NOTICE: This is the author's version of a work that was accepted for publication in **Computers in Human Behavior**. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms, may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in **Computers in Human Behavior**, Volume 50, September 2015, Pages 508-519, DOI: http://dx.doi.org/10.1016/j.chb.2015.03.074

Please cite as:

Chittaro L., Sioni R. Serious games for emergency preparedness: Evaluation of an interactive vs. a non-interactive simulation of a terror attack, *Computers in Human Behavior*, *50*, 2015, pp. 508-519.

Serious Games for Emergency Preparedness: Evaluation of an Interactive vs. a Non-Interactive Simulation of a Terror Attack

Luca Chittaro and Riccardo Sioni

Human-Computer Interaction Lab
Department of Mathematics and Computer Science
University of Udine
Via delle Scienze 206, 33100, Udine, Italy

Abstract

Emergency preparedness is a relevant emerging application of serious games. A general issue in exploiting such approach concerns the breadth of the population that can be reached by serious games. Indeed, serious games need to be actively played and this can restrict their user population, because there are people who have no experience with video games or do not like them or do not have the proper hardware to play them. Moreover, there are organizational contexts in which a non-interactive presentation is preferred because it can be given in a more convenient and less costly way with simple traditional media, i.e., printed materials, slides or videos. This paper deals with the possibility of generating and using a non-interactive version of the experience provided by serious games. First, we propose a serious game that simulates a mass emergency caused by a terror attack in a train station. To obtain design guidance, we explore psychological models that explain how people are motivated to protect themselves from danger. Then, we generate the non-interactive version of the terror attack simulation and we contrast it with the interactive version. Results of our study show that both versions of the simulation can provide positive outcomes in learning, risk severity perception and self-efficacy, but they differ in how much they affect user's threat appraisal and emotional response.

Keywords: serious games, simulation, interactivity, emergency preparedness, evaluation, Protection Motivation Theory

1. Introduction

Serious games are video games that further training and education objectives (Zyda, 2005). Emergency preparedness is a very promising application for serious games, because traditional preparedness instructions are often not attended to or not adequately understood by people, and presenting them through a game could possibly attract more attention as well as prepare people in more clear and informal ways. However, the number of serious games for emergency preparedness aimed at citizens is still limited in the literature (e.g., Chittaro & Ranon; 2009; Smith & Trenholme, 2009; Chittaro, 2012; Ribeiro, Almeida, Rossetti, Coelho, & Coelho, 2012; Silva, Almeida, Pereira, Rossetti, & Coelho, 2013; Silva, Almeida, Rossetti, & Coelho, 2013).

A critical issue we want to address with the current work is the breadth of the population that can be reached by a serious game for presenting preparedness recommendations. Serious games need to be actively played by users and this may restrict the number of individuals who could benefit from the preparedness presentation, because there are people who have no experience with video games or do not like to play them or do not have the proper hardware to play them. Moreover, there are organizational contexts in which a non-interactive presentation is preferred because it can be given in a more convenient and less costly way, e.g., emergency preparedness presentations in workplaces are typically given with simple traditional media, i.e., printed materials, slides or videos. This led us to consider the possibility of creating non-interactive versions of the serious game experience. An interesting aspect of this approach is that the same game engine and assets used to produce the serious game could straightforwardly be reused for generating the non-interactive version of the experience. Such experience could progress by itself along the game storyline by making the avatar automatically perform predetermined actions, or can be obtained by simply video-recording game sessions in which a human player carries out the possible actions in the serious game.

A second relevant issue concerning the presentation of emergency preparedness instructions as a game is the theoretical grounding for their design and evaluation, which is lacking in the literature. For this reason, a second goal of our work is to start exploring psychological models that explain how people are motivated to protect themselves from dangers. In designing serious games for emergency preparedness, such models could provide design guidance about the organization of the game elements as well as the in-game presentation of the content and recommendations. Moreover, they could provide guidance about which users' psychological variables are predictors of the motivation to protect themselves from threat and assess the effect of the gaming experience on such variables. In particular, we follow *Protection Motivation Theory* (PMT) (Rogers, 1975; 1983), a leading theory that models protection motivation on the basis of threat appraisal and coping appraisal processes.

In this paper, we propose a serious game for emergency preparedness that simulates a mass emergency caused by a terror attack in a train station and provides recommendations about emergency evacuations. We evaluate the interactive vs. the non-interactive simulation of such scenario. In the interactive simulation, users actively play the game with controllers; in the non-interactive simulation, they watch the same storyline progress automatically, showing them all the meaningful events which players must deal with for completing the corresponding interactive simulation. In other words, the non-interactive version can be considered as an instructional video made by recording video game action. In addition to assessing knowledge gain as other evaluations of serious games do, we measure users' perceived self-efficacy (which has been shown to be a predictor of real-world performance) (Bandura, 1997) as well as other variables that are significantly linked with positive changes in protective behaviors (Floyd, Prentice-Dunn, & Rogers, 2000).

The paper is organized as follows: in Section 2, we briefly review the literature on serious games, interactivity and PMT. Section 3 describes in detail the two proposed versions of the terror attack simulation. Then, Sections 4 and 5 describe in detail our experiment and its results, while Section 6 critically discusses the results. Finally, Section 7 presents conclusions and future work.

2. Related Work

Serious games for emergency preparedness aimed at citizens have been described in the literature, with a special focus on evacuation: see, as an example, (Chittaro & Ranon, 2009; Smith & Trenholme, 2009; Chittaro, 2012; Ribeiro et al., 2012; Silva et al., 2013).

In general, there is a consensus on the potential of serious games as emergency preparedness tools, but the design and evaluation of their effectiveness is still understudied. First, to the best of our knowledge, no serious game for emergency preparedness has followed existing psychological theories that model human protection motivation, which could provide a more solid theoretical basis to the design and evaluation of such games. Second, evaluations of emergency preparedness games tend to focus only on performance measures, e.g., time required for an emergency evacuation (Smith & Trenholme, 2009) or the number of times the nearest exit is chosen (Kobes, Oberijé, Groenewegen-Ter Morsche, 2009; Ribeiro et al., 2012). More generally, as pointed out by analyses of the serious game literature (Bellotti, Kapralos, Lee, Moreno-Ger, & Berta, 2013; Girard, Ecalle, & Magnant, 2013), lack of a rigorous assessment is a current issue for serious games in any domain and might be due to the complexity encountered in assessing intangible measures such as learning, emotions or motivation. Physiological measures (e.g., electrodermal and cardiovascular measures) are considered as a novel and promising instrument in the evaluation of serious games (Bellotti et al., 2013). For example, measures of physiological arousal could be used as an index of user stress and anxiety as well as immersion in realistic game experiences (Parsons et al., 2009; Kim, Rosenthal, Zielinski, & Brady, 2014).

Interactivity plays a major role in serious games, and has a significant impact on the promotion of behaviors and attitudes. This effect has been observed for both risk-averting and risk-inducing behaviors. For example, Peng (2008) observed how a serious game for healthy diet promotion was more effective in influencing user's self-efficacy than a passive observation of game play through a pre-recorded video clip. The video game employed in the study is inspired by commercial third-person simulation games like "The Sims", and allows players to choose a character and take dietary decisions on its behalf. However, the experiment was not able to control the sequence of actions taken by players: different sequences of actions that users might perform in the interactive condition (i.e., deciding how to feed the avatar and how to make the avatar exercise) could cause the game to progress in very different ways and with a different final outcome with respect to the pre-recorded video used in the non-interactive condition. Interactivity also seems to enhance the effect of risk-glorifying experiences in terms of increasing risk-promoting cognitions more than non-interactive media such as movies and music, as pointed out by a meta-review of 88 studies (Fischer, Greitemeyer, Kastenmüller, Vogrincic, & Sauer, 2011). In particular, the studies of interactive game experiences in the meta-review analyzed the effects of violent games such as FPSs, fighting games and racing games that encourage risk-taking behaviors, while the studies of non-interactive media focused on violent and risk-glorifying movies, music and ads. Effects of interactivity have been observed also for educational outcomes in serious games. Ritterfeld, Shen, Wang, Nocera, and Wong (2009) showed that a serious game for health allowed participants to increase their knowledge about the presented concepts more than a pre-recorded video clip of game play. Cole, Yoo, & Knutson (2012) analyzed effects of game interactivity at a neural level. They studied participants' brain activity with functional magnetic resonance imaging (fMRI), highlighting that personal involvement and agency in a serious game for health activates neural circuits related to motivation more markedly than a passive exposure to the same experience. Interactivity may also enhance users' self-identification with avatars in a simulation, especially in first-order mediated action, i.e., when the player uses a proximal tool (e.g., a joystick) to exert an action upon an external object (e.g., moving an avatar and swinging a virtual sword in the game) (Riva & Mantovani, 2012). In particular, first-order mediated actions allow players to perceive the proximal tool as an extension of their body and, in the case of a joystick, to use it intuitively as they use their hands and fingers (Riva & Mantovani, 2012). Self-identification has been hypothesized by Peng (2008) to have a mediating role in the relationship between interactivity and self-efficacy.

