiment

profit in this round will be the money you have at the beginning of the round minus the fixed cost of moving all the products. If you choose to not move your products, then your profit in this round will be the money that you have at the beginning of this period minus the loss caused by the storm if the storm happened or all the money that you had at the beginning if the storm did not happen.

After you made your decisions, you will see information about the level of the storm, whether or not the storm occurred and your profit in this period similar to this.

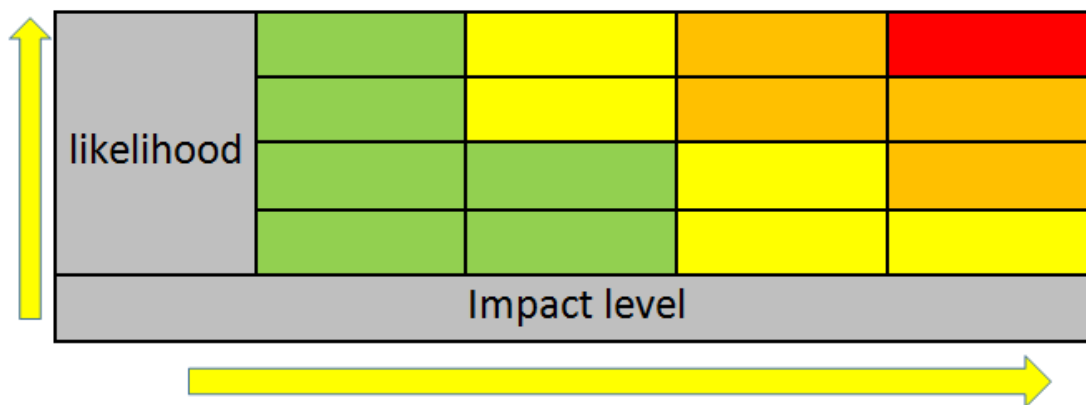|	
Endowment for this period	2000.0
The level of the storm	level 3 (medium impacct)
The storm	Did not occur
Your decision was	Not move
Your profit from this round will be	2000(endowment)
Your payoff for this period	2000.0

You will receive a 5-pound show-up fee plus any profit that you make in 5 randomly chosen rounds. The exchange rate is 1 pound for every 2000 token you make in the experiment.

Instructions for Treatment 2

You will participate in an experiment involving extreme weather warnings. If you follow the instruction and make good decisions based on the warnings, your payoff may be higher. Your earnings will depend on your decisions only, the other participants decisions will not affect your payoff. All the decisions that you make during this experiment will remain anonymous and will be

recorded by the computer. You need to make decisions at the computer that you are seated at. Your earnings will be paid to you in cash, after you finish this experiment.

This experiment consists of 60 round. In each round you will see a graphic of a storm warning, similar to the one below. You are to choose whether or not to move your product to protect it from a potential storm. Moving involves a fixed cost of moving but no risk from the storm. Not moving has not fixed cost of moving but risks damage from the storm. **In this experiment we successfully predicted the color of the warning with 100% certainty. The vertical axis represents the likelihood that a storm going to happen, the horizontal axis represents the impact level of the potential storm.**

You will also see information about the amount of the money that you have at the beginning of this round, the fixed cost of moving all products to a safe place, and the amount of the loss that may be caused by each level of the storm (all in tokens). For example

Endowment for this period	2000.0
The fixed cost of moving your products to a safe place is	500.0
The amount of damage caused by level 1 (very low impact) storm	925.0
The amount of damage caused by level 2 (low impact) storm	975.0
The amount of damage caused by level 3 (medium impact) storm	1025.0
The amount of damage caused by level 4 (high impact) storm	1075.0
The color of the warning	Amber
What is your decision?	Do not move
	Move

Your task is to study the graphic and information of the storm warning you then should choose whether or not to move all your products to a safe place.

