Journal of Environmental Psychology 80 (2022) 101764

Contents lists available at ScienceDirect

Journal of Environmental Psychology

journal homepage: www.elsevier.com/locate/jep

Immersive storm surge flooding: Scale and risk perception in virtual reality

Mark Simpson [a,1,*], Lace Padilla [b], Klaus Keller [c,2], Alexander Klippel [a,3]

[a] Department of Geography, Pennsylvania State University, 304 Walker, University Park, PA, 16802, USA
[b] Cognitive and Information Sciences, UC Merced, 5200 N Lake Road, Merced, CA, 95343, USA
[c] Department of Geosciences, The Pennsylvania State University, 503 Deike Building, University Park, PA, 16802, USA

ARTICLE INFO

Handling Editor: L. McCunn

Keywords:
Virtual reality
Immersion
Natural hazards
Flooding
Risk communication

ABSTRACT

Immersive virtual reality (iVR) can enable users to experience phenomena at real-world scale. This attribute may be useful for communicating the risks of many natural hazards. Storm-surge is a flood hazard whose risk has proven challenging to communicate through traditional means, such as maps. When it comes to storm-surge flooding, iVR experiences have shown promise in increasing awareness of their danger. However, it is currently unclear whether iVR enhances risk perception over standard display methods, and how such experiences affect the interpretation of map products. To address these questions, we ran a between-participants experiment comparing the impact of display type (desktop versus iVR) on risk perception and spatial learning, using a custom-developed immersive simulation of storm-surge flooding. We measured perceived risk by having participants rate damage on a series of hypothetical storm-surge maps and making evacuation decisions in response to notional flooding. To understand if more accurately sized iVR representation led to better comprehension of flood heights, we measured participants' ability to point to flood heights in a real environment. We found that iVR increases map damage-ratings and real-world height estimation accuracy, but that iVR leads participants to report that they would evacuate (and evacuate others) at higher water levels, indicating a disconnect between their understanding of environmental and bodily danger. We found that both desktop and iVR experiences aid map interpretation while iVR improves recall of flood heights in the real world. These results provide an avenue towards greater understanding of how iVR may be used for hazard-risk communication and conveying spatial information from immersive environments.

1. Introduction

Immersive virtual reality (iVR) allows users to experience seemingly dangerous situations without actual risk. This has made training and education applications of iVR popular (Bowman & McMahan, 2007; Jensen & Konradsen, 2018), and recent advances make iVR increasingly accessible for public outreach and hazard-communication applications. While adoption of iVR appears lagging compared to some projections (Laurell et al., 2019), the market for immersive technologies has grown by two orders of magnitude between 2016 and 2020 (Grand View Research, 2021). In the context of natural-hazard communication, users often struggle with traditional communication products, such as maps (Bostrom et al., 2018; Huang et al., 2016). Research has found that images of a hazard can even influence evacuation behavior more than official warnings (Burnside et al., 2007). iVR experiences promise to engage users at much higher levels than traditional media by placing them (perceptually) into simulated environments, an effect sometimes termed *immersion* (Bowman & McMahan, 2007; Slater & Sanchez-Vives, 2016). However, it remains an open question whether the notoriously challenging topic of communicating risk associated with natural hazards, such as storm-surge flooding, would benefit from integrating iVR experiences.

To inform the application of iVR technologies to hazard risk communication, this study seeks to understand how learning hazard-related spatial information (i.e., the height of flood waters) in an immersive environment affects risk perception. iVR experiences may be

* Corresponding author.
E-mail address: mbsimpson@mitre.org (M. Simpson).
[1] Present address. mbsimpson@mitre.org, 202 Burlington Road, MITRE Corporation, Bedford, MA 01730, USA.
[2] Present address: Thayer School of Engineering, 14 Engineering Drive, Dartmouth College, Hanover, NH 03755, USA.
[3] Present address: Laboratory of Geo-information Science and Remote Sensing, Wageningen University and Research, Droevendaalsesteeg 3, 6708 PB, Wageningen, The Netherlands.

https://doi.org/10.1016/j.jenvp.2022.101764
Received 28 March 2021; Received in revised form 15 January 2022; Accepted 17 January 2022
Available online 25 January 2022
0272-4944/

especially effective for flood-risk communication, as higher levels of technological immersion (Slater, 1999) enable realistic representations that appear properly scaled—that is, as large or small as they are in real life (Merchant et al., 2014). Flooding in turn is a hazard with easily perceived metric properties such as water depth, which is apparent at natural human scales and closely associated with risk. iVR has been thoroughly investigated for a variety of training and education applications, with considerable–but not universal–success (Freina & Ott, 2015; Slater & Sanchez-Vives, 2016). However, little empirically-tested guidance exists for how to apply iVR to the communication of natural-hazard risks, and there are reasons to be skeptical. For example, the known positive affective responses to technological immersion (Hou et al., 2012; Sundar et al., 2017) may interfere with risk messaging. More fundamentally, it remains largely unknown if (and *how*) iVR experiences influence individuals' perceived risk over comparable non-immersive desktop experiences.

We examine the impact of an iVR experience on risk perception through a between-participants experiment comparing an iVR simulation of storm-surge flooding to a non-immersive desktop experience with identical content. Storm surge occurs when the water level in coastal areas greatly exceeds its normal height during a storm—influenced by factors such as wind, atmospheric pressure, and tides (Strahler & Strahler, 2005). Researchers have explored flood simulations as educational tools, and iVR (Bernhardt et al., 2019) and interactive desktop simulations (Lindner & Alsheimer, 2019) of flooding have separately been found to be effective risk-communication tools, but their relative impact on risk perception remains uncertain. We investigate the impact on risk perception by measuring *perceived risk* primarily using a map-based damage-rating task (having participants rate the amount of damage they expected at a map location subject to notional flooding). Participants also made a pair of notional evacuation decisions (asking participants to indicate at what flood height they would evacuate and order the evacuation of others). We based the map task on work in Mason (2018), that found participants using National Hurricane Center (NHC)-style maps made notional evacuation decisions that appeared to systematically ignore the very flood-risk information the maps were designed to present. In order to assess the impact of the different levels of immersion, we also examine how each experience impacts real-life estimation of flood heights, as well as how such real-life tasks impact later responses in the map-based damage-rating task.

This work contributes to iVR, environmental psychology, and hazard-risk communication in several ways. We provide insight into potential causal mechanisms of how iVR experiences impacts perceived risk of natural hazards, including scale cues (for true water depth) provided by increased immersion and affective responses to immersion. We conducted an experimental evaluation of the impact of immersion (desktop versus iVR) on risk perception, with this paper organized as follows: We begin by outlining the research background motivating our hypotheses and our experiment, before discussing the experimental methods, summarizing the results, and discussing our findings and their broader impacts on spatial learning in iVR and hazard communication applications.

1.1. Storm surge risk communication

Improving risk communication for flooding due to tropical storm surges has been the subject of considerable research (van Valkengoed & Steg, 2019). For instance, the National Oceanic and Atmospheric Research Administration (NOAA) invests in improving storm-surge communication (Zachry et al., 2015), as well as hazard risk communication more generally through initiatives such as the National Weather Service's *Weather Ready Nation* initiative (Uccellini & Ten Hoeve, 2019). Members of the public commonly underestimate the risk of storm-surge flooding, and often pay more attention to wind risk, such as that represented in the well-known Saffir-Simpson scale for Hurricanes (Morrow et al., 2015). For example, in semi-structured interviews of residents of Miami-Dade County, Florida, conducted by Bostrom et al. (2018) in 2010, two-thirds of interviewees mentioned wind hazards during the interview, but only 7% mentioned storm surge. This is despite storm surges being responsible for about half of the loss of life associated with tropical cyclones in the United States (Rappaport, 2014), largely due to a few extreme events such as hurricane Katrina (Camelo & Mayo, 2021). According to Dash and Gladwin (2007), more effectively incorporating flood height information into risk-communication products could lead to higher evacuation rates, but effective methods for doing so remain unclear.

Several studies have investigated using desktop and immersive systems to create 3D flood visualizations (simulations), but few measure the impact of such visualizations on flood risk perception. Lindner and Alsheimer (2019) evaluated an interactive tool for visualizing storm-surge flood heights by rendering water over images of local landmarks. They found that while users of the tool did not report being more likely to evacuate at low levels of storm surge after using it, they report being more likely to evacuate at higher levels. Reyes and Chen (2017) presented an immersive storm surge experience of Miami Beach, Florida, but their work presented no formal evaluation. Wang et al. (2019) found that stakeholders had a positive impression of 3D flood visualization and thought it was superior to equivalent 2D visualizations, but did not formally compare the two in terms of impact on risk perception or decision-making.

Bernhardt et al. (2019) examined an iVR experience for hurricane-risk communication. They showed a set of traditional warning products depicting hypothetical hurricane scenario to 124 participants. They then gave half the participants an iVR experience of a hurricane (that included storm-surge flooding). They found participants who had the iVR experience were statistically significantly more likely to indicate that they would prepare for a hurricane compared to participants who did not have the experience. However, it remains unclear whether iVR simulations have any advantages over similar but less-immersive simulations using 2D desktop or mobile displays. As iVR experiences present cost and deployment challenges over media products such as video, we need to establish whether iVR provides improvement in risk communication outcomes (e.g., actually leads to changes risk perception as intended) before recommending its use by general public or policymakers.

1.2. Maps and risk communication

Maps are commonly used to communicate potential hazards, but the general public often misinterprets depicted hazard information (Padilla et al., 2017). For example, numerous large-scale user studies found that when viewing a hurricane-forecast visualization entitled the *Cone of Uncertainty*, people incorrectly believed it showed the size of the storm growing over time (e.g., Padilla et al., 2017; Ruginski et al., 2016). Subsequent research found that more modern hurricane-forecast visualizations have problems of their own (Padilla et al., 2017). For example, recent work demonstrated that when hurricane forecasts were visualized as multiple paths, individuals overreacted when one path was shown directly hitting their town (Padilla et al., 2017). Moreover, maps are subject to simplistic heuristic-based (Gigerenzer & Gaissmaier, 2011; Kahneman et al., 1982) interpretations, such as equating distance from a hazard with danger even when not appropriate (Mason, 2018; Ruginski et al., 2016). In contrast, Cuite et al. (2017) in a survey-based study found that highly-localized, geographically-explicit warnings can increase evacuation intentions for those in danger and reduce undesirable "shadow evacuation." Visual realism in a display artifact can also influence map interpretation. Zanola et al. (2009) found that higher realism increased viewers' confidence in the depicted spatial data – a phenomenon which Kostelnick et al. (2013) point out may also have the power to mislead when it comes to depicting risk. Together, this work illustrates the importance of testing risk-communication visualizations to ensure that they do not produce unintended beliefs about hazards.

One specific series of map products are the NHC-created storm-surge maps for use by the public and decision-makers (Sherman-Morris et al., 2015). These maps show the likely areas of storm surges, the potential height of the storm surge, and the relative uncertainty that the storm surge will exceed a given height (see Fig. 2). However, few studies have attempted to evaluate how such maps impact risk perception. Sherman-Morris et al. (2015) examined NHC map design, evaluating different color schemes and legends but found no statistically significant differences in terms of interpretation accuracy (e.g., which marked map location had higher forecasted storm surge). Moreover, non-experts generally have difficulty interpreting 3D information presented on 2D maps, such as elevation (Rapp et al., 2007), which may contribute to issues in flood-map comprehension.