When looking into user's motivation with respect to avoidance of danger, *Protection Motivation Theory* (PMT) (Rogers, 1975; 1983) is a leading and useful theory that models how individuals are motivated to protect themselves from risks on the basis of threat appraisal and coping appraisal processes. According to PMT (Milne, Sheeran, & Orbell, 2000; Norman, Boer, & Seydel, 2005), threat appraisal is characterized by two variables: perceived *vulnerability* (i.e., how personally susceptible an individual feels to the threat) and perceived *severity* (i.e., how serious the individual believes that the threat would be to his/her own life). *Fear* is a third intervening variable that stands between perceptions of severity and vulnerability and the level of the appraised threat. Fear (and emotional response to threat) can be measured by using subjective questionnaires as well as through physiology: self-report measures are correlated with heart rate (HR) and electrodermal activity (EDA) (Mewborn & Rogers, 1979). Coping appraisal is instead characterized by perceived *response efficacy* (i.e., whether the recommended coping response will be effective in reducing threat) and *self-efficacy* (i.e., the individual's beliefs about whether he/she is able to perform the recommended coping response). *Recommendation simplicity* concerns the costs of performing the recommended response in terms of the resources of the individual. These variables have been shown to be significantly linked with positive changes in protective behaviors (Floyd et al., 2000) and, among them, self-efficacy showed the most robust association with protection motivation, as reported by Milne et al. (2000) in their meta-analysis of 27 PMT studies. Figure 1 shows a schematic representation of PMT variables and processes we mentioned.

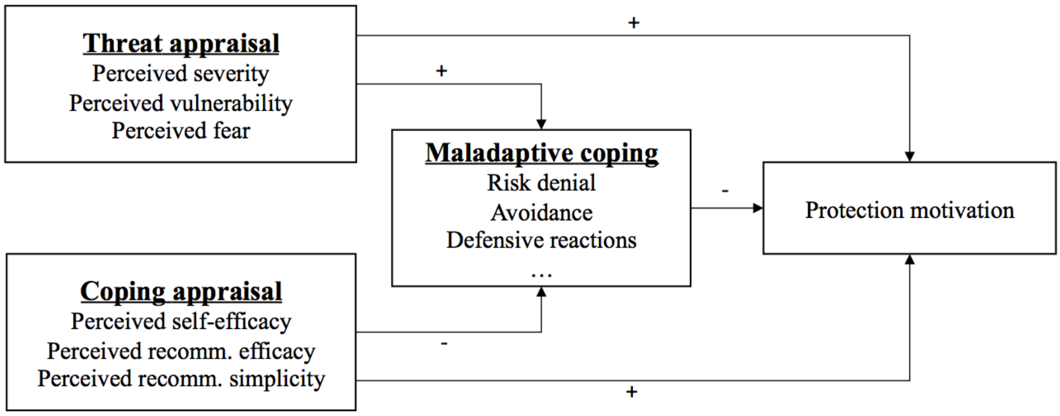

Figure 1. A schematic representation of the PMT processes and variables mentioned in the paper. The "+" and "-" sign on arrows represents respectively a positive or a negative association.

When designing an intervention concerning emergency preparedness that aims to change recipients' attitudes and behaviors with respect to a given risk, PMT recommends to first threaten the individual by highlighting the severity of the risk and the vulnerability of the individual to it. If the individual perceives the risk as not serious or perceives himself/herself as not vulnerable, the model predicts that (s)he will not feel threatened and will not be motivated to consider how to cope with the risk. After having threatened the individual, the intervention must highlight that there is an effective action that can avert the risk and that the individual is capable of taking that action. If the individual perceives the action as ineffective or perceives himself/herself as not capable of doing it, the model predicts that (s)he will try to reduce the negative emotions induced by threat through maladaptive coping, i.e., processes such as risk denial, avoidance and defensive reactions which are detrimental to learning proper safety behavior to face the risk.

In the following section, we outline how PMT could provide guidance in the design of a serious game for emergency preparedness, while in Section 4 we will describe how we took the theory into account in the evaluation of the interactive vs. non-interactive simulation by including an assessment of the above described PMT variables.

3. The Proposed Terror Attack Simulation

The experience we considered simulates, from a first-person perspective, a mass emergency caused by a series of explosions in a train station. This scenario is highly representative of citizen preparedness needs for different kinds of real-world emergencies. Indeed, terror attacks with explosives have been often carried out in different countries, in train stations (e.g., Madrid, Paris, London, Moscow, Bologna) as well as other urban areas. Moreover, such intentional attacks bear a strong resemblance in effects (and thus in the required individual preparedness for facing them) to accidental disasters that lead to large-scale explosions (e.g., involving chemical storage tanks and facilities).

As motivated in Section 1, we developed and contrasted an interactive version of the simulation (Interactive Simulation, IS in the following) and a non-interactive version (Non-Interactive Simulation, NIS in the following). The safety recommendations provided by the simulation have been taken from traditional, publicly available civil defense materials. In particular, we referred to video recommendations created by the Singapore Civil Defence Force (www.scdf.gov.sg) and to the emergency guide of the Michigan Technological University (http://www.mtu.edu/publicsafety/reports/emergencyguide/).

In this section, we first describe user interaction with the game (Section 3.1), then we discuss in detail how we used PMT in designing the terror attack simulation (Section 3.2). Finally, we provide a detailed storyline of the simulation in Section 3.3.

3.1. Game Control


a
b

Figure 2. Two screenshots from the terror attack simulation showing the train station before the explosions, from the outside (a) and inside (b). In (b), the blue triangle-shaped arrow indicates the direction of the next checkpoint.

IS starts with a short introduction to the game controls: this is done inside the game by starting just outside the train station (Figure 2a), on the other side of the road, and asking users to follow a series of checkpoints in a predefined order that takes them inside the train station (Figure 2b) to a specific seat on a train. During this introduction, the next checkpoint that the user must reach is always highlighted by a blue aura in the 3D world and its direction is indicated by a blue triangle-shaped arrow in the current view (Figure 2b). Once players have reached the seat, the explosions struck the station, ending the familiarization with controls and starting the terror attack simulation (Figure 3b). The initial familiarization with controls is not needed in NIS because users do not operate game controls in NIS.

a

b

Figure 3. Two screenshots from the terror attack simulation showing the train station after the explosions, from the outside (a) and inside (b).

The simulation of the emergency in IS and NIS progresses by following the same sequence of events. At the beginning, an on-screen message indicates for a few seconds that the goal is to evacuate the train station. Along the path, users experience a sequence of six dangerous situations that illustrate different threats that can be encountered during a real-world emergency evacuation. The simulation ends when the user has seen the entire evacuation path, which ends outside the train station where emergency medical services are gathering (Figure 3a). In the design of the simulation, we were particularly careful not to promote any possible risky behavior, emphasizing the necessity for users to pay attention and be careful, to avoid encouraging risk-positive cognitions and attitudes, as pointed out by Fischer et al. (2011).

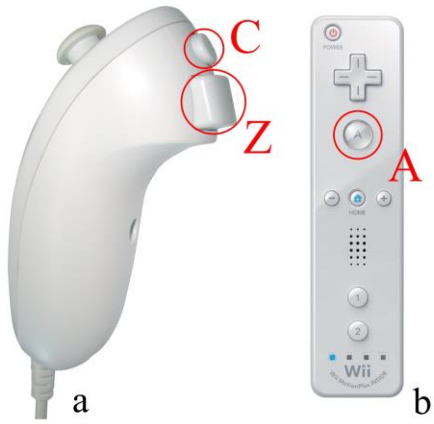

Figure 4. The Nintendo Nunchuck controller (a) and the Nintendo Wii Remote (b). The C, Z, and A buttons are circled.

To move in the environment, IS users employ the Nintendo Nunchuck controller, equipped with a joystick and two buttons (Figure 4). By moving the joystick forward or backward, users walk respectively forward and backward in the virtual world; by moving the joystick to the left or to the right, users rotate respectively counter-clockwise and clockwise. Button Z is used to run, while button C is used to crouch. The Nintendo Wii Remote, to which the Nunchuck is connected, is instead employed to interact with objects by pressing the A button (see Figure 4).

For NIS, a 3-min video obtained by recording a full play session of the interactive simulation shows the same sequence of six dangerous situations and provides the same recommendations given by IS, without requiring any interaction. More specifically, for each dangerous situation, the recorded game play (during which the avatar makes a mistake and dies) is in NIS about 20 s long, followed by the 10 s that are devoted (in NIS as well as IS) to show the recommendations, as described in detail in the next section.

