Don’t worry, be active: how to facilitate the detection of errors in immersive virtual environments

Sara Rigutti Corresp., [1] , **Marta Stragà** [1] , **Marco Jez** [2] , **Giulio Baldassi** [1] , **Andrea Carnaghi** [1] , **Piero Miceu** [2] , **Carlo Fantoni** Corresp. [1]

[1] Department of Life Sciences, Psychology Unit "Gaetano Kanizsa", University of Trieste, Trieste, Italy
[2] Area Science Park, Arsenal S.r.L, Trieste, Italy

Corresponding Authors: Sara Rigutti, Carlo Fantoni
Email address: srigutti@units.it, cfantoni@units.it

The current research aims to study the link between the type of vision experienced in a collaborative immersive virtual environment (active vs. multiple passive), the type of error one looks for during a cooperative multi-user exploration of a design project (affordance vs. perceptual violations), and the type of setting in which multi-user perform (natural in Experiment 1 vs. controlled in Experiment 2). The relevance of this link is backed by the lack of conclusive evidence on an active vs. passive vision advantage in cooperative search tasks within software based on immersive virtual reality. Using an ecologically valid yoking paradigm, we found that the likelihood of error detection in a complex 3D environment was characterized by an active vs. multi-passive viewing advantage depending on:

(1) The degree of knowledge dependence of the Type of Error the passive/active observers were looking for (*low* for perceptual violations, vs. *high* for affordance violations), as the advantage tended to manifest itself irrespectively from the setting for affordance, but not for perceptual violations;

(2) The degree of social desirability induced by the setting in which the task was performed, as the advantage occurred irrespectively from the Type of Error in the controlled (Experiment 2) but not in the natural (Experiment 1) setting.

Results are relevant to future development of cooperative software based on immersive virtual reality used for supporting the design review. A multi-user design review experience in which designers, engineers and end-users, all cooperate actively within the immersive virtual reality wearing their own head mounted display, seems more suitable for the detection of relevant errors, than standard systems characterized by a mixed usage of active and passive viewing.

Don't worry, be active: how to facilitate the detection of errors in immersive virtual environments

Sara Rigutti[1*], Marta Stragà[1*], Marco Jez[2], Giulio Baldassi[1], Andrea Carnaghi[1], Piero Miceu[2], Carlo Fantoni[1*]

[1] Department of Life Sciences, Psychology Unit "Gaetano Kanizsa," University of Trieste, Trieste, Italy

[2] Arsenal S.r.L., Area Science Park, Trieste, Italy

*Equally contributing authors

Corresponding Author:

Carlo Fantoni[1]

via Weiss 21, Trieste, 34128, Italy

Email address: cfantoni@units.it

Abstract

The current research aims to study the link between the type of vision experienced in a collaborative immersive virtual environment (active vs. multiple passive), the type of error one looks for during a cooperative multi-user exploration of a design project (affordance vs. perceptual violations), and the type of setting in which multi-user perform (natural in Experiment 1 vs. controlled in Experiment 2). The relevance of this link is backed by the lack of conclusive evidence on an active vs. passive vision advantage in cooperative search tasks within software based on immersive virtual reality. Using an ecologically valid yoking paradigm, we found that the likelihood of error detection in a complex 3D environment was characterized by an active vs. multi-passive viewing advantage depending on:

(1) The degree of knowledge dependence of the Type of Error the passive/active observers were looking for (*low* for perceptual violations, vs. *high* for affordance violations), as the advantage tended to manifest itself irrespectively from the setting for affordance, but not for perceptual violations;

(2) The degree of social desirability induced by the setting in which the task was performed, as the advantage occurred irrespectively from the Type of Error in the controlled (Experiment 2) but not in the natural (Experiment 1) setting.

Results are relevant to future development of cooperative software based on immersive virtual reality used for supporting the design review. A multi-user design review experience in which designers, engineers and end-users, all cooperate actively within the immersive virtual reality wearing their own head mounted display, seems more suitable for the detection of relevant errors, than standard systems characterized by a mixed usage of active and passive viewing.