The motivation of our study derives partially from the results of a series of experiments by Mason (2018) using NHC-style storm-surge maps. She found that participants largely discounted flood-height information on NHC storm-surge maps when making hypothetical evacuation decisions, instead defaulting to simplistic heuristics (Gigerenzer & Gaissmaier, 2011; Kahneman et al., 1982). In order to evaluate how a general audience interprets NHC maps, she undertook three experiments. In each experiment, she showed participants NHC-style maps of hypothetical storm-surge scenarios, and asked if they would evacuate if in a home at a specific location on the map. Inconsistent results in the first two experiments hinted that participants were paying more attention to the location's distance to the shore than the predicted height of flooding. The third experiment examined the influence of distance explicitly and found that the location's distance from the shore statistically significantly predicted participants' evacuation decisions. Further, the results suggested that participants accounted for storm-surge height only in areas that were far away from the shore. Essentially, participants decisions appeared to be dominated by a relatively simplistic distance-from-shore heuristic over the flood-height information indicated on the map—a concerning finding since the flood heights are much more indicative of actual risk. However, evacuation decisions are subject to several potential confounds, which make them difficult to interpret, as discussed below in the *Risk Perception and Evacuation Decision-Making* section.

1.3. Risk Perception and Evacuation Decision-making

In this study, we focus on measuring risk perception rather than evacuation decision-making. Evacuation decision-making is a complex process subject to many factors (Dash & Gladwin, 2007; Riad et al., 1999), such as whether one lives in a mobile home (Huang et al., 2016), has a high degree of social support (Riad et al., 1999), or whether one owns pets (Whitehead et al., 2000). Risk perception is generally considered a component of evacuation decision-making (Riad et al., 1999; Trumbo et al., 2016), and here refers to the subjective judgment of the likelihood and severity of danger (e.g., high or low perceived risk), inclusive to the entire process of developing a risk judgment (Dash & Gladwin, 2007). In their review, Dash and Gladwin (2007) found that risk perception was a strong indicator of evacuation behavior. This relationship has been found in subsequent research (e.g., Trumbo et al., 2016), although there is debate over its relative importance as a behavioral predictor (Bubeck et al., 2012; van Valkengoed & Steg, 2019). In the context of risk communication, Kellens et al. (2012) argued that communication should focus on increasing perceived risk, as they found that perceived risk predicts information-seeking behavior. Morss et al. (2018) compared different types of textual hurricane warnings and found that fear- and impact-focused messages increased risk perception and evacuation intentions, but only slightly. They also found that fear-based messaging led to a higher perception that warnings were exaggerated. Moreover, they found that messaging explicitly including storm surge information did not alter evacuation intentions or increase risk perception. van Valkengoed and Steg (2019), in reviewing climate change and hazards adaptation literature, found that risk perception generally motivates adaptive behavior, but point out that the effect size of risk perception varies considerably between studies. Risk perception also varies with individual characteristics, including age, gender, and religiosity (Wachinger et al., 2013). Wachinger et al. (2013) conclude that the two most important factors in risk perception are personal experience of a natural hazard, and trust (or mistrust) of authorities—external information only impacts perceived risk when individuals have no personal experience.

1.4. Immersion and scale perception

iVR conveys spatial scale (i.e., physical size) information directly through more depth cues than 2D displays (Bowman & McMahan, 2007; McIntire et al., 2014), which may allow people to more easily comprehend the true size of real-world phenomenon. In the context of flood hazards, this is potentially useful as flood depth is a direct indicator of the potential danger of flooding (Jonkman & Vrijling, 2008). iVR allows for digital representations at different perceptual scales (perceived sizes relative to the user), unlike traditional displays which are limited to specific perceptual scales such as screen dimensions (Barba & Marroquin, 2017). Importantly, perceptual scale shapes how we perceive and understand spatial information, including that in virtual spaces (Montello, 1993; Newcombe, 2018). In the context of hazards, there may be a benefit in matching perceptual scale to the actual scale of the depicted phenomenon, e.g., showing flood waters that appear to be their actual height. iVR affords different perceptual scales through technological immersion (Barba & Marroquin, 2017). *Immersion* in this sense refers to the degree to which a system substitutes sensory information from the real-world with information generated by the system itself, following Slater's (1999) definition. Unsurprisingly, highly immersive head-tracked, stereoscopic head-mounted displays (HMDs) enable users to discern the relative sizes of objects in a virtual environment more easily compared to 2D displays (McIntire et al., 2014). In particular, the ability to use one's own height as a measuring tool is an important affordance of immersive environments (Dixon et al., 2000; Leyrer et al., 2011). While previous generations of HMDs tended to result in participants underestimating the size of virtual environments compared to equivalent real environments, modern HMDs like the HTC Vive appear to have largely attenuated this effect—for example, eliminating size underestimation in blind walking tasks (Kelly et al., 2017).

1.5. Immersion and affect

In order to better understand how iVR influences risk perception, it may be important to account for affective responses to immersion. Even relatively simple increases in technological immersion, such as increased screen size, induce positive affective responses to media (Hou et al., 2012). For example, Yeo et al. (2020) examined how experiencing natural underwater scenes impacted positive and negative affect in different delivery modes: 2D video, 360° video viewed in an HMD, and an interactive HMD VR experience. They found that while all conditions increased positive affect and reduced negative affect, positive affect was significantly improved in the HMD-VR condition over the 2D video. Affective responses in turn can influence decision making and risk perception, with positive affect being associated with lower perceived risk (Slovic et al., 2007). Both positive and negative affective responses have a strong influence on risk perception (Dixon, 2016). Media technology can induce affective responses beyond the messages they carry (Sundar, 2009), and understanding the impact of positive affective responses to novel technology has been a longstanding challenge in iVR research.

1.6. Experiment

To investigate the above issues, we ran a between-participants experiment comparing how an immersive and less-immersive (desktop) experience of storm-surge flooding impacts risk perception and whether this is linked to accurate real-world understanding of flood height. We also investigate affective emotional responses to the flooding experiences, which may influence risk perception. People tend to underestimate the damage of flooding (Dash & Gladwin, 2007), which is closely linked to flood height. Bernhardt et al. (2019) found that iVR experiences combined with traditional print products increased perceived risk compared to print products alone, and we expect iVR to similarly increase perceived risk over a non-immersive experience. We hypothesize that showing participants potential flood heights at real-world scale through an iVR experience increases perceived risk, measured through having participants make a series of judgements regarding expected damage ("damage ratings") at notional flooded locations on a series of maps (H1). Likewise, we expect self-reported confidence in those ratings to be higher (H2), based on the findings of Zanola et al. (2009) that higher degrees of immersion lead to greater confidence in an uncertainty visualization.

Mason (2018) found that participants appeared to largely discount flood-height categories when interpreting NHC maps. We expect the simulated experience to link realistic-appearing flood heights to the color categories indicating flood height in the minds of participants. This more accurate understanding of flood heights will result in higher damage ratings corresponding to greater flood heights (H3). The true-scale representation (i.e., rich in depth cues and more similar to the real-world —McIntire et al., 2014) in iVR will allow for easier transfer of spatial-scale information (e.g., *depth*) to the real-world (Bailey & Witmer, 1994). People use their own height as a unit of measurement to judge the height of objects in the real world (Twedt et al., 2012) and immersive environments (Leyrer et al., 2011). People do not use their height in this way in desktop virtual environments, reducing size estimation accuracy (Dixon et al., 2000). Therefore, we expect participants in the iVR condition to more accurately point to given heights (e.g., 9 ft) on an obscured measuring tape hung from the ceiling in the real world (H4).

However, because the height task asks them to imagine the height of flood water in a real room, we also expect that in performing the task participants will use height information from the real world to measure and remember the flood heights. The height task will have a small effect on iVR participants since they 1) already have been provided accurate scale information in terms of depth cues (Loyola, 2018; McIntire et al., 2014) and 2) have been able to integrate their own height into their estimations in a way similar to a real environment (Leyrer et al., 2011; Twedt et al., 2012). Therefore, in the desktop condition only, participants will be *first* cued to the real-world height of the flood waters when performing the height task. Size in the past has been underestimated in virtual environments (Renner et al., 2013), but this has shown to be less of an issue with more modern equipment like the HTC Vive (Kelly et al., 2017). Thus, we expect desktop participants will believe flood heights are effectively lower, leading to lower damage ratings when not cued in the height task. Therefore, we hypothesize that there will be an interaction between condition and order, such that desktop participants who completed the real-world height task before the map task will have higher map damage ratings than participants who completed the map task first, but no such difference within the iVR condition (H5).

While not a primary focus of this study, previous iVR research has found immersion can create positive affective responses (Sundar et al., 2017; Yeo et al., 2020). Moreover, results and participant feedback in our pilot studies indicated affective responses may have influenced results in the iVR condition. Positive affect can reduce perceived risk in some circumstances (Dixon, 2016), but the potential impact in the context of iVR and hazard risk perception is unknown. We measured affective responses with a self-report questionnaire immediately following exposure to the flood experience in both iVR and desktop conditions, but we treat our analysis in this area as exploratory.

2. Methods

2.1. Participants

We recruited fifty-six students and staff from a northeastern U.S. university, comprising of 23 males and 33 females with a mean age of 25.4 years ($SD = 11.2$), with half (28) randomly assigned to either the iVR or desktop condition, with 18 set as the minimum age for participation. Within each condition, half (14) were assigned to either complete the map damage-rating task before the real-life height-estimation task, or vice-versa (the manipulation of *order*), for 14 participants per condition/order combination. The required sample size was estimated from two pilot studies ($N = 8$, and $N = 10$, respectively), and assuming a small-to-medium effect size commonly found in iVR literature (Cummings & Bailenson, 2014). Individuals were paid $10 US for participating and university institutional review board approval was obtained for this study.

2.2. Design and procedure

We varied flood height condition (iVR or desktop) and order (map task first or real-life height task first) between participants. Participants were randomly assigned to a condition and order of tasks. After the consent process, the experimenter measured participants' height and interpupillary distance. After fitting of the HTC Vive iVR headset (used for all iVR experiences in the experiment), the participants completed an iVR informational experience. Participants then completed a questionnaire-based map training task to gain familiarity with map stimuli. Participants then moved to the flood simulation, either with the desktop or iVR system. See Fig. 3 for a diagram of the full procedure, which took about 50 min to complete. Before participants entered the flood experience, the experimenter described the features of the system and explicitly told the participant that the experience was not highly realistic in terms of flood behavior, especially the wave action. The realism statement served to control any extreme expectations regarding the fidelity of the simulation. After the flood experience ended, participants completed the post-experience questionnaire. Their next task depended on the order assigned, either completing the real-world height-estimation task first and the map-based damage ratings second, or vice-versa. Once those two tasks were completed, regardless of order, participants completed the demographic questionnaire and open-answer strategy questions before receiving their compensation and leaving.

2.3. Stimuli: virtual experiences

We created two types of virtual experiences in Unity (Unity Technologies, 2018), an informational experience that displayed facts about storm surge taken from an NHC pamphlet (National Hurricane Center, 2019), and the main experience which simulated storm-surge flooding. Both experiences took place in a generic American suburban neighborhood (see Fig. 1).