3.2. Using PMT in the Design of the Simulation

To use PMT as guidance in the design of a serious game for emergency preparedness, one can consider all PMT variables introduced in Section 2 and define the game elements that are meant to appropriately affect each variable. To do so, we first formulated these possible guidelines:

- *Severity*. As reported in (Norman et al., 2005), many studies in the literature have successfully manipulated the perception of the threat by emphasizing the fact that it may cause severe harm. In the serious game context, to convey threat severity, the simulation should thus use visual and audio stimuli that emphasize the negative consequences suffered by the user's avatar. For example, every time the avatar is injured, audio and video depictions of the severity of the consequences (e.g., non-verbal vocal sounds of human distress, sounds of breaking bones, blood squirts, simulation of temporary blindness, tinnitus or dizziness, a decreasing life bar…) should be presented. Regarding vocal signals, Bachorowski and Owren (2008) underline the importance and effectiveness of their emotional features in influencing listener's affect and modulating his/her behavior. Vocal signals can be coupled with visual stimuli to depict the severity of an emergency situation, e.g., Chittaro and Zangrando (2010) used coughing sound coupled with visual dark vignetting to emphasize the suffering of users' avatar caused by smoke in a fire simulation.

- *Fear*. The stimuli employed to increase perceived severity should be designed to be threatening and emotionally strong for increasing the fear component of PMT. As reported by Öhman (2008), fear factors can be grouped in four categories: (i) fears about interpersonal events or situations; (ii) fears related to death, injuries, illness, blood and surgical procedures; (iii) fear of animals; (iv) agoraphobic fears. In the serious game context, to increase fear perception, the design of the simulation should thus focus on realistically reproducing the fearful elements of the specific real-world emergency that belong to one or more of the above mentioned categories. In the terror attack scenario we considered, such elements belong to the second category. For example, vocal sounds of distress (e.g., coughs and pain screams) for the avatar and other characters should be realistically recorded with human actors and blood squirts should be graphically rendered in a clear and disturbing way. Stimuli like full-screen blood squirts and realistic non-verbal sounds of suffering and suffocation have been successfully employed in (Chittaro, Buttussi & Zangrando, 2014; Chittaro & Zangrando, 2010) to elicit anxiety and fear in simulations of other kind of emergencies (aircraft ditching and structure fire, respectively).
- *Vulnerability*. As reported in (Norman et al., 2005), studies in the PMT literature have increased perceived vulnerability by providing information that stresses how the considered threat may put people similar to participants (e.g., with respect to gender and age) at risk. In the serious game context, a possible approach is to include in the virtual environment elements (such as buildings, people, cars, vegetation, signs, sounds…) that match those of the contexts in which the intended users find themselves in their real-world experience.
- *Self-efficacy*. To manipulate self-efficacy, it is necessary to argue that the individual has or lacks the ability to perform the recommended response (Norman et al., 2005). A serious game for emergency preparedness should thus provide information that emphasizes players' ability in putting into practice the risk-avoiding behaviors. Gaining experience in

performing the given behavior is a major factor that contributes to increase self-efficacy (Bandura, 1997). In this sense, the simulations that people can experience with a serious game could allow them to actually succeed in applying safety knowledge in the real world.

- *Recommendation simplicity*. As suggested in (Maibach & Parrott, 1995), to make users perceive the considered recommendations as simple, their costs (in terms of money, time, physical effort, inconvenience…) must be anticipated and correctly communicated, because high perceived costs can inhibit the effectiveness of fear appeals. Therefore, to be perceived as simple, the recommendations provided during the simulation should be simple per se in the real world, otherwise their cost must be clearly communicated. Moreover, in the context of serious games, one should be careful in designing a simple interaction to prevent that feelings of difficulty due to the complexity of game controls could end up in being associated to the recommended behaviors.
- *Recommendation efficacy*. To increase the perception of recommendation efficacy, one must emphasize the effectiveness of the recommendation in reducing or preventing the effects of the considered threat (Norman et al., 2005). In the case of a serious game for emergency preparedness, recommendations should clearly present the proposed behaviors as effective in reducing any risk (e.g., serious injuries or even death) related to the threat considered by the simulation.

We followed the above proposed guidelines in designing the terror attack simulation. In particular, to address the severity and fear guidelines, the proposed simulation exploits visual and audio stimuli inspired to those employed in commercial action games, especially *first-person shooter* (FPS) games. More specifically:

- When users come into contact with smoke, coughing sounds are played, and a dark vignetting is shown on screen (Figure 5a). When users are crouched under smoke, the vignetting is still shown, but the coughing sound is played less frequently.
- When users are injured by metal or concrete debris, blood squirts are shown on screen (Figure 5b) and the sound of a human screaming in pain is played;
- When users are hit by a nearby explosion, the screen flashes in white for a fraction of a second, and tinnitus sound is played;
- When users come into contact with fire, a red vignetting is shown on screen (Figure 5c), and the sound of human voice, screaming in pain, is played;
- During the entire simulation, users can hear voices of wounded people in pain, ambulance sirens, and the noise of helicopters flying over the station;

- A life bar is always shown on the top left of the screen to indicate avatar's health and possibly make users more aware of the severity of the threat. When the avatar is healthy, the bar is completely green; every time the avatar is injured, health decreases, i.e., part of the bar becomes red, starting from the left side; when the bar is completely red, the avatar dies. When the avatar comes into contact with smoke, health decreases constantly (if the avatar is standing, the bar decreases faster with respect to crouched avatar).

Figure 5. Three screenshots from the terror attack simulation, showing the user coming into contact with: smoke (a), metal debris (b), and fire (c).

To address vulnerability, we took extreme care in modeling the virtual environment to make it clearly look like a typical train station of the country in which participants live.

To address perceived efficacy and simplicity of the recommendations, they have been designed to be as concise and clear as possible, and to convey the message that, by paying sufficient attention, users can safely survive an emergency evacuation. Also, the controls of IS have been designed to be easy to learn and understand also for users who do not play video games frequently, to allow them to follow the recommendations easily.

To present the recommendations, every time the avatar dies during the simulation, we show on screen both a message that explains why the avatar has died (*cause-of-death message*) and then the recommendation that must be followed to avert the threat or to reduce its impact on health. A picture of a skull and crossbones accompanies the cause-of-death message, and a glowing light bulb accompanies instead the recommendation (Figure 6).

We created six pairs of cause-of-death message and recommendation, one for each of the six different threats that users face during the terror attack simulation. While we used the same recommendations in IS and NIS, we had to slightly differentiate the cause-of-death messages to be consistent with the presence or absence of interactivity: in IS, users are in control of the actions carried out by the avatar, while in NIS they are passive spectators of a non-interactive simulation. The cause-of-death messages, therefore, have been phrased to reflect the fact that the actions carried out by users in IS are directly related to the negative consequences suffered by the avatar, while in NIS they observed a possible negative outcome (that is why we use the word “can”) that can be avoided by carrying out the correct behavior suggested by the recommendations. The full cause-of-death message and recommendation we used are the following:

1) When users die of toxic smoke inhalation, the cause-of-death messages are respectively “You have died because you have inhaled smoke and toxic fumes” (IS), “Inhaling smoke and toxic fumes can kill you” (NIS), while the recommendation is “During the evacuation, be careful of smoke and toxic fumes. If there is too much smoke, keep low (30-60 cm). Walk away for about 500 m”. The pair of cause-of-death message and recommendation used in IS is shown in Figure 6.
2) When users are killed by sharp-edged debris, the cause-of-death messages are respectively “You have died because of the wounds from sharp-edged debris or wreckage” (IS), “Sharp-edged debris and wreckage can fatally wound you” (NIS), and the recommendation is “During the evacuation, pay attention to sharp-edged debris and wreckage”.

3) When users are crushed by falling debris, the cause-of-death messages are respectively "You have died, crushed by debris falling from buckled structures" (IS), "Debris falling from buckled structures can crush and kill you" (NIS), and the recommendation is "During the evacuation, pay attention to damaged structures and overhanging or buckled slabs".
4) When users are hit by the explosion of a tank wagon, the cause-of-death messages are respectively "You have died because of a tank wagon explosion" (IS), "Going near objects that may explode can kill you" (NIS) and the recommendation is "During the evacuation, if you notice objects that may explode, keep away from them".
5) When users die by fire, the cause-of-death messages are respectively "You have died because of the burns caused by fire" (IS), "Burns caused by fire can kill you" (NIS), and the recommendation is "During the evacuation, pay attention to flames and fires".
6) When users die because a vehicle runs over them, the cause-of-death messages are respectively "You have died because you have been run over by a vehicle" (IS), "Vehicles can run over you and kill you" (NIS), and the recommendation is "When you cross the road during the evacuation, pay attention to the transit of vehicles".

Each pair of cause-of-death message and recommendation is shown to users for 10 s in both IS and NIS.