1. Introduction

To achieve an efficient visualization of the external world functional to the detection of relevant 3D environmental features, the brain has to integrate retinal information with extraretinal and proprioceptive information about the observer's ego-motion (Wallach, 1987; Braunstein & Tittle, 1988; Ono & Steinbach, 1990; Wexler, 2003; Fetsch et al. 2007; Fantoni, Caudek, & Domini 2010). Only when the changes in retinal projections are accurately accounted for the sensed ego-motion information, a stable perception of the environment can be established, and environmental features can be efficiently detected. To what extent the brain integrates retinal and extra-retinal information, and whether such integration provides performance advantages in ecologically valid tasks when interacting within a complex 3D environment is still under debate (Jaekl, Jenkin, & Harris, 2004; Liu, Ward & Markall, 2007; Teramoto & Riecke, 2010; Chrastil & Warren, 2012; Fantoni, Caudek & Domini, 2014; Bülthoff, Mohler, & Thornton, 2018). In order to provide an empirical answer to this question, we rely on Immersive Virtual Reality (IVR), and compare active vs. passive seeking performance for different types of design errors. In two experiments we contrasted participants' performance resulting from the dynamic stereoscopic/cyclopean view of a complex 3D environment, during either an active exploration (i.e., active viewing condition) or a passive replay of the exact same optic information, which was self-generated by the active observer (i.e., passive viewing condition).