The informational experience took place in iVR, and had two purposes: 1) to give *all* participants an iVR experience to keep potential novelty effects of iVR constant across conditions, and 2) provide individuals with background information about storm surge. For the iVR experiences (the informational experience in both conditions, and the flood experience in the iVR condition), participants used an HTC Vive HMD, which has a resolution of 1080 x 1200 per eye (HTC Corporation, 2019). Within the environment, a virtual screen displayed a navigable series of facts and illustrations taken directly from a NHC pamphlet (National Hurricane Center, 2019) available online (e.g., "During the peak of a storm surge event, it is unlikely that emergency responders will be able to reach you if you are in danger."). Participants navigated through the information using an HTC Vive controller, which allowed them to end the experience after viewing all information, which

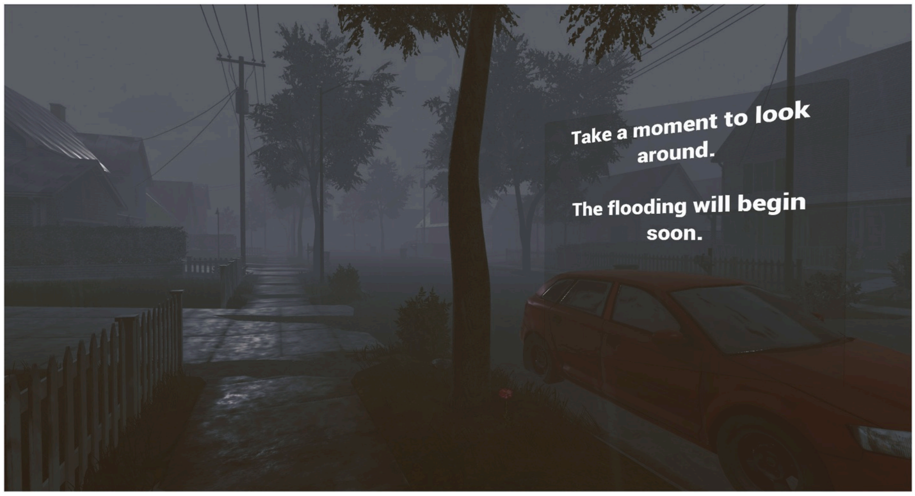

Fig. 1. The virtual-neighborhood simulation shortly before being inundated by storm-surge flooding up to twelve feet high (about 3.6 m). Participants viewed the experience with either a desktop monitor or an immersive virtual reality headset before completing tasks to measure their perceived risk of storm-surge flooding and ability to remember flood heights in the real world.

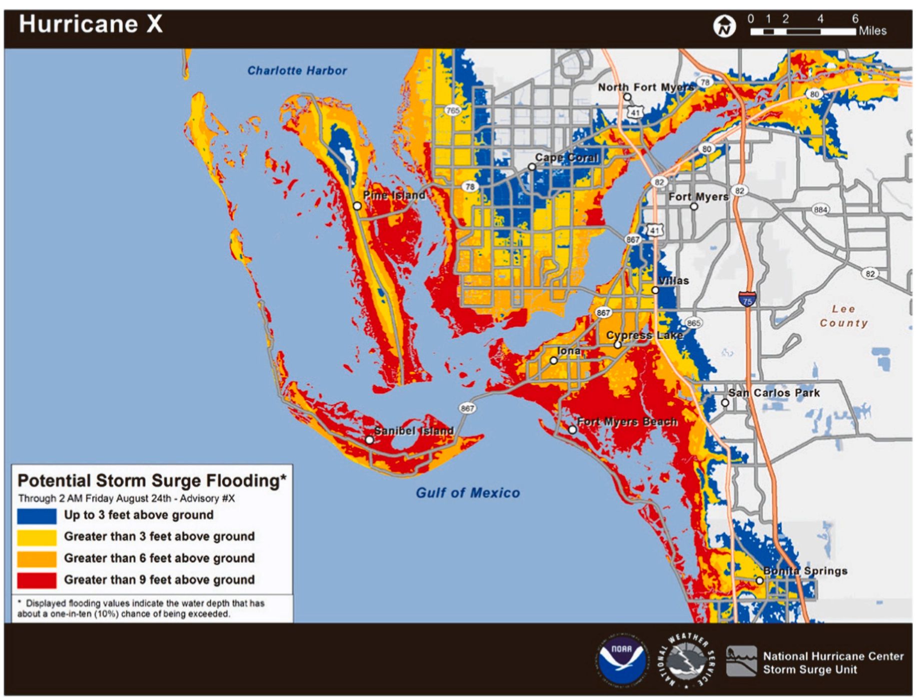

Fig. 2. Example storm-surge map reproduced from a National Hurricane Center pamphlet (The National Hurricane Center, 2017). Public domain image.

typically took about five minutes.

We implemented two nearly identical versions of the flooding simulation which differed primarily in terms of display and interaction: 1) Desktop-based, with participants using an Xbox One controller to look around (but not move) within the environment, and 2) iVR-based (HTC Vive) with no active interaction (participants were allowed to look around using the HMD). We used a Dell U2715Hc 27in monitor with 2560 x 1440 resolution on a stand (sitting approximately 1.6 m above the ground) as our desktop display. In both experiences, the virtual camera position was matched to participant height, and participants remained standing throughout. Desktop participants stood approximately 53 cm from the monitor and had a virtual field-of-view of 60° when viewing the flooding experience. iVR participants were instructed to stand in one location and rotate their bodies without moving from that location. For audio, desktop participants used Koss UR20 wired headphones while the iVR participants used the HTC Vive Deluxe Audio Strap. Aside from the display and input technology and height adjustment outlined above, all other aspects of the iVR and desktop conditions were nearly identical. That is, the virtual environments were the same in terms of content, such as 3D models, positioning of objects and the participant object behaviors, sound effects, lighting, and underlying code. However, the rain and raindrop visual effects were adjusted between experiences in order to achieve similar visual salience (the same settings resulted in a much brighter effect in iVR).

We designed the flood experience to illustrate the storm-surge flood heights depicted on the NHC maps in a semi-realistic manner in a realistic context. In order to emulate a realistic hurricane scenario, we replicated different aspects of severe weather, such as heavy rain, wind, lightning, and associated sound and visual effects (see Fig. 4 for screenshots of two flood stages). The environment flooded to the two

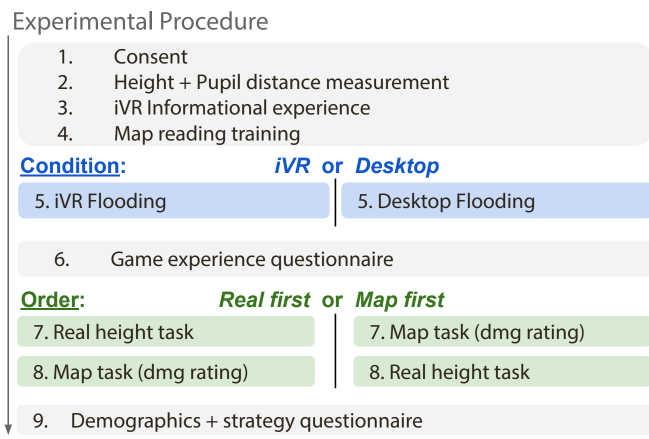

Fig. 3. Diagram of the experimental procedure. Condition and order are manipulated between subjects, for four potential combinations of condition-order.

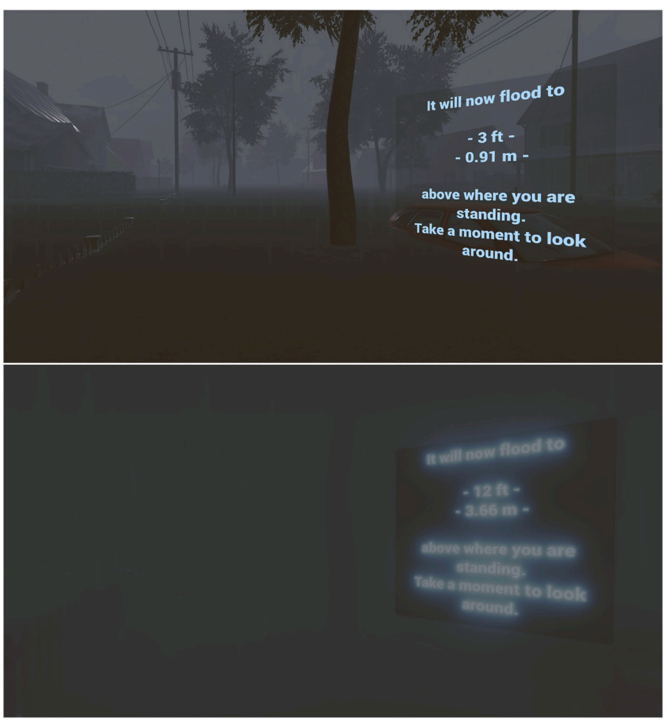

Fig. 4. A comparison of two flood height stages experienced by participants in the storm-surge simulation, from the same perspective as Fig. 1. The top screenshot shows three feet of storm surge, while the bottom screenshot shows twelve feet of storm surge.

heights mentioned in the NHC pamphlet (The National Hurricane Center, 2017), six inches (15.2 cm) and two feet (61.0 cm) and the three flood heights shown on the NHC maps in timed stages: three feet (0.91 m), six feet (1.83 m), nine feet (2.74 m), and 12 feet (3.7m) in lieu of "beyond nine feet." The water paused at each height for 25 seconds before flooding to the next height, starting from no flooding and proceeding to the maximum height. The flooding consisted of the water surface rising to the designated level, with no obvious source or directional flow but simulated waves. A highly visually realistic, real-time fluid simulation of flooding (and its effect on the environment) was prohibitively complex to implement. Instead, we opted to use a standard surface-based water model, with specific objects within the scene scripted to float. We also limited the severity of wave action and the opacity of water in order to ensure visibility of the water levels. The flooding experience lasted approximately three minutes.

The experience attempted to communicate the value of the different flood heights in several ways beyond the position of the water surface itself. A "screen" in the environment displayed text describing the current stage of flooding and indicated when the next stage would start. In the scene, we placed a 12 ft (3.7 m) tall rectangular "scale bar," that was colored to match to levels matching the color classes on the NHC maps (i.e., the bar was colored blue up to three feet, then yellow up to six feet, etc.). This provided a visual reference to the maps the participants would be viewing later. We allowed some objects near the participant to float, such as a garbage can and trash bags, to provide additional height cues.

2.4. *Material*

2.4.1. *Map damage-rating task*

We created a map-based task adapted from the evacuation decision task in Mason's (2018) study, that had participants interpret a series of hypothetical NHC storm-surge maps. We substituted map-based damage ratings for the evacuation questions posed in Mason's (2018) work. Damage ratings have been extensively studied in prior work with risk-communication visualizations resulting in a clearer picture of how those influence perceived risk (Ruginski et al., 2016; Padilla et al., 2017; Liu et al., 2016, 2018). For each map scenario, participants were asked to rate the amount of damage they expected a location marked "A" to incur on a Likert-style scale from one to seven (with one being "no damage" and seven being "severe damage"). We also asked participants to rate their confidence for each response on a scale from one to seven (one being "not at all confident," and seven being "very confident"). Each participant was presented with 32 maps and associated questions in a randomized order.

Each map represented a hypothetical storm-surge scenario on an area of north-south coastline, with their appearance based closely on NHC maps. Potential storm-surge flooding was represented by four colored categories: "Less than 3 feet above ground" (blue), "Greater than 3 feet above ground" (yellow), "Greater than 6 feet above ground" (orange), and "Greater than 9 feet above ground" (red). The maps varied in terms of 1) which of the four color categories contained the location of interest, 2) whether the point was offset north or south, 3) whether the coast faced east or west, an 4) whether the point was near or far from the shore, for a total of 32 maps. See Fig. 5 for examples. We included north/south and east/west alternations to provide variation for repeated measures. We varied distance to account for the distance effect found in Mason (2018).