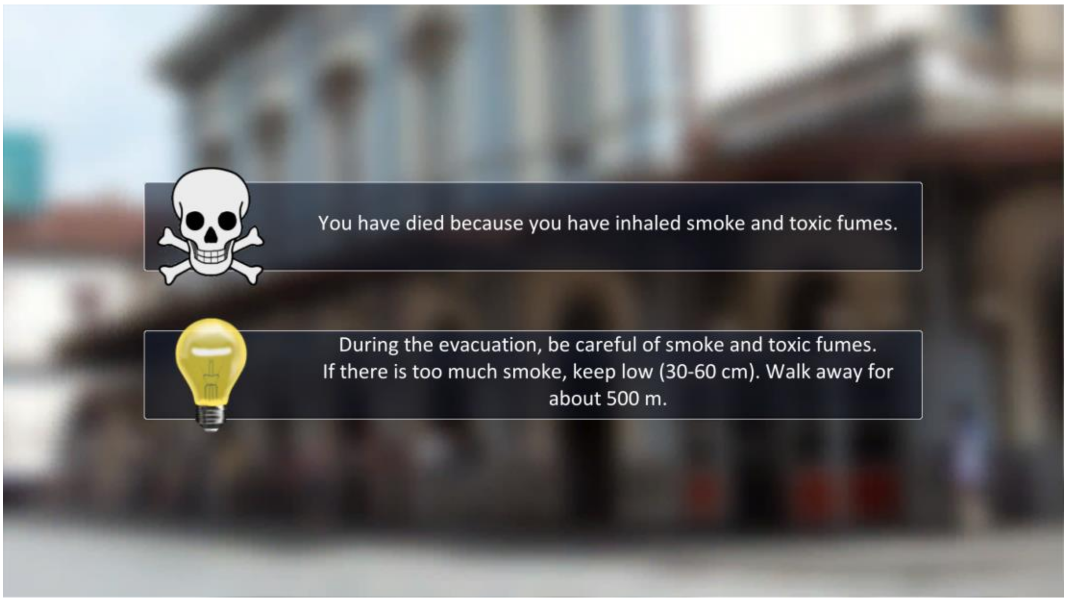

Figure 6. The pair of cause-of-death message (accompanied by a skull and crossbones) and recommendation (accompanied by a glowing light bulb) for toxic smoke inhalation in IS. The original text was in Italian and it has been translated for this picture.

3.3. Storyline of the Terror Attack Simulation

In the following, we describe the six dangerous situations depicted by the terror attack simulation. While NIS is a recording that shows each of the six situations and the associated mistake and recommendation to every user, in IS we had to take explicit game programming steps to guarantee that every IS user actually did the mistake in each of the six situations and was thus shown each of the six recommendations. We explain how we achieved this by examining the detailed storyline of the simulation:

1) To exit the train, users must proceed along coaches and find an exit. Coaches are filled with thick, dark smoke. To inhale the smallest amount of smoke and survive until the exit is reached, users must proceed crouched. However, the first time users try to exit the train, their life bar decrease is set to a speed that makes it impossible to reach the goal. In this way, users see the recommendation for smoke inhalation at least once during the play session with IS. In subsequent attempts, the life bar decrease in the crouched position is slow enough to make it possible to exit the train.
2) After exiting the train, users are on the station platform. The first time, the only visible escape route requires them to proceed through a small passage close to a wagon wreck. The sharp-edged debris of the wagon fatally wounds the avatar. In subsequent attempts, another escape route is made visible, allowing users to proceed along the platform.
3) Part of the platform roof along users' path is unstable: concrete dust is falling from it, and some slabs appear to be on the point of falling. The first time users try to pass under the slabs, an explosion makes the roof collapse and a slab falls over them, killing them on the spot. In subsequent attempts, the explosion occurs before users pass under the dangerous part of the roof, so they can then proceed safely and reach an underpass that leads outside the station.
4) Users walk in the underpass until they find it blocked by a large slab fallen from a platform roof, and they must go out to the next platform. Here, the only escape route is close to a tank wagon that is intact, but partially covered in flames. A plaque on the tank wagon indicates that it contains flammable material. When users try to pass close to it, the wagon explodes and the explosion kills them. In the subsequent attempt, another escape route is made visible, allowing users to proceed along the platform, cross two train tracks, and reach the next platform.
5) Users see a new entrance to the underpass that leads out of the station but is partially surrounded by flames. When they approach this entrance, they try to run in the small space between two flames, but the flames burn and kill them. In the subsequent attempt, another

escape route is made visible, allowing users to proceed along the platform and reach the train station lounge.

6) From the lounge, users can see ambulances and first responders taking care of victims on the bus stops outside the station. The first time users try to reach the first responders by crossing the road, a running ambulance leaving the station hits and kills them. In the subsequent attempt, the ambulance leaves a few moments in advance, allowing users to avoid it if they cross the road carefully.

4. Experimental Evaluation

The evaluation of the terror attack simulation followed a between-subject design, with interactivity (IS or NIS) as the independent variable.

4.1. Materials and Measures

Both versions of the simulation were displayed in full-screen mode on a 30'', 2560 x 1600 pixel LCD monitor. The distance between the screen and the participant was about 1 m. The lights in the room used for the evaluation were turned off: as reported in the literature (Dekker & Champion, 2007), room brightness may have negative effects on a virtual experience. Volume of sound was kept identical in the two conditions.

A PC was devoted to run the simulation, while a second PC recorded participants' physiological data at 2048 Hz with a Thought Technologies Procomp Infiniti encoder. We employed two physiological sensors, following the placement suggestions described in (Andreassi, 2007):

- An EDA sensor, placed on the intermediate phalanges on the middle and ring fingers of one hand. From EDA, the skin conductance level (SCL), which represents the electrical conductivity of the skin at a given point in time, can be extracted through decomposition analysis (Benedek & Kaernbach, 2010). SCL is commonly employed in the literature as a physiological parameter that measures arousal (Boucsein, 2006; Andreassi, 2007; Nacke & Lindley, 2008). As reported in (Mewborn & Rogers, 1979), EDA is also a correlate of fear.
- A photoplethysmograph (PPG) for blood volume pulse (BVP), placed over the distal phalanx of the index finger of the same hand. Increases in HR (which is calculated from BVP signal) are generally related to emotional activation. As reported by Boucsein (2006), compared to EDA, HR is well suited as an indicator for the higher arousal range and for pronounced arousal processes. Furthermore, HR is positively correlated to tension and

negative affect in stressful first-person experiences (Drachen, Nacke, Yannakakis, & Pedersen, 2010).

To collect participants' demographic data and their subjective opinions, we employed the following questionnaires:

- *Demographic questionnaire*. We asked participants about (i) their age, (ii) gender, (iii) if they had ever been to a city train station in our country and, if that was the case, (iv) how often they visit such station on a 5-point scale (1 = less than once a year; 2 = from 1 to 6 times a year; 3 = at least once a month; 4 = at least once a week; 5 = almost every day). As a final question, they were asked how often they play 3D video games on a 7-point scale (1 = never; 2 = less than once a month; 3 = about once a month; 4 = more than once a month; 5 = more than once a week; 6 = everyday, for less than an hour; 7 = everyday, for more than an hour).
- *Knowledge.* Participants were asked two questions before and after the simulation to assess their emergency preparedness knowledge for the considered type of scenario. To avoid suggesting possible answers (e.g., as a multiple-choice questionnaire would do), participants were asked to answer orally the questions. They were asked to imagine themselves in an emergency evacuation of a train station after an explosion. The first question was "What should one pay attention to during the evacuation?" We compared participants' answers with a checklist of six items, one for each threat described in Section 3: smoke and toxic fumes, objects that may explode, flames and fires, damaged structures and slabs that may fall, debris and sharp-edged wrecks, transit of vehicles. When participants mentioned a threat, we checked the corresponding item. The second question was "What is the correct behavior that must be carried out when one is inside a closed environment filled with smoke and toxic fumes?" We compared participants' answers with the recommendation described in Section 3.2. If participants mentioned this behavior, we checked a seventh item. The knowledge score was calculated as the total number of items checked, and is thus in the 0-7 range.
- *Vulnerability*. We measured perceived vulnerability before and after the simulation by using the three items employed by (de Hoog, Stroebe, & de Wit, 2008), changing the name of the considered risk into "emergency evacuation". The three items asked how vulnerable respondents perceived themselves to be with respect to an emergency evacuation; how high they thought their risk of being involved in an emergency evacuation was; and how high the probability of suffering personal negative consequences from an emergency evacuation was.

Participants answered on a 7-point Likert scale (1 = “not at all”, 7 = “very”) and the three answers were averaged.