At the beginning of each round, you need to look at the storm warning and the payoff information and then choose to move or not move all your products. If you choose to move your products, your profit in this round will be the money you have at the beginning of the round minus the fixed cost of moving all the products. If you choose to not move your products, then your profit in this round will be the money that you have at the beginning of this period minus the loss caused by the storm if the storm happened or all the money that you had at the beginning if the storm did not happen.

After you made your decisions, you will see information about the level of the storm, whether or not the storm occurred and your profit in this period similar to this.

Endowment for this period	2000.0
The level of the storm	level 3 (medium impacct)
The storm	Did not occur
Your decision was	Not move
Your profit from this round will be	2000(endowment)
Your payoff for this period	2000.0

You will receive a 5-pound show-up fee plus any profit that you make in 5 randomly chosen rounds. The exchange rate is 1 pound for every 2000 token you make in the experiment.

Instructions for Treatment 3

You will participate in an experiment involving extreme weather warnings. If you follow the instruction and make good decisions based on the warnings, your payoff may be higher. Your earnings will depend on your decisions only, the other participants decisions will not affect your

payoff. All the decisions that you make during this experiment will remain anonymous and will be recorded by the computer. You need to make decisions at the computer that you are seated at. Your earnings will be paid to you in cash, after you finish this experiment.

This experiment consists of 60 round. In each round you will see a graphic of a storm warning, similar to the one below. You are to choose whether or not to move your product to protect it from a potential storm. Moving involves a fixed cost of moving but no risk from the storm. Not moving has not fixed cost of moving but risks damage from the storm. **In this experiment we successfully predicted the color of the warning with 100% certainty. The vertical axis represents the likelihood that a storm going to happen and the impact level of the potential storm (expectation of the damage of the storm).**

You will also see information about the amount of the money that you have at the beginning of this round, the fixed cost of moving all products to a safe place, and the amount of the loss that may be caused by each level of the storm (all in tokens). For example
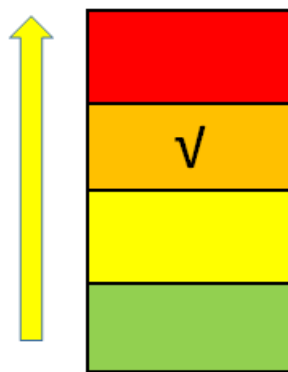
Your task is to study the graphic and information of the storm warning you then should choose whether or not to move all your products to a safe place.

Endowment for this period	2000.0
The fixed cost of moving your products to a safe place is	500.0
The amount of damage caused by level 1 (very low impact) storm	300.0
The amount of damage caused by level 2 (low impact) storm	600.0
The amount of damage caused by level 3 (medium impact) storm	900.0
The amount of damage caused by level 4 (high impact) storm	1200.0
What is your decision?	Do not move
	Move

At the beginning of each round, you need to look at the storm warning and the payoff information and then choose to move or not move all your products. If you choose to move your products, your

profit in this round will be the money you have at the beginning of the round minus the fixed cost of moving all the products. If you choose to not move your products, then your profit in this round will be the money that you have at the beginning of this period minus the loss caused by the storm if the storm happened or all the money that you had at the beginning if the storm did not happen.

After you made your decisions, you will see information about the level of the storm, whether or not the storm occurred and your profit in this period similar to this.

Endowment for this period	2000.0
The level of the storm	level 3 (medium impacct)
The storm	Did not occur
Your decision was	Not move
Your profit from this round will be	2000(endowment)
Your payoff for this period	2000.0

You will receive a 5-pound show-up fee plus any profit that you make in 5 randomly chosen rounds. The exchange rate is 1 pound for every 2000 token you make in the experiment.

Test questions

Test question 1

In one period of this experiment, one participant had initial endowment A

They will be faced with the problem of choosing whether or not to move their product to protect it from a potential storm. Moving involves a fixed cost of moving, but no risk from the storm. Not moving does not have the fixed cost of moving, but risks damage from the storm.