1.1. Passive and active observers in standard collaborative IVR systems

IVR offers a new human-computer interaction paradigm, in which users actively participate in, and interact with a computer-generated environment. The immersive experience offered by technologies, such as CAVE systems or Head Mounted Displays (HMD), leads to a sense of presence in the virtual environment that is experienced to some extent as it were real (e.g., Bowman & McMahan, 2007). IVR opens up the possibility to reproduce a variety of contexts that in the real world would be risky, costly or unreachable. For this reason, IVR has been successfully applied to a variety of fields, beyond the entertainment industry, like gaming (Durlach & Mavor, 1995; Brooks, 1999; Schell & Schochet, 2001). Traditionally, IVR systems have been developed for single users, following a user-centred design approach. The user wears the HMD, interacts within immersive virtual environment and self-generates a 3D dynamic optic information by controlling the viewpoint motion. Viewpoint motion can be obtained by a combination of ego-motion and control devices generally guided by hand (i.e., SpaceMouse, Oculus Touch and Leap Motion). The user-centred approach has been relevant to support and improve the immersive experience in different contexts, such as design, aesthetics, and 3D objects' visualization (Bayon, Griffiths & Wilson, 2006; Chen, 2014). However, it disregards the possible co-presence of more users that share the same immersive virtual environment, as in the case of cooperative working activities. The co-presence of the users in immersive virtual environments is typical of design review activities, in architecture, engineering and shipbuilding. In these working environments, a multi-user IVR system is required to support the activity of design review participants, which is mainly based on the collaborative recovery of design errors (Fernández & Alonso, 2015). Nowadays an optimal collaborative IVR system, in which all participants in the design review task wear their own HMD, is quite far from being used in real working environments for technological and economical drawbacks (Chen et al., 2015). To overcome this issue, standard collaborative IVR systems are based on a mixed usage of active and passive viewing and supported by HMD and projection screens respectively (e.g., Bayon, Griffiths, & Wilson, 2006). The visual immersive experience is generated by one observer who wears the HMD (i. e., the project leader of the design review session) and moves his point of view by body movements (encoded by HMD's translations and rotations), and manual control (encoded by different types of devices). His/her virtual experience is paralleled in real-time on a wide screen for collaborators. In this kind of collaborative IVR system, a single participant actively views the scene through self-controlling the viewpoint translation within the immersive virtual environment, while all remaining participants passively observe the scene from an external viewpoint (Bayon, Griffiths, & Wilson, 2006; Shao, Robotham, & Hon, 2012). This mixed usage of simultaneous active and multiple passive viewing constitutes the standard system for remote collaboration nowadays, in which users from different geographical locations share virtual environments (Jones, Naef, & McLundie, 2006; Bassanino et al., 2010; Chen et al., 2015; Fleury et al., 2015; Tanaya et al., 2017). Although this mixed setting provides a sustainable technological solution, it is characterized by a drawback. Indeed, there is now considerable evidence suggesting that active and passive observers, although receiving similar visual input, rely on a different visual experience (Wallach, Stanton & Becker, 1974; Braunstein & Tittle, 1988; Wexler, Lamouret & Droulez, 2001; Fantoni, Caudek & Domini, 2010; Caudek, Fantoni, & Domini, 2011; Fantoni, Caudek & Domini, 2012; Fantoni, Caudek & Domini, 2014). This difference is known to be due to the dissimilar sensori-motor and cognitive information resulting from an active exploration of the spatial layout relative to the passive observation, see supplemental text in Text S1 for further details (Sherrington, 1906; Ono & Steinbach, 1990; Wilson et al., 1997; Chance et al., 1998; Christou & Bülthoff, 1999; Harman, Humphrey & Goodale, 1999; Wang & Simons, 1999; Ujike & Ono, 2001; James et al., 2002; Peh et al., 2002; von Helmholtz, 2002; Wilson & Péruch, 2002; Wexler, 2003; Jaekl, Jenkin, & Harris, 2004; Naji & Freeman, 2004; Waller, Loomis & Haun, 2004; Ono & Ujike, 2005; Wexler & van Boxtel, 2005; Colas et al., 2007; Liu, Ward & Markall, 2007; Waller & Greenauer, 2007; Fantoni, Caudek, & Domini 2010; Riecke et al., 2010; Teramoto & Riecke, 2010; Caudek, Fantoni & Domini, 2011; Meijer & Van der Lubbe, 2011; Ruddle, Volkova & Bülthoff, 2011; Ruddle et al., 2011; Chrastil & Warren, 2012; Fantoni, Caudek & Domini, 2012; Fantoni, Caudek & Domini, 2014; Bülthoff, Mohler, & Thornton, 2018). On the basis of these claims, we formulated the main expectation at the basis of our study: the compatibility between retinal and extra-retinal information resulting from ego-motion, together with agency and intentionality, that are at the basis of active but not of passive viewing, which ultimately mimics a standard collaborative IVR environment, should enhance the allocation of attention to relevant features of the 3D spatial layout. We tested this expectation comparing a possible bolstering effect of active vs. passive viewing in collaborative IVR systems based on the mixed usage of simultaneous active and multiple passive viewing. So far, it remains unclear whether this type of advantage occurs in task used in cooperative software based on IVR (Bayon, Griffiths & Wilson, 2006).

2. General Method

We addressed our aim by using, for the first time, an ecologically valid task relevant for the design review of large scale digital mock-ups, like those employed in ship design (Jones, Naef, and McLundie, 2006; Fernández & Alonso, 2015): namely a modified *hide and seek* task (Chandler, Fritz, & Hala, 1989). In particular, independent groups, which comprised both an active and multiple passive participants, were asked to simultaneously seek for different types of design errors in two different views, namely a dynamic stereoscopic (for the active observer) and dynamic cyclopean (for the passive observers) of a 3D ship corridor. These views resulted either from an active exploration of the IVR (in active viewing condition) or from the passive replay of the exact same optic information self-generated by the active observer (in passive viewing condition). The active observer self-generated the 3D view combining head (through HMD rotations and translations) and hand (through SpaceMouse controller) movements from a sitting position which was similar to the one adopted by passive observers. As the observers performed the task in a multi-user modality, we encoded the number of detected errors at the end of the exploration/observation session (lasting 4 minutes). A series of screenshots (nine in Experiment 1 or ten in Experiment 2, with only eight of them referring to actual design errors and the remaining serving as *catch* trials) were shown to the group of participants. After each screenshot, participants were asked to raise their hand to indicate whether the screenshot included a design error they saw during the exploration phase.