- *Severity*. We measured perceived severity before and after the simulation by using the three items employed by (de Hoog et al., 2008), changing the name of the considered risk into “emergency evacuation”. Participants rated severity of risk on three items which respectively asked how severe, harmful, and serious the consequences of an emergency evacuation would be. Ratings were given on a 7-point Likert scale (1 = “not at all”, 7 = “very”), and the three answers were averaged.
- *Self-efficacy*. We measured self-efficacy before and after the simulation, by using a 5-items questionnaire we designed by using the Generalized Self-Efficacy Scale (Schwarzer & Jerusalem, 2005) as an inspiration. The items in our questionnaire are: “I am confident that I am able to effectively deal with an emergency evacuation”, “Thanks to my resources, I know how to manage an emergency evacuation”, “I would be able to deal with an emergency evacuation even if there are smoke and toxic fumes in the environment”, “I would be able to deal with an emergency evacuation even if I find flames and fires along the way”, “I would be able to deal with an emergency evacuation even if I find objects that may explode along the way.” Ratings were given on a 7-point Likert scale (1 = “not at all”, 7 = “very”), and answers were averaged to form a scale. To assess the reliability of our questionnaire we used Cronbach’s alpha (pre-test: 0.88; post-test: 0.83 for IS, 0.93 for NIS).
- *Recommendation efficacy*. We measured recommendation efficacy after the simulation, by using a 3-items questionnaire we designed. The items were: “The provided recommendations are useful for my safety”, “The provided recommendations will allow me to effectively deal with an emergency evacuation”, “By following the provided recommendations, I can strongly reduce the probability of being injured during an emergency evacuation.” Ratings were given on a 7-point Likert scale (1 = “not at all”, 7 = “very”), and answers to items were averaged to form a scale. To assess the reliability of our questionnaire we used Cronbach’s alpha (0.75 for IS, 0.75 for NIS).
- *Recommendation simplicity*. We measured recommendation simplicity after the simulation, by using a 3-items questionnaire we designed. The items were: “The provided recommendations can be easily learned”, “The provided recommendations can be easily remembered”, “The provided recommendations can be easily carried out.” Ratings were given on a 7-point Likert scale (1 = “not at all”, 7 = “very”), and answers were averaged to form a scale. To assess the reliability of our questionnaire we used Cronbach’s alpha (0.61 for IS, 0.60 for NIS).

- *Attention*. We measured attention before and after the simulation by using the two items employed by (van den Bos et al., 2008). Participants rated how much the simulation put them in a condition of respectively alertness and attention. Ratings were given on a 7-point Likert scale (1 = “very weakly”, 7 = “very strongly”) and were averaged.
- *Agitation*. We measured attention before and after the simulation by using the two items employed by (van den Bos et al., 2008). Participants rated how much the simulation put them in a condition of respectively agitation and tension. Ratings were given on a 7-point Likert scale (1 = “very weakly”, 7 = “very strongly”) and were averaged.

4.2. Participants

The evaluation involved a sample of 44 participants (30 M, 14 F) recruited among graduate and undergraduate students at our university and people from other occupations. Participants were volunteers who received no compensation. All participants had been in a city train station in our country at least once.

The data collected through the demographic questionnaire was used to assign participants to the two conditions in a balanced way with respect to the demographic indexes (see Section 4.1). As a result, mean age was 23.59 (SD = 3.83) in the IS group and 23.77 (SD = 2.72) in the NIS group; there were 15 M and 7 F in each group; mean visiting frequency of train station was 3.14 (SD = 1.21) in the IS group and 3.09 (SD = 1.48) in the NIS group; mean video game playing frequency was 3.59 (SD = 2.20) in the IS group and 3.68 (SD = 2.03) in the NIS group. These small demographic differences between the two groups are not statistically significant, as confirmed by independent samples t-test.

4.3. Procedure

Participants were clearly informed that the collected experimental data was going to be analyzed anonymously for research purposes. They were asked to fill the demographic questionnaire, then they answered the knowledge questions and the written items for vulnerability, severity, and self-efficacy.

In IS, the Wii Remote and Nunchuck were handed to participants before the play session and they were asked if they preferred to hold the Wii Remote with the left hand and the Nunchuck with the right hand, or vice versa. The skin of the palm of the index, middle, and ring fingers of the hand chosen for holding the Wii Remote were cleaned with a pad of cotton wool and alcohol, and the physiological sensors were applied. The hand holding the Wii Remote was chosen because we wanted to minimize the possibility of signal artefacts due to hand motion in physiological recordings: the simulation is indeed played mostly with the Nunchuck, while the Wii Remote is used much less frequently, only for the A button. In NIS, the sensors were always applied to the participants' right hand after skin cleaning.

After applying the sensors, participants were asked to relax while watching a 2-min video with relaxing pictures and music, to record the baseline data for the physiological signals, i.e., the signals' values that can be observed when participants are in a resting state. When analyzing data, baseline values have to be subtracted from the data recorded during the experimental conditions, to separate the physiological responses to experimental stimuli from the intrinsic physiological differences among participants. Participants were asked to relax as much as possible during the 2-min video, and they were allowed to close their eyes if they preferred to do so. Then, participants were told that they were going to try a video game (for participants in the IS group) or watch a video (for participants in the NIS group) that illustrates safety recommendations for emergency evacuation.

After trying IS or NIS, participants answered the written items for attention and agitation, then they answered the knowledge questions orally. Finally, they answered the written items for vulnerability, severity, self-efficacy, recommendation efficacy and recommendation simplicity. After completion of the questionnaires, participants were debriefed about the experiment and thanked for their participation.

5. Results

Figure 7 shows mean values of knowledge, vulnerability, severity, and self-efficacy. As suggested in (Cohen, 2001) for dealing with pre-test and post-test scores, we analyzed differences in knowledge, vulnerability, severity, and self-efficacy means using ANCOVA, with the pre-test scores as the covariate and post-test scores as the dependent variable. In this way, the analysis controls for pre-test scores. Results revealed significant differences between IS and NIS in post-test scores of vulnerability (IS: M = 5.00, SD = 1.09; NIS: M = 4.00, SD = 1.18; $F(1, 41) = 7.77$, $p < 0.01$, $\eta_p^2 = 0.16$) and severity (IS: M = 6.06, SD = 1.00, NIS: M = 5.27; SD = 0.95; $F(1, 41) = 7.10$, $p < 0.05$, $\eta_p^2 = 0.15$). No significant differences in post-test scores between IS (M = 6.14, SD = 1.42) and NIS (M = 6.64, SD = 1.89) were found for knowledge and self-efficacy (IS: M = 4.07, SD = 1.10; NIS: M = 4.36, SD = 1.33).

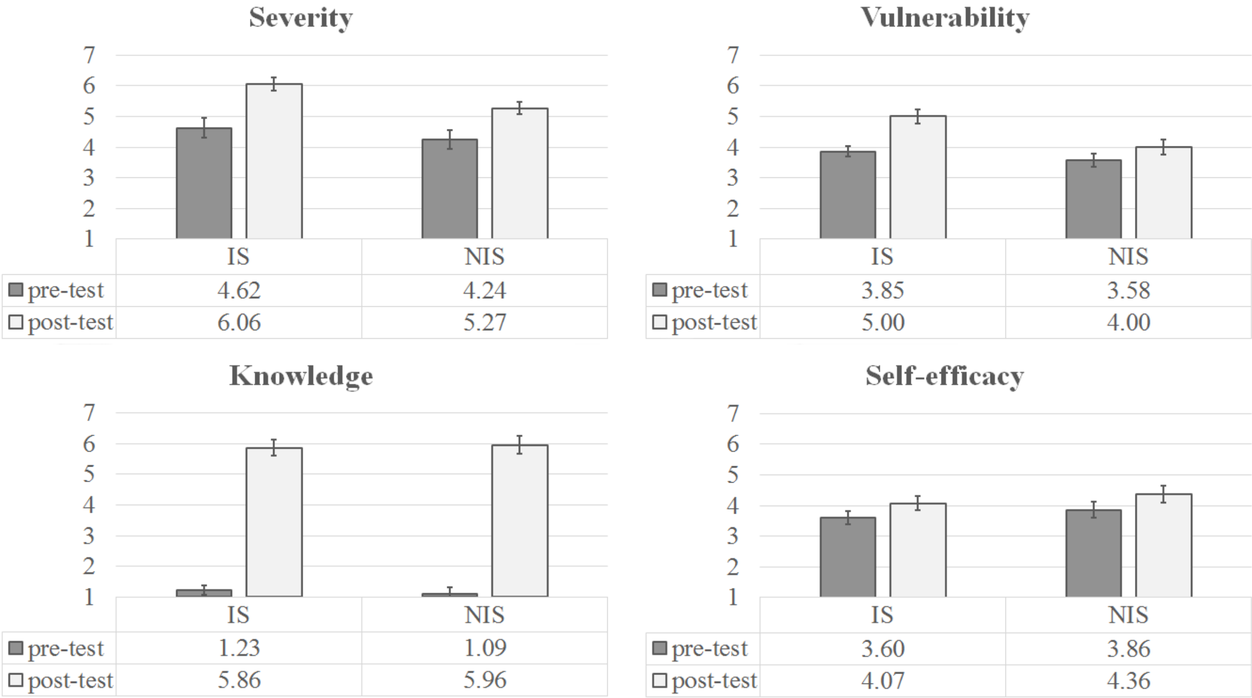

Figure 7. Mean pre-test and post-test severity, vulnerability, knowledge and self-efficacy. Error bars indicate standard error of the mean.