2.4.2. *Real-world height tasks*

The real-world height task had two purposes in the final experiment: 1) measure knowledge transfer from the flood simulation to real life, and 2) provide (real-world) scale information that can be compared to that of the iVR experience. In the task, participants used a laser pointer to point to an obscured measuring tape hung away from any walls to minimize height references. The experimenter verbally instructed participants to point to where water would reach on the tape if the room flooded to the following heights: two feet (0.61 m), six feet (1.83 m), five feet (1.52 m), three feet (0.91 m), and seven feet (2.13 m). Participants stood 3.4 m from the measuring tape.

Participants performed two evacuation-decision tasks using the same setup as the height estimations. They were asked to point to the potential flood heights at which 1) they would order an evacuation (*manager*) and 2) they would personally evacuate (*personal*). These were intended to serve as secondary measures of perceived risk, although as discussed in the section *Risk Perception and Evacuation Decision-Making*, risk perception is only one potential predictor of evacuation behavior. In addition to testing for any issues specific to map comprehension, it provided risk-perception information more related to participants' personal risk perception, in contrast to the more abstract and environment-based

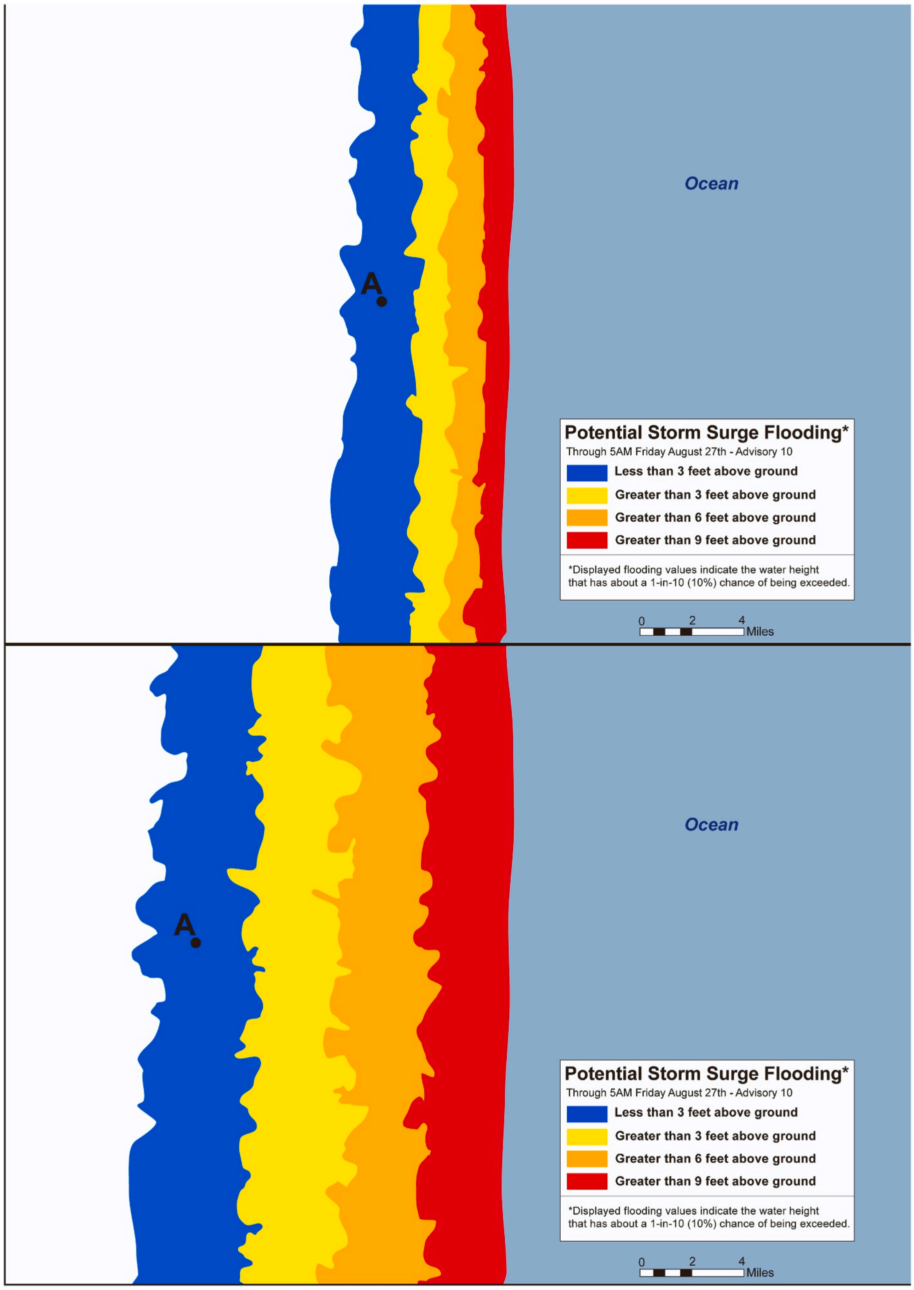

Fig. 5. Examples of the map stimuli used for the damage-rating task in the experiment. Participants were asked to rate the damage at location A. Both panels show a map with a point location in the blue category ("Less than 3 feet above ground"), with the top image having a near-shore point, and the bottom image having a point far from shore. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

damage-rating task.

2.4.3. *Affective measures*

To evaluate the potential role of affective responses to the virtual experiences, we used a modified Game Experience Questionnaire (GEQ) from IJsselsteijn et al. (2013). The GEQ is a popular questionnaire for measuring player experience in digital games (Johnson et al., 2018), and was selected for its inclusion of affective measures and its inclusion of items relevant to an immersive experience. However, the validity of the factors has been called into question in analyses by Law et al. (2018), and we caution that our analysis should be treated as exploratory. We used the core and post-game modules of the questionnaire, dropping questions related to Competence and Challenge since these were not applicable to a nearly non-interactive experience. Our modified version included Sensory and Imaginative Immersion (hereafter: *SI Immersion*), Flow, Tension/Annoyance, Negative Affect, and Positive Affect from the core module, and all components from the post-game module (Positive Experience, Negative Experience, Tiredness, and Returning to Reality). The non-affective measures were retained as it was thought they may provide insight into the perceived experience of the simulation not otherwise captured.

2.4.4. *Questionnaires*

Training, map damage rating, and post-questionnaires (including the GEQ) were administered through a Qualtrics web-based questionnaire (Qualtrics, 2019). We included a series of open-ended questions regarding participant strategies for the map damage ratings, confidence of those ratings, and the real-world height tasks (both estimation and evacuation questions), but do not analyse these results in the present work. Finally, we also included basic demographic questions and questions regarding previous experience with flooding, iVR, and video gaming.

3. Results

This section describes the results of the experiment, and the raw data, analysis scripts (R), and additional results are available in an Open Science Framework repository linked here: https://osf.io/6bgt3/?view_only = 20613851ce8d4ea0a8397e46b4aaf5f5.

3.1. Damage ratings

We used a multilevel model in R (lme package; Bates et al., 2015) to fit the data with restricted maximum likelihood estimation procedures (Raudenbush & Bryk, 2002). A multilevel model is appropriate for analyzing the variance in experimental outcomes predicted by both within-participant (e.g., given Color category) and between-participant (in our case, iVR condition) variables. In the primary damage-rating model, we included the predictors of flood height category (evaluating H1), order, and the interaction between condition and order (evaluating H5) to predict damage ratings. While not related to the main questions of the current study the effects of distance-from-shore, iVR experience, and video game experience were included as covariates to control for these effects in the model. The variables in the model were dummy coded such that the referents were; iVR (for Condition), the blue regions (for Color), the closest distance (for Distance), and real-world task first (for Order). The intercepts for each participant were allowed to vary in the model. Our damage rating model took the form of Equation (1) (using notation from Gelman & Hill, 2007).

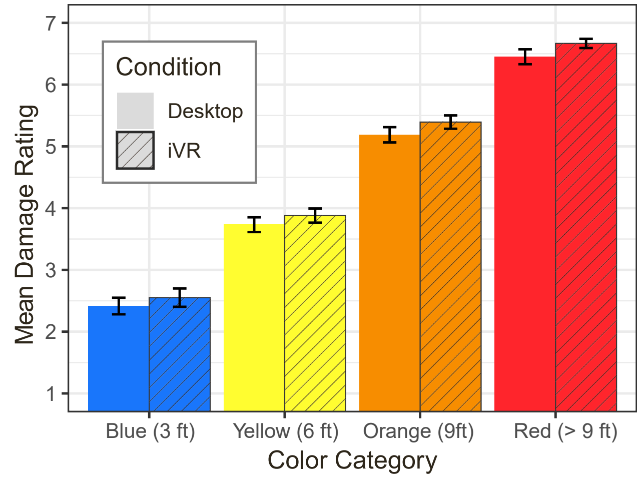

Fig. 6. Average confidence ratings (on a scale of 1–7) for each color height category and condition. Error bars represent 95% confidence intervals around mean damage ratings. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

higher map damage ratings, we included the main effect of condition. We found that individuals in the iVR condition had statistically significantly ($p < .05$) higher damage ratings ($M = 4.62$, $SD = 1.77$) compared to those in the desktop condition ($M = 4.45$, $SD = 1.79$), ($b = 0.38$, $t = 2.34$, $p = .019$, 95% CI [0.06, 0.70]). The resulting model explained 81% of the variance in damage ratings (conditional R^2). This result supports our hypothesis that the iVR experience would increase perceived risk (as

$$\begin{aligned} Rating_i &\sim N\alpha_{j[i]} + NA(Color_Y) + NA(Color_O) + NA(Color_R) + NA(Distance_{Near}), \sigma^2 \\ \alpha_j &\sim N\gamma_0^\alpha + \gamma_1^\alpha(Condition_{iVR}) + \gamma_2^\alpha(Order_{Real\_First}) + \gamma_3^\alpha(VR\_Exp_{Yes}) + \gamma_4^\alpha(GamingTotal) + \gamma_5^\alpha(Condition_{iVR} \times Order_{Real\_First}), \sigma^2_{\alpha_j}, \text{ for ID } j = 1, \ldots, J \end{aligned} \quad (1)$$

measured through damage ratings), suggesting higher immersion increases perceived risk when interpreting maps.

Distance (*near* or *far*) was included to account for the distance-from-shore effect found in Mason (2018). Other iVR studies have shown that previous experience with iVR systems and 3D visualization tools (Slater & Sanchez-Vives, 2016) and extensive video game experience (Lages & Bowman, 2018) can influence results. The iVR experience question consisted of a yes/no question asking if the participant had every used an iVR system with several examples (e.g., "Oculus Rift"). The game experience questionnaire was adapted from Oprean et al. (2017) and consisted of six questions asking the participant to rate their familiarity on a scale of 1–9 with 3D game genres and other 3D software, with the mean of responses used as the predictor variable. In the body of the paper, we will not report the results from the distance, iVR experience, and video game experience but they are available in the supplemental material and linked Open Science Repository. All the results reported in the following section do, however, control for the effects of those covariates.