The damage of level 1 storm is	B1
The damage of level 2 storm is	B2
The damage of level 3 storm is	B3
The damage of level 4 storm is	B4
The fixed cost of moveing the product is	C

At the end of this period, the participant decided to move the product. Eventually the storm did not occur.

What is the participant's payoff at this period?

- A+C
- −C
- A+B3
- A−C

Test question 2

In one period of this experiment, one participant had initial endowment   A

They will be faced with the problem of choosing whether or not to move their product to protect it from a potential storm. Moving involves a fixed cost of moving but no risk from the storm. Not moving has not fixed cost of moving but risks damage from the storm.

The damage of level 1 storm is   B1
The damage of level 2 storm is   B2
The damage of level 3 storm is   B3
The damage of level 4 storm is   B4
the fixed cost of move the product is   C

At the end of this period, the participant decided not to move the product. Eventually a level 3 storm occurred.

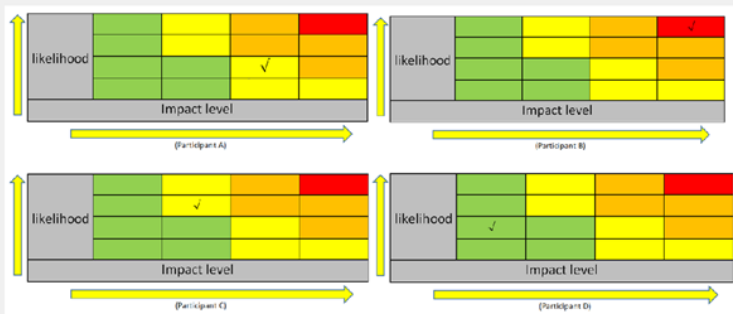What is the participants' payoff?

| A+C |
| B3 |
| A−B3 |
| A−C |

Test questions for Treatment 1

In one period of the experiment, participants A, B, C and D had four different warnings.
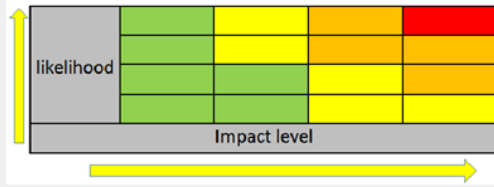
The potential storm which participant A is facing is one of the least likely to happen.   ○ True ○ False

The potential storm which participant C is facing definitely will cause less damage than the storm participant A is facing.   ○ True ○ False

Participant B definitely will suffer more damage than the other participants.   ○ True ○ False

OK

Test questions for Treatment 2

In one period of the experiment, participants A, B, C and D had four different warnings.
The color of the warning participant A had is yellow, the color of the warning participant B had is red,
the color of the warning participant C had is amber, the color of the warning participant D had is green.
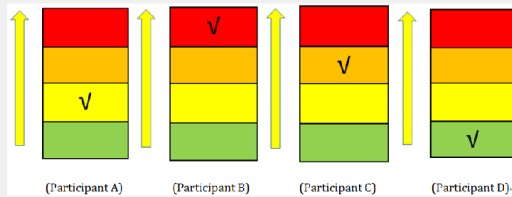
The potential storm which participant D is facing has one of the least impact.   ○ True
                                                                                  ○ False

The potential storm which participant C is facing definitely will cause less damage than the   ○ True
                                                storm participant D is facing.   ○ False

Participant B definitely will suffer no less damage than the other participants.   ○ True
                                                                                  ○ False

OK

Test questions for Treatment 3

(Participant A)   (Participant B)   (Participant C)   (Participant D)

In one period of the experiment, participants A, B, C and D had four different warnings.

The potential storm which participant D is facing has one of the least impact.   ○ True
                                                                                  ○ False

The potential storm which participant C is facing definitely will cause less damage than the   ○ True
                                                storm participant D is facing.   ○ False

Participant B definitely will suffer no less damage than the other participants.   ○ True
                                                                                  ○ False

OK