Moreover, we performed two-way mixed-design ANOVAs with time (before the simulation, after the simulation) as the within-subjects variable and interactivity (IS, NIS) as the between-subjects independent variable to test if the increase in knowledge, vulnerability, severity, and self-efficacy from pre-test to post-test results was significant. The analysis revealed: a main effect of time ($F(1, 42) = 454.79$, $p < 0.001$, $\eta_p^2 = 0.92$) and no significant interaction between time and interactivity for knowledge; a significant interaction ($F(1, 42) = 6.08$, $p < 0.05$, $\eta_p^2 = 0.13$) for vulnerability; a main effect of time ($F(1, 42) = 48.72$, $p < 0.001$, $\eta_p^2 = 0.54$) and no interaction for severity, a main effect of time ($F(1, 42) = 7.72$, $p < 0.01$, $\eta_p^2 = 0.16$) and no interaction for self-efficacy. For the interaction concerning vulnerability, Bonferroni post-hoc analysis shows a significant increase in IS (adjusted p value < 0.001) but not in NIS.

Figure 8 shows mean values of attention, agitation, recommendation efficacy, and recommendation simplicity. T-tests with Welch correction did not reveal significant differences between conditions in attention (IS: M = 5.91, SD = 1.04; NIS: M = 5.55, SD = 1.03), agitation (IS: M = 5.02, SD = 1.49; NIS: M = 4.34, SD = 1.76), recommendation efficacy (IS: M = 5.83, SD = 0.89; NIS: M = 5.97, SD = 0.91) or recommendation simplicity (IS: M = 5.92, SD = 0.94; NIS: M = 6.36, SD = 0.63).

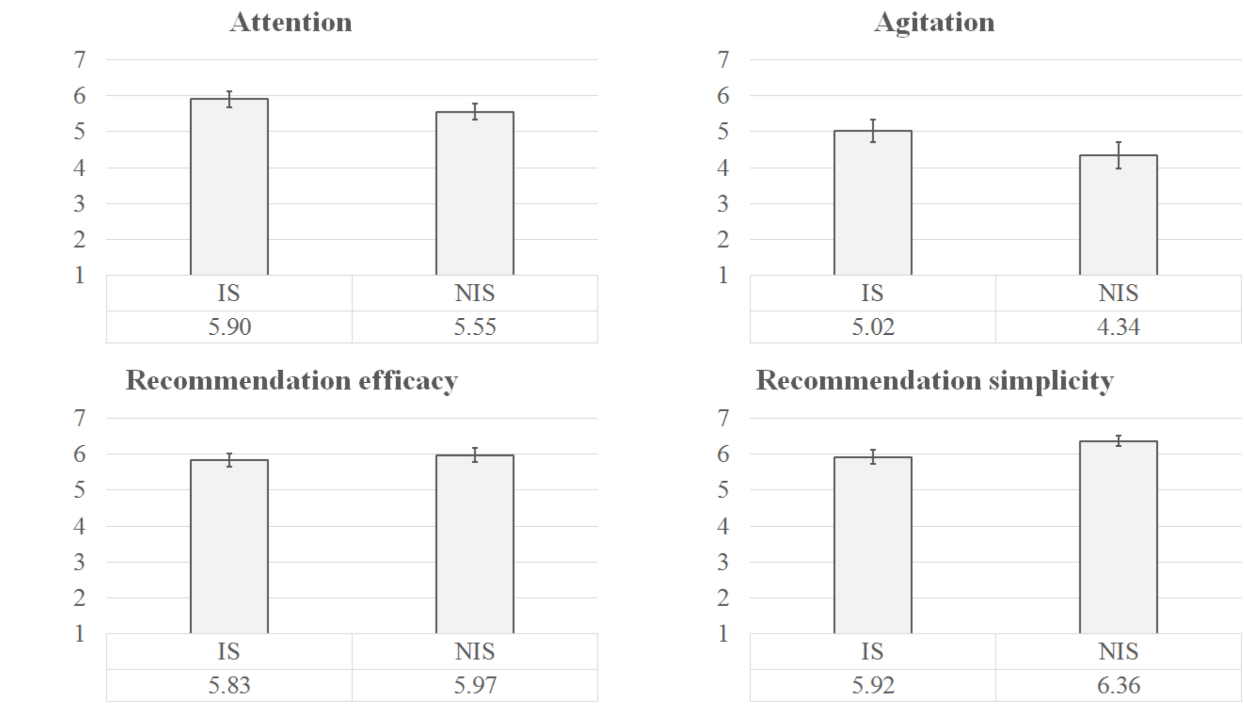

Figure 8. Mean values of attention, agitation, recommendation efficacy and recommendation simplicity. Error bars indicate standard error of the mean.

Figure 9 shows mean values of HR and SCL. T-tests with Welch correction revealed significant differences between conditions in HR (IS: M = 4.70, SD = 7.75; NIS: M = -0.20, SD = 2.79; t(26.36) = 2.79, $p < 0.01$, two-tailed) and SCL (IS: M = 1.77, SD = 1.99; NIS: M = 0.31, SD = 0.30; t(21.94) = 3.39, $p < 0.01$, two-tailed).

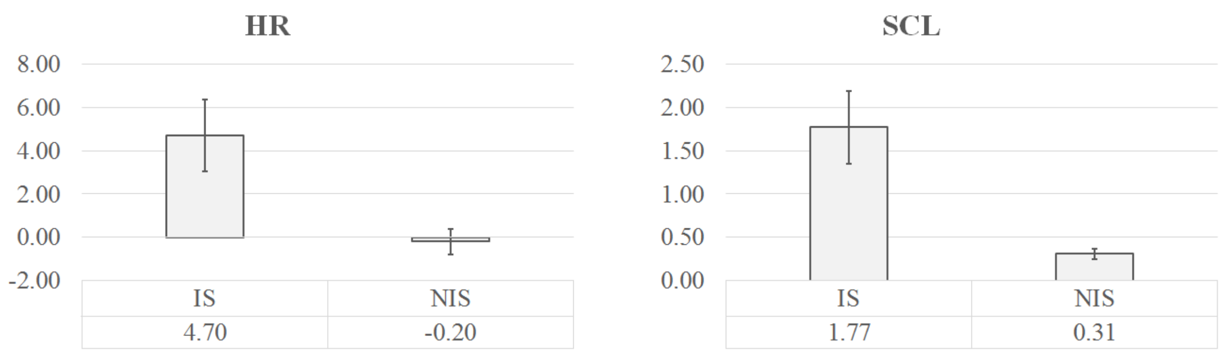

Figure 9. Mean values of HR and SCL, after baseline subtraction. Error bars indicate standard error of the mean.

6. Discussion

Results show that the interactive simulation had a larger impact on PMT variables related to the threat appraisal process with respect to the non-interactive simulation. More specifically, perceived severity and vulnerability were significantly larger in IS than NIS. The higher values of HR and SCL in IS, two physiological measures that indicate arousal and are also related to fear (Mewborn & Rogers, 1979) are consistent with a more intense emotional response as a result of the threat appraisal process. Taken together, these results highlight possible advantages of IS with respect to NIS, indicating that interactivity enhanced users' perception of threat, which is a desirable outcome in psychological models of users' protection motivation, as long as simple and effective recommendations are included in the simulation. This seems to be the case in our study, considering the relatively high scores we obtained for perceived simplicity and efficacy of recommendations, two PMT variables involved in coping appraisal.

Results also show a significant increase in emergency preparedness knowledge and self-efficacy in both conditions, suggesting that IS as well as NIS helped participants learn from the simulation. These results seem to indicate that both simulations could be effectively employed for emergency preparedness purposes, and NIS could thus provide an alternative to IS in those contexts for which serious games are unsuitable (see Section 1).

