PROCEEDINGS OF THE INTERNATIONAL
CRISIS AND RISK COMMUNICATION CONFERENCE
MARCH 10-12, 2025. CLEMSON SC, USA

Evidence for Overthinking and Poor Decision Making after Mistakes: Implications for Crisis and Risk Communication

Scott C. Merrill
University of Vermont
Burlington, Vermont, USA

Christopher Koliba
University of Kansas
Lawrence, KS, USA

Rodrigo Soares
University of Vermont
Burlington, Vermont, USA

Ruth Quainoo
University of Vermont
Burlington, Vermont, USA

Eric Clark
University of Vermont
Burlington, Vermont, USA

Trisha Shrum
University of Vermont
Burlington, Vermont, USA

Masood Ali Khan
University of Vermont
Burlington, Vermont, USA

Jessica Balerna
University of Vermont
Burlington, Vermont, USA

Asim Zia
University of Vermont
Burlington, Vermont, USA

Abstract: Extreme weather events, such as floods, are increasing in frequency, severity, and extent, which creates additional pressure to effectively communicate during crises. Moreover, effective communication should only be considered effective if it generates appropriate responses. Here we sought to understand how different types of flood hazard warnings were perceived and acted upon over the course of an experiment. Of specific interest is how the decision making process evolved and changed throughout the experiment (behavioral dynamism), as both the exposure to the warnings and the potential for making significant response errors increased. To test flood warning message efficacy and behavioral dynamism, we used a serious game methodology to gather data from the public. Specifically, using video game software, we developed a virtual neighborhood with an approaching storm, and then provided participants with flood hazard warning messages. Participants would then decide how to respond to the messages by either evacuating or preparing for the storm at home. Data were collected while participants experienced a series of these simulated events. A nationally representative sample of 746 participants completed the experiment. Our findings revealed that participants made significantly worse decisions after making a costly mistake, providing evidence of cascading errors. Moreover, on average, they took longer than expected to make the poor decisions, indicating that they may have been "overthinking". Overall, results provide evidence of a difficult-to-measure phenomenon, where individuals make sequentially bad decisions. Finding ways to reduce or break the cycle of bad decisions could help protect lives and property.

Keywords — Serious game, Flood Hazard, Risk Communication, Cascading Errors

---

SUGGESTED CITATION: Merrill, S.C., Koliba, C.J., Soares, R., Quainoo, R., Clark, E., Shrum, T., Ali-Khan, M., Balerna, J. & Zia, A. (2025). Evidence for overthinking and poor decision making after mistakes: Implications for crisis and risk communication. *Proceedings of the 2025 International Crisis and Risk Communication Conference, 13(*1) 9-12. https://www.doi.org/10.69931/001c.142819

ISSN: 3068-6539

https://doi.org/10.69931/001c.142819

INTRODUCTION

Weather-related crises including floods are becoming more frequent, intense, and predicted to affect an increasing number of people in the future (1). This increases the pressure for federal, state, and local officials to communicate with the public, ideally using strategies that generate appropriate and timely responses. Like many weather-related forecasts, flood forecasts, and subsequent communications, are challenging because of the inherent uncertainties associated with i) the models driving the forecasts, ii) the interpretations of those models, and iii) the decision-making behavior stemming from information provided about the flooding events. Currently, the National Oceanic Atmospheric Administration (NOAA) including the National Weather Service (NWS) and Office of Water Prediction (OWP) are developing and testing new forecasting tools for the United States, (e.g., flood inundation maps which depict the geographical extent of flooding). Moreover, NOAA and the NWS are highly trusted by the public and are a primary source for flood forecasts (2).