To evaluate our hypothesis (H1) that iVR participants will have

3.2. Confidence ratings

In order to assess whether higher levels of immersion between condition affected risk perception in terms of decreasing perceived uncertainty (H2), we examined confidence ratings associated with damage estimates. We computed a multi-level model which used the interaction between condition * color, along with lower order terms to predict confidence (see Equation (2)). The intercepts for each participant were allowed to vary. The model explained 57% of the variance in damage ratings (conditional R^2).

$$\begin{aligned} Confidence_i &\sim N\mu, \sigma^2 \\ \mu &= \alpha_{j[i]} + \beta_1(Color_Y) + \beta_2(Color_O) + \beta_3(Color_R) \\ \alpha_j &\sim N\gamma_0^\alpha + \gamma_1^\alpha(Condition_{iVR}) + \gamma_2^\alpha(Color_Y \times Condition_{iVR}) + \gamma_3^\alpha(Color_O \times Condition_{iVR}) + \gamma_4^\alpha(Color_R \times Condition_{iVR}), \sigma^2_{\alpha_j}, \text{ for ID } j = 1, \ldots, J \end{aligned} \quad (2)$$

This analysis revealed that there was no evidence for an interaction between condition * color, or a main effect of condition, i.e., we found no support for H2. But there was a main effect of color, such that

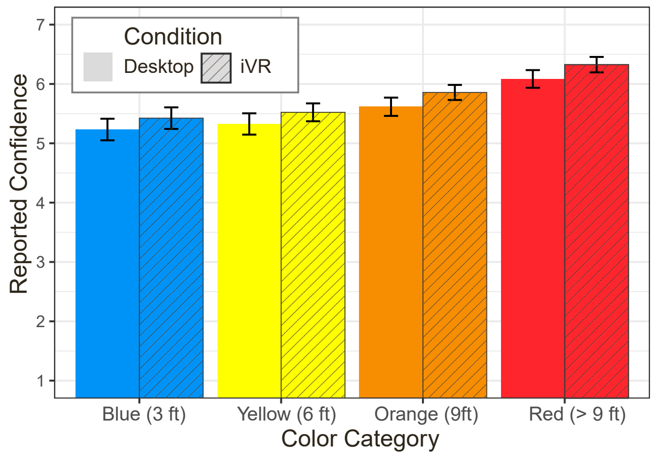

Fig. 7. Damage rating (on a scale of 1–7) by color category and condition. Error bars represent 95% confidence intervals around mean damage ratings. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

individuals were more confident in their damage ratings when judging the orange ($b = 0.38$, $t = 4.98$, $p < .001$, 95% *CI* [0.23, 0.54]) and red regions ($b = 0.85$, $t = 11.05$, $p < .001$, 95% *CI* [0.70, 1.00]) compared to the blue region (see Fig. 6).

To evaluate our hypothesis that flood-height category would predict map damage ratings (H3), we included a main effect of color (which represents the different flood height categories) in the damage-rating model. We found evidence to suggest that individuals' damage ratings increased relative to the heights indicated by the colored height categories as seen in Fig. 7. In our model, the blue category was coded as the referent ($M = 2.48$, $SD = 1.06$) and damage ratings increased significantly compared to blue; yellow ($M = 3.8$, $SD = 0.88$, $b = 1.32$, $t = 25.58$, $p < .001$, 95% *CI* [1.22, 1.42]), orange ($M = 5.29$, $SD = 0.88$, $b = 2.81$, $t = 54.28$, $p < .001$, 95% *CI* [2.71, 2.91]), and red ($M = 6.56$, $SD = 0.77$, $b = 4.08$, $t = 78.81$, $p < .001$, 95% *CI* [3.97, 4.18]). To confirm that the significance of color did not change with different referent colors, we re-ran the model with yellow, orange, and red color codes as the referent (reported in supplemental material), and found no changes to significance. For the red category, we observed a larger difference between the mean damage rating for desktop ($M = 6.45$, $SD = 0.92$) and iVR conditions ($M = 6.67$, $SD = 0.55$), which is apparent in Fig. 7. In a post-hoc exploratory linear regression analysis using only the red trials, we found evidence that the iVR condition elicited statistically significantly higher damage ratings than the desktop ($b = 0.22$, $t = 2.97$, $p = .002$, 95% *CI* [0.07, 0.36]).

3.3. Height estimates

To evaluate our hypothesis (H4) that participants in the iVR condition would perform more accurately in the real-world height estimation task, we performed a linear regression predicting error by condition (taking the form: $HeightError = \alpha + \beta_1(Condition_{iVR}) + \varepsilon$). We found participants in the iVR group ($MeanError = 3.03$ cm, $SD = 8.29$) had statistically significantly better real-world flood height estimates compared to the desktop group ($MeanError = 4.79$ cm, $SD = 11.25$, $b = -1.76$, $t = -3.72$, $p < .001$, 95% *CI* [−2.69, −0.83]). See Fig. 8 for a visualization of these results. This result supports H4, that participants in the iVR condition would be better able to point to flood heights in the real world due to superior spatial learning.

3.4. Condition and task-order interaction

To evaluate our hypothesis that there would be an interaction between condition and order (H5), we included an interaction term between condition and order in the map damage-rating model. We did not find statistically significant support for the interaction ($b = -0.37$, $t = -1.61$, $p = .107$, 95% *CI* [−0.83, 0.08]). We elected to investigate potential differences between conditions within each order set (real-world task first or map task first), and we ran the same model as before (Equation (1)) but on each order set separately. For the individuals that completed the map first (i.e., without having done the real-world height task), there was a main effect of iVR ($b = 0.33$, $t = 3.23$, $p = .001$, 95% *CI* [0.13, 0.53]), where the iVR condition ($M = 4.52$, $SD = 1.77$) elicited higher damage ratings than the desktop condition ($M = 4.16$, $SD = 1.86$) as seen in Fig. 9. In contrast, when we conducted the same analysis with the real-world task first participants, there was not a significant main effect of iVR ($b = 0.01$, $t = -0.30$, $p = .975$, 95% *CI* [−0.4, 0.42]).

3.5. Evacuation decisions

We ran two exploratory linear models (taking the form: $EvacuationHeight = \alpha + \beta_1(Condition_{iVR}) + \varepsilon$) using condition to predict the two evacuation decisions (the height at which they would personally evacuate, and the height at which they would order an evacuation). We expected that the iVR experience would lead partici-

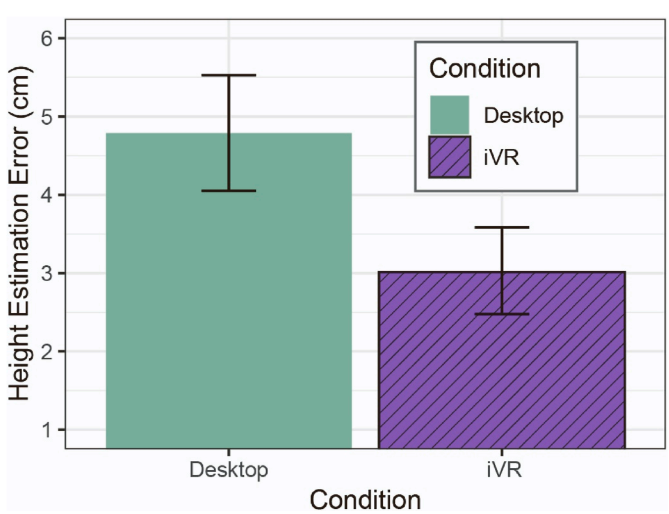

Fig. 8. Average height error from the real-world height task. Error bars represent 95% confidence intervals around mean damage ratings.

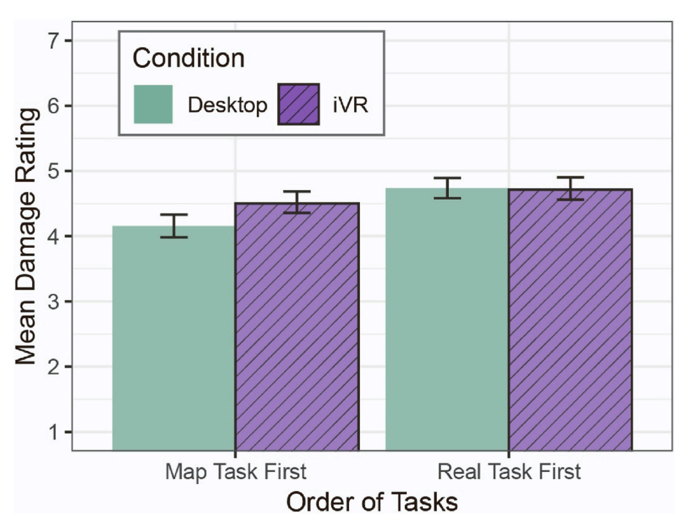

Fig. 9. Damage rating by order (real-world height task first or map damage-rating task first) and condition (iVR or desktop). Error bars represent 95% confidence intervals around mean damage ratings.

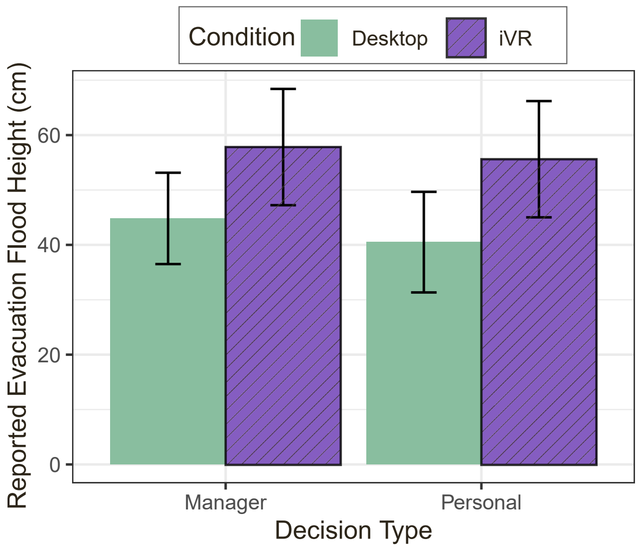

Fig. 10. Reported flood heights for the two evacuation decisions posed to participants (manager and personal) by condition (iVR or desktop). Error bars represent 95% confidence intervals around mean damage ratings.

Table 1
Results of Wilcoxon Rank Sum tests comparing GEQ affective responses between conditions. Null hypotheses are that there were no differences between condition in median response. p-values are adjusted using the Holm-Bonferroni method for multiple testing comparison.

Wilcoxon Rank Sum Tests of GEQ Response

	Median Value		*W*	*p*	*95% CI*
	Desktop	iVR			
SI Immersion	2.83	3.67	189.0	***.007***	[-1.00, −0.33]
Flow	2.20	2.80	153.5	***.001***	[-1.20 -0.40]
Tension	0.0	0.0	450.0	*.387*	[-0.00, −0.00]
Negative Affect	0.50	0.25	504.0	*.239*	[0.00, 0.00]
Positive Affect	0.25	3.0	251.0	*.126*	[-1.20, −0.40]
Negative Experience	0.17	0.0	480.0	*.387*	[-1.20, −0.20]
Positive Experience	0.67	1.17	234.0	*.126*	[0.00, 0.00]
Tiredness	0.0	0.0	350.0	*.444*	0.00. −0.00]
Diff. Return. Real.	0.67	1.33	211.5	***.020***	[-1.00 -0.33]

pants to responding that they would evacuate and evacuate others at lower water heights. Instead, individuals in the iVR condition reported ordering a hypothetical evacuation at significantly higher water heights ($M = 55.73$ cm, $SD = 24.99$) compared to the participants in the desktop group ($M = 44.82$ cm, $SD = 21.09$, $b = 10.91$, $t = 9.91$, $p < .001$, 95% *CI* [8.75, 13.07]). Individuals in the iVR condition also reported that they would personally evacuate at significantly higher water heights ($M = 53.06$ cm, $SD = 23.77$) compared to the desktop group ($M = 40.5$ cm, $SD = 23.22$, $b = 12.56$, $t = 11.21$, $p < .001$, 95% *CI* [10.37, 14.76]). The disparity between responses between conditions for both questions can be seen in Fig. 10. Assuming that evacuating at higher water heights is associated with *lower* perceived risk, the effect of condition appears to be inverted for this measure compared to the map damage-rating task.