Importantly, our modified *hide and seek* task, beyond being applicable in both natural (i.e., in the form of demonstrative group game during a science and technology exhibit as in Experiment 1), and controlled settings (i.e., in the form of group experiment in a laboratory setting as in Experiment 2), involves relevant visual and cognitive components generally characterizing cooperative technological applications based on IVR. First, it involves visual spatial attention, which is required to perform a visual search of design errors within a complex IVR scene. Visual spatial attention with active exploration of the actively generated/passively observed visual layout regulates the detection of a target feature (i.e., the error) among all potential distracters. Error detection is often involved in cooperative multi-user activity as in the case of design review sessions in architecture, engineering and shipbuilding domains (Fernández & Alonso, 2015). Secondly, the detection of errors within a complex 3D environment implied that the observer must form a representation of the detected object (i.e., erroneous), retrieve from memory a conventional representation of the object and finally compare the two representations to decide whether the detected object is erroneous or not (MacKay & James, 2009).

As far as we know, no studies, to date have tested the effect of active vs. multi-passive viewing in such a type of task when different types of design errors are involved in the 3D layout. To this purpose, we followed the Norman's categorization of design errors in artifacts (Norman, 1988) and included two different types of design errors in our task with different degree of dependence on knowledge about the environment structure (Fig. 1):

(1) Affordance errors in Fig. 1A (AE, high degree), based on affordance violation such as a door handle in the same side of the hinges;

(2) Perceptual errors in Fig. 1B (PE, low degree), based on violation of perceptual organization principles like good continuation (such as a misaligned handrail) and colour similarity (such as a blue handrail embedded within a layout of yellow handrail).

According to the Norman~~'s categorization~~ (1988), the detection of AE is a knowledge-based process. Indeed, the detection of AE results from incongruence between the mental interpretations of artifacts, based on our past knowledge and the experience applied to our perception of an artifact. By contrast, the detection of PE is perceptual based, as it is knowledge independent and results from incongruence directly stemming from the level of perceptual attributes, which are automatically processed, thus not requiring any learning process (i.e., a colour or shape discontinuity in the 3D layout). An AE type of error can thus be detected if the observer possessed knowledge about the conventional structure of the target object (e.g., the observer must know how a door looks like and how it works in order to recognize that the door handle is misplaced). Otherwise, a PE type of error can be detected even in absence of knowledge about target object structure (e.g., the observer does not need to know how the handrail should look like in conventional setting in order to detect the error). Hence, we expected that the likelihood of detecting an error in our task would depend on the Type of Error (AE vs. PE) irrespectively from Viewing Conditions (active or passive). The likelihood of detecting AEs was expected to be smaller than the likelihood of detecting PEs (Expectation 1).

We selected the immersive virtual environment from a section of a digital mock-up of a ship, in order to keep our task ecologically valid for cooperative software based on IVR. In particular, we choose a 3D ship corridor suitable for real multi-user session of ship design review (Fig. 1D). The structure of the immersive virtual environment was chosen so that the exploration path of the active observers was sequential relative to the ordering of appearance of design errors. This structure assured us that during the hide and seek task different active observers would have: (1) followed similar pathways; (2) passed at least once through all of the relevant places needed for the potential detection of all errors.