Effective messaging requires a number of components, such as an explanation of the problem, essential action steps, and information as to why the message is important to the person receiving the message (3). One goal is to ascertain and quantify likely responses to warning messages that are currently or in plans for use in order to appropriately resource emergency responses. Yet, multiple processes may drive decision-making, and determining the distribution of likely behaviors within a context is dependent upon the experience, attitudes, and perceptions of those making the decisions as well as the social-ecological context of the decision. Moreover, responses and decisions may change as experience accrues, including learning from previous decisions. One methodological approach to gathering dynamic decision-making data is by conducting serious game experiments. These experiments can be deployed to simulate pressure and conflict, increasing the potential salience and engagement with the process. Importantly, they allow for hypothesis testing, especially in the context of decision-making and behavior under conditions of risk and crisis. The serious game methodology, with a controlled experimental design, is effective for risk communication research because it allows for collection of dynamic response data under simulated crisis conditions (e.g., under pressure and with stakes tied to performance) without harming the participants. Here we are interested in extending our investigation of flood hazard message efficacy to include responses after a serious mistake has been made. Evidence suggests that decision-making can be improved by providing feedback about the consequences of previous decisions e.g. (4). That is, we assume that people will improve their decision-making after a mistake if we tell them the consequences of their decisions and the costs of the mistake. However, we do observe people in crises making multiple, cascading poor decisions. Here, we explore aspects of behavioral dynamism, such as responses directly following a mistake, to determine if decision-making improves with experience, as well as to test the assumption that reflexivity is positively associated with improved decision-making. To test these hypotheses, we developed a serious game that simulated a flooded street-level environment, including sound, storm effects and flooding. A demonstration version of the game can be found at: https://segs.w3.uvm.edu/demos/CIROH-Flood-Evac1.6/

METHODS OR PROCEDURES

An experiment was developed to test the efficacy of flood warning message types, with an emphasis on collecting dynamic behavioral data, such as how people learned to interpret and respond to the forecasts over time, as well as how people responded to mistakes. Data were collected using a serious game methodology which is detailed in Merrill et al. (5). The serious game experiment was developed using Unity (2021.3.18f1) (Fig. 1). The experiment included 24 treatment rounds. During each round participants could make or lose experimental dollars, which were translated to real-world $USD at the end of the experiment. The experiment had a straightforward mechanic, where the participant was exposed to a flood warning message and then simply had to choose to evacuate or prepare at home. Time pressure was induced by increasing the cost of the decisions over time (after an initial nine second grace period). Each treatment round consisted of participants being provided with flood forecast information after which they were required to make a decision to either press the evacuate button (Icon with a running individual in Fig. 1) or the prepare at home button (Icon with a house in Fig. 1). Participants were told that they should evacuate if there was a “Moderate” or “Major” flood and prepare at home if there were a “Minor” or a “Limited” flood was forecast. As each treatment round progressed, the costs for evacuating or preparing at home increased, adding a time pressure component to participant decisions. The 24 experimental treatments were comprised of six different flood warning message types (such as the flood inundation map infographic depicted in Fig. 1), with each message type used for each of four flood severity levels: a) Major Flooding, b) Moderate Flooding, c) Minor Flooding, and d) Limited Flooding. Treatment order was randomized. Critically, after each round participants were provided with feedback that informed them of the appropriate decision (i.e., which button they should have pressed) and how much money they made or lost during that round. During each round, data collected were: 1) Decision to prepare at home or to evacuate, 2) Length of time before the decision is made,

Figure 1 shows a screen grab of one of the serious game treatment rounds. Not selecting the evacuation button when a Moderate or Major flood was predicted could lead to a *Shock* and loss of real-world $USD.

and 3) Earnings or losses associated with the decision. Additionally, we created a variable labeled Appropriate Decision, that was coded as "1", when the participant chose to evacuate when there was either a "Moderate" or "Major" flood, and prepare at home when there was no flood, a "Minor" flood, or a "Limited" flood. Appropriate Decision was coded as "0" if they evacuated when they should have prepared at home ("Minor" flood or "Limited" flood) or prepared at home when they should have evacuated ("Moderate" or "Major" Flood). As noted, a key variable that was collected was how much money was made or lost in each round. At the start of each round, participants were provided with \$3,000. Evacuation initially cost \$2,000 and could increase to \$3,000 if the participant waited until the last second before deciding. Thus, if a participant chose to evacuate by pressing the Evacuate button, then they would not lose money that round, but instead earned a maximum of \$1,000 experimental dollars (e.g., the initial \$3000 minus \$2000 evacuation costs). Conversely, pressing the prepare at home button, would cost the participant between \$0–\$200 experimental dollars plus any costs associated with flooding. Thus, if there is no flood and the participant quickly chose to prepare at home, they would have the potential to earn \$3,000 experimental dollars. However, if they decided to prepare at home and a "Moderate" or "Major" flood hit, the associated flood costs could push them into experimental dollar losses. For example, if they prepared at home when a "Major" flood hit, and that flood cost \$5,200, then that participant would lose \$2,200 in that round (\$3,000 - \$5,200). For the purposes of this analysis, we defined a "*Shock*" as a round in which a participant lost money. All decisions made before a participant's first *Shock* were coded as "Before", decisions made in the round immediately after the first *Shock* were coded as "Just After", and all decisions made after that were coded as "After". As appropriate for non-parametric, and non-normal distributions, differences in the distribution of decisions between the *Shock* categories (Before, Just After and After) were tested using two-sided Wilcoxon rank sum test with continuity correction (WRST), and all analysis was completed using R.