3.6. *Affective measures*

We collected self-reported measures related to experience using a modified GEQ questionnaire (IJsselsteijn et al., 2013): SI Immersion, Flow, Tension, Negative Affect, Positive Affect, Positive Experience, Negative Experience, Tiredness, and (difficulty) Returning to Reality. As an exploratory post-hoc analysis, we ran Wilcoxon Rank Sum tests comparing the affective responses between conditions (see Table 1). One iVR participant was dropped in the analysis of Flow due to missing data (non-responses), as were two participants in the analysis of Positive Experience (one in each condition) for the same reason. iVR participants reported having significantly higher SI Immersion, Flow, and Difficulty Returning to Reality. However, as can bee seen in Fig. 11, many of the measures have large standard deviations.

4. Discussion

Our results show that experiencing an iVR flooding simulation increases the map damage ratings of participants more than the same experience with a traditional, non-immersive, desktop display. However, participants who had an iVR experience also reported that they would evacuate and evacuate others at higher (i.e., more dangerous) flood levels. We found that iVR increases participants' ability to accurately point to flood heights in a real environment, and that both experiences led participants to formulate their responses using flood height information on maps more than in previous studies by Mason (2018). However, we did not find the hypothesized interaction between the presented order of the tasks and condition, although a follow-up exploratory analysis revealed a potential effect.

Regarding perceived risk, the effect of the conditions depends on the measure used. We found direct support for our hypothesis (H1) that participants who experienced the iVR version of the flooding simulation had statistically significantly higher map damage ratings (our primary measure of perceived risk). Participants also reported being similarly confident about their damage ratings in both conditions, against our expectations that the higher level of immersion of the iVR condition would lead to more confidence in the later damage ratings (H2). However, iVR had an unexpected impact on the pair of notional evacuation decisions. iVR participants tended to rate damage as higher on maps, but iVR participants also indicated that they would evacuate or order an evacuation at higher (i.e., more dangerous) flood heights. Thus, our map-based perceived-risk measures and our evacuation decision-based measure seem to be triggering different risk perception processes. This of practical concern, as evacuations decisions often mirror real-world responses Huang et al. (2016), and implies that adding an iVR experience for hazard communication may be counterproductive if the goal is to increase risk perception. However, in some situations reducing risk perception is desired. For example, *shadow evacuation*, the evacuation of people who have not been instructed to, has been an ongoing challenge in evacuation planning (Dash & Gladwin, 2007). iVR in such cases may then serve to more appropriately "calibrate" risk perception.

The differing responses to the map and evacuation-based perceived risk measures may be linked to the relative abstractness of the map task and the more embodied nature of the evacuation decisions made in a real environment. iVR participants may have more fully appreciated the potential *environmental* harm, but having survived a realistic-seeming flood unscathed, associated it with less *bodily* harm. iVR participants may have been recalling the physically innocuous iVR representation of the flooding when making their evacuation decisions, rather than creating their own conception of a real flood event.

The lengthy exposure to deadly conditions (in the form of virtual submergence) may have led to a form of *extinction*. Extinction is the reduction in fear response due to repeated harmless exposure to a previously-learned harm-associated stimulus (Dunsmoor et al., 2014). Our system may have inadvertently led to an extinction response that in this context led to the unexpected evacuation responses. This effect has been harnessed for therapeutic purposes, and viewing realistic-*appearing* but non-real perceived threats in iVR has been used successfully in the treatment of phobias (Botella et al., 2017; Maples-Keller et al., 2017). There is evidence that repeated exposure to a virtual threat stimulus is required for long-lasting extinction as desired in a therapeutic context (Dunsmoor et al., 2014), so the effect of our environment on bodily risk perception is likely transitory, but this is unknowable without further investigation. For example, there is strong indication that exposure to

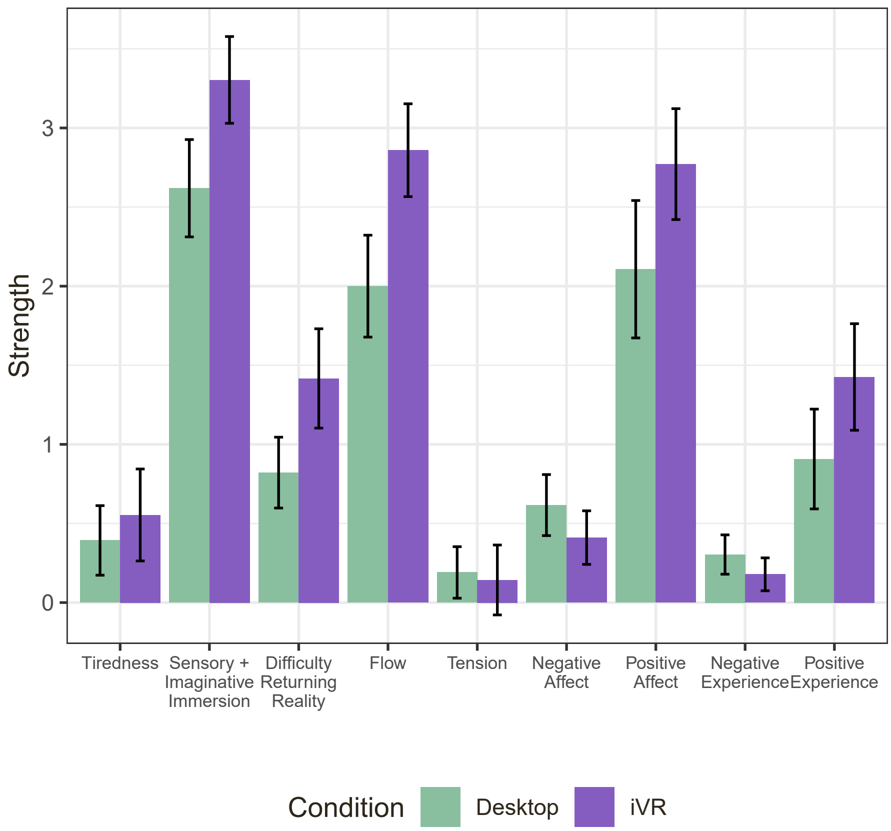

Fig. 11. Plot of Game Experience Questionnaire (GEQ) affective responses by condition and dimension. Error bars represent 95% confidence intervals around mean response value.

moderate adversity can lead to greater psychological resilience (Seery et al., 2010). Regardless of source, this effect should be further investigated before recommending iVR for hazard risk-communication applications. In terms of the measures, it should be noted that each participant made only these 2 evacuation decisions while they each made 32 damage ratings, and evacuation decisions are fraught with potential confounds as discussed in the section *Risk Perception and Evacuation Decision-Making.*.

We found iVR helped participants understand (i.e., accurately point to) the physical dimensions of flood heights. iVR participants were more successful in gaining correct spatial knowledge of the hazard (i.e., learning its true size) and applying that knowledge to the real world. However, such understanding does not appear to be closely linked to risk perception. We found support for our hypothesis, H5, that participants would be able to more accurately point to flood heights in the real-world. This was despite both conditions performing surprisingly well at the height estimation task, with a mean error of less than 5 cm for both groups (see Fig. 8). In addition, these results show that, in our study, scale mis-perception common of immersive virtual environments experienced through previous generations of hardware (Kelly et al., 2017) was not an issue. Moreover, the apparent ease of the height task for all participants indicates the task may be subject to a ceiling effect. The universally high accuracy of participants seems to indicate that any lack of recognition of flood risk in the general population has little to do with poor understanding of flood height values alone.

When examining all possible orders and conditions together, we did not find the hypothesized interaction effect between condition and task order (real-height estimation first or map damage-rating task first). However, post-hoc exploratory analyses revealed differences in damage ratings between conditions in the map-task first group; when participants completed the map damage ratings *before* before the real-world height task, the condition appears to have affected responses. When participants completed the map damage ratings *after* the real-world height, the difference between conditions is absent (see Fig. 9). Since the real-world height task lacked the same effect on iVR participants, iVR participants may have already had equivalent knowledge to what desktop participants gained in performing the real-world task (e.g., a relatively strong understanding of flood heights in the real world). If this is true, it may indicate that non-iVR interventions, such as having participants imagine a space being flooded, may be a low-cost alternative to an iVR experience for increasing flood-risk perception. However, given the apparent non-significance of this interaction in the model including both conditions, this analysis is still somewhat speculative. Even so, we would encourage further studies to examine this question as scale perception is a potentially important advantage of iVR representations (Barba & Marroquin, 2017).

The main effect of color (flood-height map category) indicates that viewing our storm-surge simulation effectively cues participants to pay attention to height categories on NHC-style potential storm-surge maps. This is in contrast to the findings of Mason (2018) who found with a similar map task that participants did not use height category as a main factor in decision-making. Overall there was no statistically significant difference between conditions in damage ratings within the same color categories. Post-hoc exploratory analysis of the red category alone did reveal a significantly higher confidence iVR participants, mirroring the findings of (Zanola et al., 2009), but this difference was relatively small. Analysis of the confidence ratings revealed a similar level of confidence in provided damage ratings for the 3 ft (blue) and 6 ft (yellow) categories, but higher confidence ratings for estimations for the 9 ft (orange) and greater-than-9ft (red) categories, with participants being most confident for red-category ratings. Experiencing a flood simulation, regardless of level of technological immersion, appears to have attenuated the distance effect/heuristic found by Mason (2018). This mirrors the results of Houston et al. (2017) who found that using either of the tested web maps appeared to eliminate differences in perceived flood risk between some user groups. In a follow-up analysis, we did find evidence that interpretation of the highest level of flooding was different between conditions, perhaps indicating some increase in perceived risk in iVR participants when the water surface was high overhead, but this effect was not large. However, these findings may indicate desktop flood simulations of flooding alone may help participants interpret map information, and iVR may not be strictly necessary for this purpose, echoing Lindner and Alsheimer (2019), who found that a desktop experience can improve storm-surge knowledge.

4.1. The role of affective responses

Our post-hoc exploratory analysis of the GEQ affective responses suggests that iVR led to greater reported SI Immersion, Flow, and Difficulty Returning to Reality. Increased involvement with the experience as measured through those constructs may have made it more difficult to maintain awareness that the lack of impact on their own body was not realistic, despite recognizing the potential impact of flooding on the environment. These results may help explain why iVR participants had lower perceived risk (as measured through the evacuation decisions). However, we regard our affective analysis as exploratory, and we recommend future studies should evaluate GEQ alternatives due to potential validity issues identified by Law et al. (2018). A potentially fruitful future research prospect is to address interactions between immersion, affective responses, and different measures of perceived risk. For example, measuring presence more directly through instruments such as the spatial presence experience scale (Hartmann et al., 2016) may provide greater insight into the degree to which participants felt to be part of the environment.