In line with the procedures outlined by previous studies, we employed a new adapted version of a yoked design. In yoked design, the passive viewing is often obtained replaying the active exploration of the same scene generated by another participant, or by the same participant (Rogers & Rogers, 1992; Harman, Humphrey & Goodale, 1999; James et al., 2002). Differently from this conventional design, our adapted design mimicked the viewing conditions typical of collaborative IVR systems based on the mixed usage of simultaneous active and multiple passive viewings. Therefore in our design, the active exploration and the multiple passive observations took place simultaneously (Meijer & Van der Lubbe, 2011). The group of passive participants observed on a large screen the dynamic layout generated by the active observer in real-time. This design assures that, within each group, in the active and passive conditions the participants rely on the exact same visual input. This design overcomes the problem of the unbalanced visual exposure of different types of viewers (i.e., active and passive) which is typical of yoked design, in which the participant passively views the optic array that results from the active exploration of a scene, which is generated by another (active) participant (Chrastil & Warren, 2012). Furthermore, differently from traditional yoked design, in which the passive view is replayed just after the recording of the active exploration (Rogers & Rogers, 1992; Wexler, Lamouret, & Droulez, 2001; Wexler, 2003; Fantoni, Caudek, Domini, 2010), our simultaneous active and passive exposition to the dynamic 3D scene prevents us from potential threats to external and internal validity. Indeed, our simultaneous viewing involves a balanced (not unbalanced) temporal ordering of the viewing conditions.

Finally, we purposely unbalanced the amount of sources of information at which the active and passive viewers had access during our task, in order to assess the main ~~applicative~~ aim of our study: to compare the effectiveness of different type of viewing resulting from collaborative IVR systems based on mixed usage of passive and active viewing. Our passive observers (subdivided in small groups) simultaneously viewed the same dynamic scene self-generated from a single active observer (Fig. 2), which was aided by stereoscopic vision (through HMD), extra-retinal and proprioceptive information regarding self-motion derived from head and hand movements, and cognitive control (see supplemental text in Text S1 for further details). We thus expected (Expectation 2) active viewing to be associated with a larger likelihood to detect design errors than passive viewing. This should result in an active vs. multi-passive error detection advantage. Notably, an opposite expectation occurs if the different amount of cognitive load involved in active exploration over passive observation is considered (Expectation 2b). In particular, the active exploration (relative to the passive observation), including demanding and distracting operation to intentionally move and translate the viewpoint, should result in a reduction of the attentional resources needed to perform the search task (Liu, Ward & Markall, 2007; see 3rd paragraph in Text S1 for further details).

We validate the generalizability of the expected active vs. multi-passive advantage by manipulating the type of setting through Experiments, with Experiment 1 involving a natural vs. Experiment 2 a controlled setting. Such a manipulation allowed us to identify the limits within which the active vs. multi-passive advantage should fall. In particular, natural settings are known to induce social desirability biases more than controlled settings as an instance of the reactivity to the experimental situation (Crowne & Marlowe 1964; Fisher, 1993; Shadish, Cook & Campbell, 2002). Such reactivity in our task might lead to a ceiling effect given that the most socially desirable response was the one hitting not missing the error.

Both Experiments were approved by the Research Ethics Committee of the University of Trieste (approval number 84) in compliance with national legislation, the Ethical Code of the Italian Association of Psychology, and the Code of Ethical Principles for Medical Research Involving Human Subjects of the World Medical Association (Declaration of Helsinki). All participants provided their oral informed consent. The request of oral consent, formulated by the experimenter, made explicit that people not willing to participate in the session should simply not participate or not respond, without any consequence, and emphasis was put on the anonymous treatment of data which was part of group instructions at the beginning of session. Written consent (implying identification of every respondent) was redundant, given that the Experiments were carried out with visitors' groups scheduled according to the program of the science exhibit within the natural setting in Experiment 1 and with classes of psychology courses within the controlled setting in Experiment 2. All participants present during the data collection sessions accepted to respond. The Ethics Committee of the University of Trieste thus approved the informed consent. Dataset is available as (Data S1).