RESULTS

The serious game experiment was completed by 746 people from February to April of 2024. Participants were recruited through Amazon Mechanical Turk. Initial findings suggest that the flood level and message type were strongly associated with the quality of decisions made by the participants. These results are detailed in (5). Behavioral dynamism associated with a *Shock* was striking, largely because data suggest that providing feedback did not uniformly or adequately improve behavior. Specifically, using the entire data set, decisions were made appropriately 80.2% of the time before a *Shock* decision was made, while after a *Shock* decision was made, decisions were made appropriately 68.6% of the time. When considering only participants who never made a *Shock* decision, those participants (n=132 out of 746 participants) made appropriate decisions at a much higher rate than the rest of the population (84.5% of decisions were appropriate).

All further analyses were completed using only the 614 (or the total 746 participants - 82.3%) who experienced a *Shock.* After removing individuals who never experienced a *Shock*, the patterns remained. Decisions made before a *Shock* were made correctly 76.2% of the time, and afterwards they were made 68.6% of the time (WRST, $W = 1017804$, p-value<0.001). Moreover, decisions were made more slowly than expected after a *Shock*. This effect is more complex to quantify because it is confounded by an effect that decisions were made more quickly on average as the game progressed (a learning effect). Specifically, results show that, on average, participants made decisions approximately 0.1 seconds faster per round of play (Effect size = -0.105 seconds/round, t-value= -13.77, p-value< 0.001), e.g., a decision made in Round 5 was approximately one second slower than a decision made in Round 15. Results show that on average decisions after a *Shock* were made approximately 0.3 seconds slower compared to expectations (WRST: W=139413, p-value<0.001).

Some evidence in the data reveals that not all message styles result in the same depression in the ability to make decisions. E.g., there does not seem to be a reduction in the rate of appropriate responses when standard text message alerts are used. This is in contrast with a substantial drop in message efficacy if the currently used infographic (i.e., depicting a polygon warning region on a map) is used. With this infographic, the rate of appropriate responses is reduced from 67.2% before a *Shock* to 55.1% after a *Shock*: For context, making an appropriate decision 50% of the time is equivalent to flipping a coin to make a decision.

DISCUSSION AND CONCLUSION

Message efficacy in response to a crisis has obvious implications for lives and property. Responses to those messages are frequently tested, allowing for quantification of preference and adequacy of understanding of the messages. As a crisis and risk communication community, we recognize that components of messages, such as detailed in the IDEA model (3), affect message efficacy. Yet, it is challenging to quantify behavioral dynamism components, how people change, learn, and adapt, or in this case, demonstrate maladaptive behavior, as they accrue experience. Here we show that making a *Shock* decision can lead to poor decision making. There are a few possible rationales for this behavior. There are individuals that take additional time to make subsequent decisions compared to the expected decision time. Some of these individuals are making better decisions (showing effective reflexivity), and some are making worse decisions (possible evidence of overthinking). Other individuals are making decisions after a *Shock* that are shorter than expected. Those that are making poor and faster decisions, may be fatalistic, or "giving up". Others are making better decisions and faster decisions, indicating that they were able to correct and learn from their mistakes. Further research should seek to identify systemic behavior associated with those who are making a series of poor decisions, (i.e., cascading errors), and quantify techniques that help break the cascading error cycle identified here. Moreover, we should attempt to identify drivers of these behaviors, and to determine if there are message components that reduce the likelihood for cascading errors. Messages could be examined for how they influence reflexivity efficacy (e.g., through improved feedback components) or perhaps research should examine adding another aspect to effective messaging – what to do if an error is made. This action component may allow us to overcome and educate people on what to do if everything goes wrong.