4.2. Limitations and open questions

There are several important limitations of this study, as well as open questions it highlights. Our participant pool pulled from a university community not necessarily reflective of a general population, and was located from an inland area not typically prone to storm-surge flooding. Prior experience with hazards can strongly influence risk perception (Morss et al., 2018; Wachinger et al., 2013), and there may be complex interactions with participants who have prior experience with actual hurricane evacuation and/or storm-surge flooding. Likewise, many individual factors, such as education and age, can influence risk perception (Dash & Gladwin, 2007; Wachinger et al., 2013). Future studies may also investigate, 1) the disconnect between perceived bodily and environmental risk in immersive environments and inform hazard-communication applications, using a wider variety of risk perception measures (e.g., selected from those reviewed and presented in Wilson et al., 2019), 2) other hazard types, 3) other methods of depicting dangerous hazard conditions, and 4) different dimensions of risk perception. For example, future studies could examine whether iVR's (apparent) low impact on behavior in a hazards context is due to the dual effect of increased risk perception but also increased belief in one's own ability to survive the hazard. It may be that an underwater virtual experience is by its very nature relaxing (Yeo et al., 2020). However, by providing a counter-example to their own immunity, such as the presence of simulated flood victims, we speculate the reduction of personal risk perception may be attenuated. Multidimensional perceived-risk measures, such as those in Houston et al. (2017), may reveal more complex interactions and causal relationships between different aspects of risk perception. There were also more subtle technological differences between conditions whose influence is unknown; for example, desktop participants were required to hold a game controller throughout the experience, and used a different type of audio device.

5. Conclusions

In performing this study, we iteratively developed an immersive experience of storm-surge flooding with input from domain experts, and found that giving people an iVR experience of flooding may have unintended and counterproductive effects on risk perception for risk communication applications. Our iVR experience effectively cued participants to the flood heights indicated on notional storm-surge maps, and increased their accuracy when estimating flood heights in the real world. However, the impact on individual risk perception is less clear and appears to depend on how it is measured. Participants who experienced flooding in iVR responded that there would be more damage (interpreted as greater risk) for a given level of flooding on a series of hypothetical storm-surge maps. However, in notional storm-surge scenarios in a real environment, participants in the iVR condition indicated that they would personally evacuate and order others to evacuate at greater flood heights. This is a concerning finding if an iVR flood simulation is intended to increase evacuation responses, and may indicate a more general challenge when attempting to communicate danger using iVR. Overall, this study reveals much remains unknown regarding iVR as a tool for risk communication, and illustrates that iVR appears prone to counter-intuitive effects on users in the context of hazards. Designers aiming to expose participants to danger in order to communicate associates risks should be wary of giving participants the false impression that hazards are more or less damaging than they really are. We also sought to understand potential causal mechanisms behind iVR's influence, and found that cuing participants explicitly to true-scale flood heights (either in the real-world or in iVR) appeared to play an important role in increasing perceived risk. The role of affective responses was less clear, and requires further investigation as the counter-intuitive results of evacuation decisions appeared to be partly attributable to such responses.

Credit author statement

Mark Simpson: Conceptualization, Investigation, Methodology, Software, Writing – Original Draft.

Lace Padilla: Conceptualization, Methodology, Formal Analysis, Writing – Review and Editing, Visualization.

Klaus Keller: Conceptualization, Funding acquisition, Writing – Review and Editing.

Alex Klippel: Supervision, Conceptualization, Resources, Writing -Review and Editing.

Declaration of competing interest

None.

Acknowledgements

We thank Jennifer Mason for her invaluable expertise as well as Caitlin Spence for inputs. This work was co-supported by the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 and by the Penn State Center for Climate Risk Management. Any opinions, findings, potential errors, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or other funding entities.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvp.2022.101764.

References

Bailey, J. H., & Witmer, B. G. (1994). Learning and transfer of spatial knowledge in a virtual environment. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting, 38*, 1158–1162. https://doi.org/10.1177/154193129403801803. http://journals.sagepub.com/doi/10.1177/154193129403801803

Barba, E., & Marroquin, R. Z. (2017). A primer on spatial scale and its application to mixed reality. In W. Broll, H. Regenbrecht, & J. E. Swan (Eds.), *2017 IEEE international Symposium on Mixed and augmented reality* (pp. 100–110). Piscataway, NJ: IEEE. https://doi.org/10.1109/ISMAR.2017.27.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01. URL http://www.jstatsoft.org/v67/i01/.

Bernhardt, J., Snelllings, J., Smiros, A., Bermejo, I., Rienzo, A., & Swan, C. (2019). Communicating hurricane risk with virtual reality: A pilot project. *Bulletin of the American Meteorological Society, 100*(10), 1897–1902. https://doi.org/10.1175/BAMS-D-17-0326.1

Bostrom, A., Morss, R., Lazo, J. K., Demuth, J., & Lazrus, H. (2018). Eyeing the storm: How residents of coastal Florida see hurricane forecasts and warnings. *International Journal of Disaster Risk Reduction, 30*, 105–119. https://doi.org/10.1016/j.ijdrr.2018.02.027

Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., & Baños, R. (2017). Recent progress in virtual reality exposure therapy for phobias: A systematic review. *Current Psychiatry Reports, 19*, 1–13. https://doi.org/10.1007/s11920-017-0788-4

Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: How much immersion is enough? *Computer, 40*, 36–43. https://doi.org/10.1109/MC.2007.257

Bubeck, P., Botzen, W. J. W., & Aerts, J. C. J. H. (2012). A review of risk perceptions and other factors that influence flood mitigation behavior. *Risk Analysis: An International Journal, 32*(9), 1481–1495. https://doi.org/10.1111/j.1539-6924.2011.01783.x

Burnside, R., Miller, D. S., & Rivera, J. D. (2007). The impact of information and risk perception on the hurricane evacuation decision-making of greater new orleans residents. *Sociological Spectrum, 27*, 727–740. https://doi.org/10.1080/02732170701534226. URL http://www.tandfonline.com/doi/abs/10.1080/02732170701534226.

Camelo, J., & Mayo, T. (2021). The lasting impacts of the saffir-simpson hurricane wind scale on storm surge risk communication: The need for multidisciplinary research in addressing a multidisciplinary challenge. *Weather and Climate Extremes, 33*, 100335. https://doi.org/10.1016/j.wace.2021.100335

Cuite, C. L., Shwom, R. L., Hallman, W. K., Morss, R. E., & Demuth, J. L. (2017). Improving coastal storm evacuation messages. *Weather, Climate, and Society, 9*, 155–170. https://doi.org/10.1175/WCAS-D-16-0076.1

Cummings, J. J., & Bailenson, J. N. (2014). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology, 19*, 272–309. https://doi.org/10.1080/15213269.2015.1015740

Dash, N., & Gladwin, H. (2007). Evacuation decision making and behavioral responses: Individual and household. *Natural Hazards Review, 8*, 69–77. https://doi.org/10.1061/(ASCE)1527-6988(2007)8:3(69). URL http://ascelibrary.org/doi/10.1061/%28ASCE%291527-6988%282007%298%3A3%2869%29.

Dixon, G. N. (2016). Negative affect as a mechanism of exemplification effects, an experiment on two-sided risk argument recall and risk perception. *Communication Research, 43*, 761–784. https://doi.org/10.1177/0093650215579222. URL http://journals.sagepub.com/doi/10.1177/0093650215579222.

Dixon, M. W., Wraga, M., Proffitt, D. R., & Williams, G. C. (2000). Eye height scaling of absolute size in immersive and nonimmersive displays. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 582–593. https://doi.org/10.1037/0096-1523.26.2.582

Dunsmoor, J. E., Ahs, F., Zielinski, D. J., & LaBar, K. S. (2014). Extinction in multiple virtual reality contexts diminishes fear reinstatement in humans. *Neurobiology of Learning and Memory, 113*, 157–164. https://doi.org/10.1016/j.nlm.2014.02.010

Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education state of the art and perspectives. *The International Scientific Conference eLearning and Software for Education, 1*, 133–141. URL https://ezaccess.libraries.psu.eduloginurl=httpssearch-proquest-com.ezaccess.libraries.psu.edudocview1681252932accountid=13158.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel. Analytical methods for social research.* Cambridge: Cambridge University Press. URL http://www.loc.gov/catdir/enhancements/fy0668/2006040566-d.html.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Grand View Research. (2021). Virtual reality market size, share & trends analysis report by technology (semi & fully immersive, non-immersive), by device (hmd, gtd), by component (hardware, software), by application, and segment forecasts, 2021-2028. URL https://www.grandviewresearch.com/industry-analysis/virtual-reality-vr-market.

Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., Böcking, S., Ravaja, N., Laarni, J., Saari, T., Gouveia, F., & Maria Sacau, A. (2016). The spatial presence experience scale (spes). *Journal of Media Psychology, 28*, 1–15. https://doi.org/10.1027/1864-1105/a000137

Hou, J., Nam, Y., Peng, W., & Lee, K. M. (2012). Effects of screen size, viewing angle, and player's immersion tendencies on game experience. *Computers in Human Behavior, 28*, 617–623. https://doi.org/10.1016/j.chb.2011.11.007. URL https://linkinghub.elsevier.com/retrieve/pii/S0747563211002512.

Houston, D., Cheung, W., Basolo, V., Feldman, D., Matthew, R., Sanders, B. F., Karlin, B., Schubert, J. E., Goodrich, K. A., Contreras, S., & Luke, A. (2017). The influence of hazard maps and trust of flood controls on coastal flood spatial awareness and risk perception. *Environment and Behavior, 51*, 347–375. https://doi.org/10.1177/0013916517748711

HTC Corporation. (2019). Vive^{TM} — vive virtual reality system. URL https://www.vive.com/us/product/vive-virtual-reality-system/.

Huang, S.-K., Lindell, M. K., & Prater, C. S. (2016). Who leaves and who stays? A review and statistical meta-analysis of hurricane evacuation studies. *Environment and Behavior, 48*, 991–1029. https://doi.org/10.1177/0013916515578485. URL http://journals.sagepub.com/doi/10.1177/0013916515578485.

IJsselsteijn, W., De Kort, Y., & Poels, K. (2013). The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven, 46*.

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies, 23*, 1515–1529. https://doi.org/10.1007/s10639-017-9676-0

Johnson, D., Gardner, M. J., & Perry, R. (2018). Validation of two game experience scales: The player experience of need satisfaction (pens) and game experience questionnaire (geq). *International Journal of Human-Computer Studies, 118*, 38–46. https://doi.org/10.1016/j.ijhcs.2018.05.003

Jonkman, S. N., & Vrijling, J. K. (2008). Loss of life due to floods. *Journal of Flood Risk Management, 1*, 43–56. https://doi.org/10.1111/j.1753-318X.2008.00006.x

Kahneman, D., Slovic, P., & Tversky, A. (1982). In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kellens, W., Zaalberg, R., & de Maeyer, P. (2012). The informed society: An analysis of the public's information-seeking behavior regarding coastal flood risks. *Risk Analysis : An Official Publication of the Society for Risk Analysis, 32*, 1369–1381. https://doi.org/10.1111/j.1539-6924.2011.01743.x

Kelly, J. W., Cherep, L. A., & Siegel, Z. D. (2017). Perceived space in the htc vive. *ACM Transactions on Applied Perception, 15*, 1–16. https://doi.org/10.1145/3106155

Kostelnick, J. C., McDermott, D., Rowley, R. J., & Bunnyfield, N. (2013). A cartographic framework for visualizing risk. *Cartographica: The International Journal for Geographic Information and Geovisualization, 48*, 200–224. https://doi.org/10.3138/carto.48.3.1531

Lages, W. S., & Bowman, D. A. (2018). Move the object or move myself? Walking vs. manipulation for the examination of 3d scientific data. *Frontiers in ICT, 5*, 1–12. https://doi.org/10.3389/fict.2018.00015

Laurell, C., Sandström, C., Berthold, A., & Larsson, D. (2019). Exploring barriers to adoption of virtual reality through social media analytics and machine learning – an assessment of technology, network, price and trialability. *Journal of Business Research, 100*, 469–474. https://doi.org/10.1016/j.jbusres.2019.01.017

Law, E. L.-C., Brühlmann, F., & Mekler, E. D. (2018). Systematic review and validation of the game experience questionnaire (geq) - implications for citation and reporting practice. In *Proceedings of the 2018 annual symposium on computer-human interaction in play CHI PLAY '18* (pp. 257–270). New York, NY, USA: Association for Computing Machinery.