2.1. Experiment 1: Hide and seek in a natural setting

2.1.1. Participants

Experimental sessions were conducted in a natural setting during a science and technology exhibit at Trieste Next 2017 of a custom made IVR system used to support ship design review. Visitors and scholars served as participants according to a well-defined scheduling of the science and technology exhibit lasting two days (self-selected sampling). All had normal or corrected to normal vision and were naive as to the purpose of the experiment. None of them reported specific visual diseases relevant for our task, like ocular dominance, colour blindness, and severe forms of eyestrain. Ninety-eight participants took part in the Study (Mean age = 14.42, *SD* = 2.76, range [12, 20]), subdivided in 10 small groups of variable sizes ranging from a minimum number of 5 to a maximum number of 12. The groups were recruited one after the other during the exhibit. Six groups were composed by middle-school students (*n* = 61), two groups by high-school students (n = 22) and two by undergraduate students (n = 15). A total of 10 active participants vs. 88 passive participants were thus recruited, being each group composed by only one active participant. The active observer was selected among those participants who were willing to wear HMD, reported some previous experience with 3D gaming and low levels of motion sickness. This was assessed by orally asking them about previous experience of motion sickness symptoms according to Kennedy et al. (1993): oculomotor (eyestrain, difficulty focusing, blurred vision, and headache), disorientation (dizziness, vertigo), and nausea (stomach awareness, increased salivation and burping).

2.1.2. Apparatus and Stimuli

The active participants' head motions were tracked in real time using an Oculus Rift CV1 Head Mounted Display (with a FOV of 93° x 104° and a frame rate of 90fps) connected with a PC equipped with an Intel Core i7 7700K 4.20 GHz processor with 16GB RAM and nVIDIA GeForce GTX 1080 Ti graphics card. The PC controlled the simultaneous active and passive visualization of our complex immersive virtual environment and sampled the tracker. The immersive virtual environment was rendered through an IVR system designed by Arsenal S.r.L and was updated in real time according to a combination of head and hand movements of the active observer (Jez et al., 2018). In particular, the translation/rotation of the point of view through the immersive virtual environment was supported by the combination of head movements and manual control through a 3D Connexion SpaceMouse Pro, with 4 Degrees Of Freedom (DOF; i.e., transition in the directions of three directional axes – forward/backward, left/right, up/down – and rotation on the vertical axis). Accelerometers within the HMD together with two camera motion sensors converging with an angle of about 40° on the active observer's position at rest (at a distance of about 1.5 m) were used to calculate the x, y, z coordinates of the observers viewpoint. These coordinates defined the head movements of the active observers and were used by our program to update in real time the geometrical projection of the 3D graphic. A precise visualization of the virtual environment was achieved by carefully calibrating the 3D layout to the height and the inter-ocular distance of each active observer following the Oculus Rift procedure. The virtual reality experience generated by the HMD was paralleled on two large LCD screens (Samsung 50" J5200 Full HD LED set at a screen resolution of 1024 X 768 pixels) connected with the PC. The two LCDs simultaneously displayed a cyclopean view of the exact same dynamic optic information generated by the movements of the active observer within the IVR.

Figure 1 provides a bird eye view of the virtual environment (Fig. 1D) used for the *hide and seek* task together with the design errors implemented throughout the path (screenshots, in Fig. 1A and 1B, for AEs and PEs, respectively). Our virtual environment consisted of a custom-made digital mock-up of an L-shaped 3D ship corridor. The corridor was 2.8 m high, 2.4 m large, and was composed by two straight segments: a first 35 m long segment connecting the starting position to the corridor's curve (Fig. 1D, horizontal part of the corridor), and a second 40 m segment connecting the corridor's curve to the corridor's end (Fig. 1D, vertical part of the corridor). The corridor included different features, like: floor texture (a green floor with lateral yellow stripes and horizontal arrows indicating the directions), long pipes on the ceiling (with different sizes and colours), doors (five along the wall and one louvered door in the middle of the corridor), security systems (such as fire extinguisher, warning lights), signals (e.g., "fire point", exit), handrails, bollards, and several technical equipment. The screenshots of the eight design errors actually included in the digital mock-up of the corridor (as presented to the participants to the experiment during the response encoding phase) are shown in Fig. 1A and 1B, together with the screenshot used as a catch trial (Fig. 1C). The catch trial was a misplaced pipe holder that was not implemented as an actual error in the digital mock-up of the corridor though being visually consistent with the features of the corridor. Within our experimental design, valid screenshots (i.e., those depicting actual errors) were equally subdivided into the two types: AE and PE (Fig. 1A and 1B, respectively).