Behavior and decision-making in crises are challenging to study and quantify. Assumptions of logical behavior are fraught with error both in times of simple decisions and in more complex, life threatening contexts. Here we use a serious game methodology to generate data with components of causality and with the ability to capture behavioral dynamism. We can use these data to test message efficacy, and importantly, look for aspects of behavioral dynamism. Here we show that assumptions of ubiquitous improvement of behavior if feedback is provided (reflexivity) are potentially seriously flawed. Moreover, we show that a better assumption may be decreased levels of appropriate responses to conflict during crises, further that a substantial component of the population may be overthinking their circumstances and making cascading errors. This work serves as a first step towards understanding the prevalence of cascading errors in crises, by identifying the challenge, and noting that patterns are dependent upon communication components. Overall, revealing these challenges may help improve our overall communication strategy in crises, ideally resulting in reduced losses to property and lives.

Author Biography

Scott C. Merrill, (Ph.D. Colorado State University, Fort Collins, Colorado) Department of Agriculture, Landscape, and Environment, University of Vermont, Burlington, U.S.A. Email: Scott.C.Merrill@uvm.edu

Christopher J. Koliba (Ph.D. Syracuse University, Syracuse, New York) School of Public Affairs and Administration, University of Kansas, Lawrence, KS, U.S.A. Ckoliba@ku.edu

Rodrigo Soares (Ph.D. University of Central Florida, Orlando, Florida) Department of Agriculture, Landscape, and Environment, University of Vermont, Burlington, VT 05405, U.S.A. Rodrigo.Soares@uvm.edu

Ruth Quainoo (Ph.D. East Carolina University, Greenville, North Carolina) Department of Agriculture, Landscape, and Environment, University of Vermont, Burlington, VT 05405, U.S.A. Ruth.Quainoo@uvm.edu

Eric M. Clark, (Ph.D. University of Vermont, Burlington, Vermont) Department of Agriculture, Landscape, and Environment, University of Vermont, Burlington, VT 05405, U.S.A. Eric.Clark@uvm.edu

Trisha Shrum (Ph.D. Harvard University, Cambridge, MA) Department of Community Development and Applied Economics, University of Vermont, Burlington, VT 05405, U.S.A.Trisha.Shrum@uvm.edu

Masood Ali Khan (M.S. Technical University Dortmund, Germany, and Ardhi University, Tanzania) Department of Community Development and Applied Economics and Department of Computer Science, University of Vermont, Burlington, VT 05405, U.S.A. Mkhan6@uvm.edu

Jessica A Balerna, (Ph.D. University of South Florida, Tampa, FL) Gund Institute for Environment, University of Vermont, Burlington, U.S.A. Email: Jessica.Balerna@uvm.edu

Asim Zia (Ph.D. Georgia Institute of Technology, Atlanta, Georgia) Department of Community Development and Applied Economics and Department of Computer Science, University of Vermont, Burlington, VT 05405, U.S.A Asim.Zia@uvm.edu

This project received funding under award NA22NWS4320003 from NOAA Cooperative Institute Program. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA.

REFERENCES

[1] Wing, O. E. J., Lehman, W., Bates, P. D., Sampson, C. C., Quinn, N., Smith, A. M., Neal, J. C., Porter, J. R., & Kousky, C. (2022). Inequitable patterns of US flood risk in the Anthropocene. *Nature Climate Change, 12*(2), 156-162. https://doi.org/10.1038/s41558-021-01265-6

[2] Quainoo, R., Merrill, S. C., Soares, R., Myers, M., Ali Khan, M., Shrum, T., & Balerna, J. A. (2024). A National Survey of Flood Hazard Crisis and Risk Perceptions in the United States. *Proceedings of the International Crisis and Risk Communication Conference*, 6, 70-74. https://doi.org/http://www.doi.org/10.69931/UBBE7512

[3] Sellnow, D. D., Lane, D. R., Sellnow, T. L., & Littlefield, R. S. (2017). The IDEA Model as a Best Practice for Effective Instructional Risk and Crisis Communication [Article]. *Communication Studies, 68*(5), 552-567. https://doi.org/10.1080/10510974.2017.1375535

[4] Brown, J. C. (2006). The effects of behavioral and outcome feedback on prudent decision-making under conditions of present and future uncertainty. *Judgment and Decision Making, 1*(1), 76-85.

[5] Merrill, S. C., Koliba, C. J., Soares, R., Quainoo, R., Clark, E., Shrum, T., Ali-Khan, M., Myers, M., & Zia, A. (2024). Simulation-Based Experiments Reveal Differences in the Efficiency of Responses to Flood Warning Messages in Crisis. *Proceedings of the International Crisis and Risk Communication Conference,* 6, 49-53. https://doi.org/https://www.doi.org/10.69931/SMHU9552