Leyrer, M., Linkenauger, S. A., Bülthoff, H. H., Kloos, U., & Mohler, B. (2011). The influence of eye height and avatars on egocentric distance estimates in immersive virtual environments. In R. McDonnell, S. Thorpe, D. Gutierrez, M. Giese, & S. N. Spencer (Eds.), *Proceedings of the ACM SIGGRAPH symposium on applied perception in graphics and visualization - APGV '11* (p. 67). New York, New York, USA: ACM Press. https://doi.org/10.1145/2077451.2077464.

Lindner, B. L., & Alsheimer, F. (2019). Assessing improvement in the public's understanding of hurricane storm tides through interactive visualization models. *Journal of Coastal Research, 35*, 130. https://doi.org/10.2112/JCOASTRES-D-17-00223.1

Liu, L., Boone, A. P., Ruginski, I. T., Padilla, L., Hegarty, M., Creem-Regehr, S. H., Thompson, W. B., Yuksel, C., & House, D. H. (2016). Uncertainty visualization by representative sampling from prediction ensembles. *IEEE Transactions on Visualization and Computer Graphics, 23*, 2165–2178.

Liu, L., Padilla, L., Creem-Regehr, S. H., & House, D. H. (2018). Visualizing uncertain tropical cyclone predictions using representative samples from ensembles of forecast tracks. *IEEE Transactions on Visualization and Computer Graphics, 25*, 882–891.

Loyola, M. (2018). The influence of the availability of visual cues on the accurate perception of spatial dimensions in architectural virtual environments. *Virtual Reality, 22*, 235–243. https://doi.org/10.1007/s10055-017-0331-2

Maples-Keller, J. L., Yasinski, C., Manjin, N., & Rothbaum, B. O. (2017). Virtual reality-enhanced extinction of phobias and post-traumatic stress. *Neurotherapeutics : The Journal of the American Society for Experimental NeuroTherapeutics, 14*, 554–563. https://doi.org/10.1007/s13311-017-0534-y

Mason, J. S. (2018). *Visual perspectives and decision-making in geospatial uncertainty*. PA: The Pennsylvania State University University Park. Doctor of philosophy dissertation.

McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2014). Stereoscopic 3D displays and human performance: A comprehensive review. *Displays, 35*, 18–26. https://doi.org/10.1016/j.displa.2013.10.004

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033. URL https://linkinghub.elsevier.com/retrieve/pii/S0360131513002108.

Montello, D. R. (1993). Scale and multiple psychologies of space. In I. Campari, & A. U. Frank (Eds.), *Spatial information theory* (pp. 312–321). Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-57207-4_21. Berlin Heidelberg volume 716 of *Lecture notes in computer science Spatial information theory*.

Morrow, B. H., Lazo, J. K., Rhome, J., & Feyen, J. (2015). Improving storm surge risk communication: Stakeholder perspectives. *Bulletin of the American Meteorological Society, 96*, 35–48. https://doi.org/10.1175/BAMS-D-13-00197.1

Morss, R. E., Cuite, C. L., Demuth, J. L., Hallman, W. K., & Shwom, R. L. (2018). Is storm surge scary? The influence of hazard, impact, and fear-based messages and individual differences on responses to hurricane risks in the USA. *International Journal of Disaster Risk Reduction, 30*, 44–58. https://doi.org/10.1016/j.ijdrr.2018.01.023

National Hurricane Center. (2019). National storm surge hazard maps [Online]. URL https://noaa.maps.arcgis.com/apps/MapSeries/index.html?appid=d9ed7904dbec441a9c4dd7b277935fad.

Newcombe, N. S. (2018). Three kinds of spatial cognition. In *Stevens' handbook of experimental psychology and cognitive neuroscience* (pp. 1–31). Wiley Online Library.

Oprean, D., Simpson, M., & Klippel, A. (2017). Collaborating remotely: An evaluation of immersive capabilities on spatial experiences and team membership. *International Journal of Digital Earth, 11*, 420–436. https://doi.org/10.1080/17538947.2017.1381191

Padilla, L. M., Ruginski, I. T., & Creem-Regehr, S. H. (2017). Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive research: Principles and Implications, 2*(1), 1–16. https://doi.org/10.1186/s41235-017-0076-1

Qualtrics. (2019). Qualtrics [Online]. URL https://www.qualtrics.com.

Rappaport, E. N. (2014). Fatalities in the United States from atlantic tropical cyclones: New data and interpretation. *Bulletin of the American Meteorological Society, 95*, 341–346. https://doi.org/10.1175/BAMS-D-12-00074.1

Rapp, D. N., Culpepper, S. A., Kirkby, K., & Morin, P. (2007). Fostering students' comprehension of topographic maps. *Journal of Geoscience Education, 55*, 5–16. https://doi.org/10.5408/1089-9995-55.1.5

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis Methods volume 1 of Advanced quantitative techniques in the social sciences* (2nd ed.). Thousand Oaks, CA and London: Sage Publications.

Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013). The perception of egocentric distances in virtual environments - a review. *ACM Computing Surveys, 46*, 1–40. https://doi.org/10.1145/2543581.2543590

Reyes, M. E. P., & Chen, S.-C. (2017). A 3D virtual environment for storm surge flooding animation. In *2017 IEEE third international conference on multimedia big data* (pp. 244–245). Piscataway, NJ: IEEE.

Riad, J. K., Norris, F. H., & Ruback, R. B. (1999). Predicting evacuation in two major disasters: Risk perception, social influence, and access to resources. *Journal of Applied Social Psychology, 29*, 918–934. https://doi.org/10.1111/j.1559-1816.1999.tb00132.x

Ruginski, I. T., Boone, A. P., Padilla, L. M., Liu, Le, Heydari, N., Kramer, H. S., Hegarty, M., Thompson, W. B., House, D. H., & Creem-Regehr, S. H. (2016). Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition and Computation, 16*, 154–172. https://doi.org/10.1080/13875868.2015.1137577

Seery, M. D., Holman, E. A., & Silver, R. C. (2010). Whatever does not kill us: Cumulative lifetime adversity, vulnerability, and resilience. *Journal of Personality and Social Psychology, 99*, 1025–1041. https://doi.org/10.1037/a0021344

Sherman-Morris, K., Antonelli, K. B., & Williams, C. C. (2015). Measuring the effectiveness of the graphical communication of hurricane storm surge threat. *Weather, Climate, and Society, 7*, 69–82. https://doi.org/10.1175/WCAS-D-13-00073.1

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence: Teleoperators and Virtual Environments, 8*, 560–565. https://doi.org/10.1162/105474699566477

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI, 3*, 74. https://doi.org/10.3389/frobt.2016.00074

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research, 177*, 1333–1352. https://doi.org/10.1016/j.ejor.2005.04.006

Strahler, A. H., & Strahler, A. N. (2005). *Physical geography: Science and systems of the human environment* (3rd ed.). New York: J. Wiley.

Sundar, S. S. (2009). Media effects 2.0: Social and psychological effects of communication technologies. In R. L. Nabi, & M. B. Oliver (Eds.), *The SAGE handbook of media processes and effects* (pp. 545–560). Los Angeles: SAGE.

Sundar, S. S., Kang, J., & Oprean, D. (2017). Being there in the midst of the story: How immersive journalism affects our perceptions and cognitions. *Cyberpsychology, Behavior, and Social Networking, 20*, 672–682. https://doi.org/10.1089/cyber.2017.0271

The National Hurricane Center. (2017). Storm surge can be deadly: 10 tips to be ready [Online]. URL https://www.nhc.noaa.gov/surge/StormSurgeCanBeDeadly10tips-single.pdf.

Trumbo, C. W., Peek, L., Meyer, M. A., Marlatt, H. L., Gruntfest, E., McNoldy, B. D., & Schubert, W. H. (2016). A cognitive-affective scale for hurricane risk perception. *Risk Analysis : An Official Publication of the Society for Risk Analysis, 36*, 2233–2246. https://doi.org/10.1111/risa.12575

Twedt, E., Crawford, L. E., & Proffitt, D. R. (2012). Memory for target height is scaled to observer height. *Memory & Cognition, 40*, 339–351. https://doi.org/10.3758/s13421-011-0166-0

Uccellini, L. W., & Ten Hoeve, J. E. (2019). Evolving the national weather service to build a weather-ready nation: Connecting observations, forecasts, and warnings to decision-makers through impact-based decision support services. *Bulletin of the American Meteorological Society, 100*, 1923–1942. https://doi.org/10.1175/BAMS-D-18-0159.1

Unity Technologies. (2018). Unity. URL https://unity.com/.

van Valkengoed, A. M., & Steg, L. (2019). Meta-analyses of factors motivating climate change adaptation behaviour. *Nature Climate Change, 9*, 158–163. https://doi.org/10.1038/s41558-018-0371-y

Wachinger, G., Renn, O., Begg, C., & Kuhlicke, C. (2013). The risk perception paradox–implications for governance and communication of natural hazards. *Risk Analysis : An Official Publication of the Society for Risk Analysis, 33*, 1049–1065. https://doi.org/10.1111/j.1539-6924.2012.01942.x

Wang, C., Hou, J., Miller, D., Brown, I., & Jiang, Y. (2019). Flood risk management in sponge cities: The role of integrated simulation and 3d visualization. *International Journal of Disaster Risk Reduction, 39*, 101139. https://doi.org/10.1016/j.ijdrr.2019.101139

Whitehead, J., Edwards, B., Vanwilligen, M., Maiolo, J., Wilson, K., & Smith, K. (2000). Heading for higher ground: Factors affecting real and hypothetical hurricane evacuation behavior. *Global Environmental Change Part B: Environmental Hazards, 2*, 133–142. https://doi.org/10.1016/S1464-2867(01)00013-4

Wilson, R. S., Zwickle, A., & Walpole, H. (2019). Developing a broadly applicable measure of risk perception. Risk analysis. *An Official Publication of the Society for Risk Analysis, 39*, 777–791. https://doi.org/10.1111/risa.13207

Yeo, N. L., White, M. P., Alcock, I., Garside, R., Dean, S. G., Smalley, A. J., & Gatersleben, B. (2020). What is the best way of delivering virtual nature for improving mood? An experimental comparison of high definition tv, 360° video, and computer generated virtual reality. *Journal of Environmental Psychology, 72*, 1–13. https://doi.org/10.1016/j.jenvp.2020.101500

Zachry, B. C., Booth, W. J., Rhome, J. R., & Sharon, T. M. (2015). A national view of storm surge risk and inundation. *Weather, Climate, and Society, 7*, 109–117. https://doi.org/10.1175/WCAS-D-14-00049.1

Zanola, S., Fabrikant, S. I., & Çöltekin, A. (2009). The effect of realism on the confidence in spatial data quality in stereoscopic 3d displays. In *Proceedings of the 24th international cartography conference (ICC 2009)* (pp. 15–21). ICC. https://doi.org/10.5167/uzh-27017. Santiago, Chile.