Figure 1. Screenshots of design errors and the immersive virtual environment used throughout Experiment 1 and Experiment 2. A and B: screenshots of the 8 actual design errors implemented in the digital mock-up of the corridor as presented to our participants during the response encoding phase, subdivided into the two types: affordance violations in A, and perceptual violations in B. Numbers indicate the relative ordering of appearance of violations along the immersive virtual environment explored during the task from the starting position to the corridor's end (in D). C: the two screenshots used as catch trials shown to the participants in an intermixed and randomized order together with the screenshots of the actual design errors used during the response-encoding phase of the experiments. Experiment 1 included the presentation of the only catch trial screenshot 1, Experiment 2 included the presentation of the both catch trials' screenshots (1 & 2). D: a bird eye view of the immersive virtual environment, from the starting position (coded by the blue star) to the corridor's end (orange circles stand for design errors). The immersive virtual environment was the rendering of a digital mock-up of an L-shaped 3D ship corridor along which the 8 design errors were sequentially implemented along the pathway the observer was required to travel (the numbering corresponds to their relative ordering of appearance along the pathway).

AE included (Fig. 1A) a door handle placed too high and in the same side of the hinges (AE1), a fire extinguisher placed too high (AE2), a door handle placed too low and in the same side of the hinges (AE3), and a door handle in the same side of the hinges (AE4). PE included (Fig. 1B) a misaligned bollard (PE1), blue handrail embedded within yellow handrails (PE2), a louvered door missing an half (PE3), and a misaligned handrail (PE4). Three independent judges categorized the eight errors into the two categories, with an inter-rater agreement of 94%. By combining our two types of valid screenshots (AE/PE) with the two Viewing Conditions (active/passive) we obtained four experimental conditions of a mixed factorial design.

As shown in Fig. 1D, the design errors depicted within the screenshots were distributed all along the length of the 3D ship corridor, at a distance from the starting point of 16 m (AE1), 25 m (PE1), 31.5 m (AE2), 40 m (PE2), 49.5 m (PE3), 52 m (AE3), 53 m (PE4), 63 m (AE4).

The spatial arrangement of the setting is schematized in Fig. 2A, with Fig. 2B showing a photograph of the actual exposition context taken during a training session on a digital mock-up different from the one used during the experiment. The active observer comfortably sat on a stool right in between the two LCD screens, whereas the passive observers were arranged in front of the LCD screens comfortably sat on stools at a viewing distance of about 100 cm (Fig. 2A). A 90 cm wide desk was interposed in between the passive groups of participants and the screens in order to control at the best the viewing distance (observers were required to sit on the stools posing the elbows on the margin of the desk). At the viewing distance of 100 cm the retinal size subtended by the 3D dynamic layout when presented in full screen was of about 57.9° horizontally and 34.5° vertically.

Figure 2. Experimental setting of Experiment 1. A: schematic view of the experimental setting with three observers (one active, green arrow, and two passives, blue arrows) implemented in the natural condition during the science and technology exhibit of Experiment 1, with superposed the minimal distance the passive observers stand (as constrained by the 90 cm desk size), and their average viewing distance (100 cm when sitting with their elbows on the margin of the desk). This distance is taken from the 2 large LCD screens (127 cm diagonal), displaying in real time the exact same 3D, though monoscopic, view the active observer self-generated combining head movements and SpaceMouse control. In this example, the view produced by the active observer that the 3 passive observers are looking at in real time is consistent with PE1. B: a photograph of the setting during the training session with the digital mock-up (a ship thruster) used to familiarize the active observer with the SpaceMouse and the passive observers with the 3D graphic. Notice that the setting of Experiment 2 reproduced in smaller scale the one shown in the current scheme, including smaller screens though smaller viewing distances in order to equate the two Experiments for the size of the passive displays in term of retinal sizes (57.9° × 34.5°). B photographed by the co-author Carlo Fantoni who gave his permission for publication.