To explain the differences in threat appraisal observed between the two conditions, we hypothesize that interactivity may have increased participants' self-identification with their avatars. This relationship between interactivity and self-identification has been analyzed in the literature (e.g., Riva & Mantovani, 2012). Moreover, Park, Lee, Jin, and Kang (2010) as well as Riva and Mantovani (2012), explored the relationship between self-identification and presence, which is usually defined as the "sense of being there" or the "feeling of being in a world that exists outside the self". In addition to having a significant effect on users' emotions (e.g., a greater level of presence is able to enhance affective states such as anxiety), the feeling of presence is greater in "emotional" environments (i.e., environments that are able to elicit anxiety, relaxation…) (Kim et al., 2014; Riva et al., 2007). Therefore, in our case, interactivity may have enhanced participants' perception of being involved in the negative events of the simulation storyline. Moreover, interactivity allows one to experience a clear connection between wrong behaviors directly acted by players and the negative effects on their avatar, because these effects are presented (through visual and audio stimuli) immediately after players' actions. In NIS, conversely, users are simply spectators of the actions presented during the simulation. The connection between action and effects may be harder, less immediate and less compelling to perceive for users in NIS rather than IS. Furthermore, in IS we tried to emphasize the direct role of participants in the simulation also through the cause-of-death messages, which were phrased to reflect the fact that the negative effects on the avatar were directly caused by users' actions in the serious game (see Section 3.2). Finally, the larger HR and SCL in IS rather than NIS could corroborate the hypothesis above, because increases in physiological arousal are related to greater sense of presence (Riva et al., 2007) and immersion (Kim et al., 2014).

Unlike Peng (2008), the interactive and the non-interactive simulations in our study increased perceived self-efficacy in a similar way. A possible factor that contributes to explain this discrepancy is that, unlike our conditions, the experiment by Peng (2008) allowed the interactive simulation to be different and more personalized in its progress with respect to the non-interactive one. Moreover, no differences between IS and NIS have been observed for the PMT variables related to coping appraisal (self-efficacy, recommendation efficacy, recommendation simplicity). As a possible explanation for these results, one can consider that the recommendations in our experiment were the same in both conditions and the graphical and auditory illustration of the effects of following or not the recommendations were identical, therefore it is not surprising that perceived recommendation efficacy and simplicity were similar in IS and NIS. The strong similarity between the two simulations could also contribute to explain the lack of differences in knowledge measurements.

Finally, no statistically significant differences were observed between conditions in self-reported attention and agitation perceived during the simulation. This result suggests that the choice of duration of NIS (180 s), a passive simulation that could leave room for distraction, was appropriate to keep participants' attention at a level that was comparable to IS. The difference in perceived agitation, which is a measure associated to negative arousal and fear, although not reaching statistical significance, is consistent with HR and SCL results.

Looking critically at physiological results, a possible alternative interpretation of the obtained measures could be that the higher values in IS might be due to the greater length of that experience ($M = 655.73$ s, $SD = 184.51$) with respect to NIS (exactly 180 s): living a longer experience (in which the avatar might possibly get injured more times than NIS) might be more scary and/or stressful. For this reason, we carried out a second analysis in which we fairly compared the 180 s of NIS with the first 180 s of IS and performed again Welch-corrected t-tests to compare HR and SCL. The results of this second analysis do not support the alternative interpretation and confirm the findings of the study: the values for IS are again larger than NIS and the difference between the two conditions is statistically significant for both HR (IS: $M = 7.38$, $SD = 11.25$; NIS: $M = -0.20$, $SD = 2.79$; $t(23.58) = 3.07$, $p < 0.01$, two-tailed) and SCL (IS: $M = 1.32$, $SD = 1.56$; NIS: $M = 0.31$, $SD = 0.30$; $t(22.53) = 2.95$, $p < 0.01$, two-tailed).

To further improve knowledge as well as severity and self-efficacy outcomes in IS with respect to NIS, a possible approach could be to introduce additional levels into the simulation to allow users to practice systematically and for multiple times the proposed recommendations, promoting repetitive rehearsal. Following such approach in NIS is instead not necessarily beneficial, because it would result in a very long passive and repetitive experience, in which it would be likely harder for the user to maintain attention engaged with respect to an interactive game.

In general, the observed results highlight the importance of using PMT to design and evaluate the proposed terror attack simulation. However, the set of audio-visual stimuli chosen to affect PMT variables was considered as a whole by our study. Therefore, the paper is not able to assess the effects of individual stimuli on PMT variables. To do so, more experiments are needed in which the single stimuli are included or removed from experimental conditions. The results from these studies would allow us to more deeply understand the specific effects of single stimuli on PMT variables and their possible statistical interaction with the interactivity variable. Another limitation of the current study is that participants' age range is relatively narrow (minimum: 18; maximum: 31; M = 23.68; SD = 3.28), and thus lacks a large demographic cross-section.

A further limitation of the study is that, although we carried out a deeper analysis of effects on users with respect to typical serious game evaluations, we did not explicitly assess one PMT variable, i.e., protection motivation. In the considered context of terror attacks, it would have been difficult to reliably measure it. Ideally, we should have put participants, after the simulation, in a real-world scenario that recreates the same terror attack simulation and evaluate their behaviors, which is impossible for safety reasons. Questionnaires would have been more practical, but would have required us to ask participants questions about their intention to pay attention to the considered threats that, if disregarded, prevent survival (as depicted in the simulation) and provide no kind of benefit. Such approach would have resulted in questions (e.g., "Do you intend to pay attention to smoke and toxic fumes during the evacuation?") that would be very suggestive of a positive response.

Another interesting aspect that was not evaluated concerns participants' engagement and enjoyment of the experience. Future studies of emergency preparedness simulations should include measures such as the ones proposed in (Brockmyer et al., 2009; Fu, Su, & Yu, 2009; Sweetser & Wyeth, 2005). Finally, as already mentioned, we did not assess self-identification and presence. Being in direct control of avatar's actions vs. passive spectator of avatar's behaviors, and the different phrasing of the cause-of-death messages that mirrors this difference, might have resulted in different levels of presence and self-identification, which might have played a role in increasing threat appraisal in the interactive condition, determining the observed differences between IS and NIS. A better assessment of participants' perceived sense of presence and self-identification would allow us to clarify if these variables can be considered as mediators in the relationship between interactivity and threat appraisal.

7. Conclusions and future work

In this paper, we have shown that the proposed simulations for emergency preparedness can be effective for learning as well as increasing the perception of risk severity and self-efficacy when they follow an interactive as well as a non-interactive format. Exploiting both formats is a strategy that deserves to be tested further (e.g., to assess uptake rates) in real educational contexts, because it could have the advantage of reaching a larger population that includes people who play video games as well as people who do not want to or cannot play them. An interesting aspect of this dual approach is that, once the contents for a serious game have been developed, they can be straightforwardly reused to generate the non-interactive simulation. The study also showed that interactivity enhanced the effects of the simulation on risk perception as well as the emotional response to the presented threat, which are desirable outcomes when the goal is to motivate people to protect themselves from danger.

Another contribution of the paper is that we showed how psychological models of protection motivation can provide guidance to the design as well as the evaluation of serious games for emergency preparedness. PMT guided us in designing and evaluating the proposed serious game, providing a theoretical grounding which is lacking in the literature on serious games for emergency preparedness.

In addition to the new studies sketched in Section 6, our future work will consider a new version of the simulation that employs a third-person visual perspective instead of a first-person one, to explore how this difference affects the measured variables (knowledge as well as PMT) and to compare the results with the literature that advocates the use of first-person with respect to third-person (e.g., Kallinen, Salminen, Ravaja, Kedzior, & Sääksjärvi, 2007). We also want to further improve the proposed serious game, focusing more on gaming elements related to users' coping appraisal such as improving recommendation presentation. In the simulations proposed in this paper, recommendations were presented to users as short sentences that interrupted for a few seconds the simulation. Presentation techniques that do not interrupt the simulation (and may also further enhance users' perceived sense of presence and self-identification with the avatar) will be evaluated, such as the use of a non-player character (NPC) that accompanies players during the evacuation and can help them by providing recommendations through speech and showing them how to put recommendations into practice. Since our serious game is meant to be played as a single-player game (that can be used without assistance, in a classroom as well as home contexts), we will consider how to extend it in such a way that multiple players can collaborate in the evacuation, possibly replacing NPCs with other players. In this way, players who have a greater experience with the simulation can help novice users during the game play, and players with similar experience level can collaborate in dealing with threats presented by the game. We will also introduce additional levels into the simulation to promote repetitive rehearsal of the recommendations in different kinds of emergencies. These enhancements to the serious game could significantly affect how the game is perceived by players and what is learned from it, possibly influencing factors that the current study did not consider, such as players' attitude towards prosocial behaviors (i.e., voluntary behaviors intended to benefit others) (Eisenberg, Fabes, & Spinrad, 2007). As reported in (Fischer et al., 2011), the literature has shown that video games with prosocial content appear to increase prosocial cognition and affect, which in turn evokes prosocial behavior. Finally, we will design a longitudinal study to better evaluate the long-term effects in knowledge and attitude retention of the improved interactive and non-interactive simulations.

References

Andreassi, J. L. (2007). *Psychophysiology*. (5th ed.). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Bachorowski, Jo-Anne, & Owren, M.J. (2008). Vocal expressions of emotions. In M. Lewis, J.M. Haviland-Jones & L. Feldman Barret (Eds.), *Handbook of Emotions* (pp. 709-729). New York, NY, USA: The Guilford Press.

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York, NY, USA: Freeman.

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: an overview. *Advances in Human-Computer Interaction – Special Issue on User Assessment in Serious Games and Technology-Enhanced Learning, 2013*(1), Article ID 136864.

Benedek, M. & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology, 47*(4), 647-658.

van den Bos K., Ham J., Lind E. A., Simonis M., van Essen W. J., Rijpkema M. (2008). Justice and the human alarm system: the impact of exclamation points and flashing lights on the justice judgment process, *Journal of Experimental Social Psychology, 44*(2), 201–219.

Boucsein, W. (2006). *Electrodermal Activity*. New York, NY, USA: Springer.

Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K., & Pidruzny, J.N. (2009). The development of the Game Engagement Questionnaire: a measure of engagement in video game-playing. *Journal of Experimental Social Psychology, 45*(4), 624-634.

Chittaro L. (2012). Passengers' safety in aircraft evacuations: employing serious games to educate and persuade. In *Proceedings of the 7th International Conference on Persuasive Technologies (PERSUASIVE 2012),* LNCS 7284, 215–226, Heidelberg, DE: Springer.

Chittaro, L., Buttussi, F., & Zangrando, N. (2014). Desktop virtual reality for emergency preparedness: user evaluation of an aircraft ditching experience under different fear arousal conditions. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology (VRST 2014),* 141-150, New York, NY, USA: ACM.

Chittaro, L., & Ranon, R. (2009). Serious games for training occupants of a building in personal fire safety skills. In *Proceedings of ISEE First International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES '09)*, 76-83, Los Alamitos, CA, USA: ISEE Computer Society Press.

Chittaro, L., & Zangrando, N. (2010). The persuasive power of virtual reality: effects of simulated human distress on attitudes towards fire safety. In *Proceedings of the 5th International Conference on Persuasive Technologies (PERSUASIVE 2010),* LNCS 6137, 58-69, Berlin, DE: Springer.

Cohen, B. H. (2001). *Explaining Psychological Statistics* (2nd ed.). New York, NY, USA: John Wiley & Sons.

Cole, S.W., Yoo, D.J., & Knutson, B. (2012). Interactivity and reward-related neural activation during a serious videogame. *PLoS ONE, 7*(3): e33909.

Dekker, A., Champion, E. (2007). Please biofeed the zombies: enhancing the game-play and display of a horror game using biofeedback. In *Proceedings of DiGRA 2007 Conference: Situated Play (DiGRA 2007),* pp. 1–9.

Drachen, A., Nacke, L. E., Yannakakis, G., & Pedersen, A. L. (2010). Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games (SIGGRAPH'10)*, pp. 49-54.

Eisenberg, N., Fabes, R.A., & Spinrad, T.L. (2007). Prosocial Development. In N. Eisenberg, W. Damon & R.M. Lerner (Eds.), *Handbook of Child Psychology* (pp. 646-718). Hoboken, NJ, USA: John Wiley & Sons.

Fischer, P., Greitemeyer, T., Kastenmüller, A., Vogrincic, C., & Sauer, A. (2011). The effects of risk-glorifying media exposure on risk-positive cognitions, emotions, and behaviors: a meta-analytic review. *Psychological Bulletin, 137*(3), 367-390.

Floyd, D. L., Prentice-Dunn, S., & Rogers, R.W. (2000). A meta-analysis of research on Protection Motivation Theory. *Journal of Applied Social Psychology, 30*(2), 407-429.

Fu, F.L., Su, R.C., & Yu, S.C. (2009). EGameFlow: a scale to measure learners' enjoyment of e-learning games. *Computers & Education, 52*(1), 101-112.

Girard, C., Ecalle, J., & Magnant, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning 29*(3), 207-219.

de Hoog, N., Stroebe, W., & de Wit, J. B. F. (2008). The processing of fear-arousing communications: how biased processing lead to persuasion. *Social Influence, 3*(2), 84-113.

Kallinen, K., Salminen, M., Ravaja, N., Kedzior, R., & Sääksjärvi, M. (2007). Presence and emotion in computer game players during 1st person vs. 3rd person playing view: evidence from self-report, eye-tracking, and facial muscle activity data. In *Proceedings of the 10th Annual International Workshop on Presence*, pp. 187-190.

Kim, K., Rosenthal, M. Z., Zielinski, D. J., & Brady, R. (2014). Effects of virtual environment platforms on emotional responses. *Computer Methods and Programs in Biomedicine 113*(3), 882-893.

Kobes, M., Oberijé, N., & Groenewegen-Ter Morsche, K. (2009). Serious gaming for behavioural assessment and research in case of emergency. Paper presented at *SimTecT 2009 Simulation Conference and Exhibition*.

Maibach, E.W., & Parrott, R.L. (1995). *Designing Health Messages: Approaches from Communication Theory and Public Health Practice.* Thousand Oaks, CA, USA: SAGE Publications.

Mewborn, C. R., & Rogers, R. W. (1979). Effects of threatening and reassuring components of fear appeals on physiological and verbal measures of emotion and attitudes. *Journal of Experimental Social Psychology 15*(3), 242-253.

Milne, S., Sheeran, P., & Orbell, S. (2000). Prediction and intervention in health-related behavior: a meta-analytic review of Protection Motivation Theory. *Journal of Applied Social Psychology 30*(1), 106-143.

Nacke, L., & Lindley, C. A. (2008). Flow and immersion in first-person shooters: measuring the player's gameplay experience. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share (FuturePlay 2008)*, pp. 81-88.

Norman, P., Boer, H., & Seydel, E. R. (2005). Protection Motivation Theory. In M. Conner & P. Norman (Eds.), *Predicting Health Behaviour: Research and Practice with Social Cognition Models* (pp. 81-126). Maidenhead, UK: Open University Press.

Öhman, A. (2008). Fear and anxiety: overlaps and dissociations. In M. Lewis, J.M. Haviland-Jones & L. Feldman Barret (Eds.), *Handbook of Emotions* (pp. 709-729). New York, NY, USA: The Guilford Press.

Park, N., Lee, K.M., Jin, S.A., & Kang, S. (2010). Effects of pre-game stories on feelings of presence and evaluation of computer games. *International Journal of Human-Computer Studies, 68*(11), 822-833.

Parsons, T. D., Courtney, C., Cosand, L., Iyer, A., Rizzo, A. A., & Oie, K. (2009). Assessment of psychophysiological differences of West Point cadets and civilian controls immersed within a virtual environment. In: *Proceedings of the 5th International conference on foundations of augmented cognition: neuroergonomics and operational neuroscience (FAC 2009), LNAI 5638 (pp. 514-523).* Heidelberg, DE: Springer.

Peng, W. (2008). The mediational role of identification in the relationship between experience mode and self-efficacy: enactive role-playing versus passive observation. *CyberPsychology & Behavior, 11*(6), 649-652.

Riva, G., & Mantovani, F. (2012). From the body to the tools and back: a general framework for presence in mediated interactions. *Interacting with Computers 24*(4), 203-210.

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., & Alcaniz, M. (2007). Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology and Behavior 10*(1), 45-56.

Ribeiro, J., Almeida, J. E., Rossetti, R. J. F., Coelho, A., & Coelho, A. L. (2012). Towards a serious game evacuation simulator. In *Proceedings of the 26th European Conference in Modelling and Simulation (ECMS '12)*, pp. 697-702.

Ritterfeld, U., Shen, C., Wang, H., Nocera, L., & Wong, W.L. (2009). Multimodality and interactivity: connecting properties of serious games with educational outcomes. *CyberPsychology and Behavior, 12*(6), 691-697.

Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change. *The Journal of Psychology 91*(1), 93-114.

Rogers, R. W. (1983). Cognitive and physiological processes in fear appeals and attitude change: a revised theory of protection motivation. In B. L. Cacioppo & L. L. Petty (Eds.), *Social psychophysiology: a sourcebook* (pp. 153-176). London, UK: Guilford.

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, *Measures in health psychology: A user's portfolio* (pp. 35-37). Windsor, UK: NFER-NELSON.

Silva, J. F. M., Almeida, J. E., Pereira, A., Rossetti, R. J. F., & Coelho, A. L. (2013). Preliminary experiments with EVA – serious games virtual fire drill simulator. In *Proceedings of the 27th European Conference in Modelling and Simulation (ECMS '13)*, 221-227.

Silva, J. F., Almeida, J. E., Rossetti R. J. F., & Coelho, A. L. (2013). A serious game for EVAcuation training. In *Proceedings of the ISEE 2nd International Conference on Serious Games and Applications for Health (SeGAH '13)*, 1-6.

Smith, S. P., & Trenholme, D. (2009). Rapid prototyping a virtual fire drill environment using computer game technology. *Fire Safety Journal 44*(4), 559-569.

Sweetser, P., & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment, 3*(3), 3-34.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer 38*(9), 25-32.