2.1.3. Procedure

The experimental procedure included six phases lasting overall 10 minutes.

First, participants were informed about the experimental setting and the instructions of the modified *hide and seek* task using a Power Point presentation. The task was described as a *game*, in which the best-performing participant would be awarded with a gadget. This procedure was also aimed to motivate each single participant to perform at his/her best. To this purpose, an experimenter first explained to the entire group of participants the role they were about to endorse in the game. Participants were then introduced with their spatial disposition depending on the group they belonged: passive participants (one in front of the LCD to the left of the active observer, the other in front of the LCD to the right of the active observer, as by Fig. 2A) or active participant (right in between the two groups of passive participants, Fig. 2A). They were then informed about the aim of the game: "to search, find and memorize for design errors they would have encountered during the exploration of an immersive virtual environment". Participants were also informed that at the end of the exploration they would have been administered a brief serial presentation of screenshots. After each screenshot's presentation they would have been asked to raise their hand if the screenshot reported an error they saw during the exploration phase of the game.

Second, participants were then assigned to an active (n = 10) or passive (n = 88) role.

In the third phase, the group of participants was familiarized with the immersive virtual environment they were going to view during the game. During this familiarization phase, all participants watched on the LCD screen a 1-minute video clip showing the same 3D ship corridor they would have been exposed during the exploration phase of the game but without including design errors. Participants were instructed to carefully watch the video clip. In order to encourage subjects to be as accurate as possible in seeking design errors, the experimenter stressed that in the video clip design errors were absent. This information also provided the subjects with a hint on the design errors they would have successively encountered during the experimental phase.

Fourth, the active observer was briefly trained in the usage of the SpaceMouse to control the translation of his/her point of view. The training was performed keeping static the HMD and visualizing its view on the LCD screen. A digital mock-up different from the one used during the experimental phase was used for such a phase (see an example in Fig. 2B). The observer following the instruction of the experimenter was trained for about 1 minute on the main commands of the SpaceMouse (rotation, lateral and back/forth viewpoint translation). During this phase, in order to similarly train the passive observers with the 3D complex structure of the environment, the passive observers were required to look at the same scene the active observer was looking at.

Fifth, the experimental phase took place lasting about 5 minutes (1 minutes of instruction + 4 minutes of game). Active and passive observers were first required to keep their positions as by instruction (Fig. 2). The experimenter then focused the observers' attention on the experimental task and instructed them to memorize the errors they found throughout the exploration phase of the game without disclosing their identification. Participants were thus explicitly discouraged to use neither verbal (i.e., claims) nor non-verbal behaviour (i.e., gestures) as soon as they detected a design error. The exploration phase of the game then started: the active observer was required to begin moving freely along the 3D ship corridor (now including the eight design errors) in search for design violations. This phase was terminated after 4 minutes of exploration. On average, this duration led to travel along the corridor back and forth for about one time.

Finally, the sixth response-encoding phase took place. Participants were serially presented with nine screenshots: the eight design errors presented serially in the order they appeared along the path-way from the start position to the corridor's end plus the catch trial presented in a random serial position. As by initial instructions, participants were then required to raise their hands after each screenshot if the screenshot reported an error that they found during the 4 minutes exploration phase of the game: the experimenter registered the number of raised hands as well as the participants who raised their hand. At last, participants were thanked, debriefed and awarded with a gadget. Importantly, participants, throughout the game, were never aware of the number and types of design errors included in the 3D ship corridor.

2.1.4. Results

We analyzed the individual likelihood of detecting a design error as an index of performance in our modified *hide and seek* task following Knoblauch & Maloney (2012) by applying a generalized linear-mixed effect model (*glmm*) with a probit link function to the whole set of binary responses on valid screenshots (1 = error detected; 0 = error not detected). In particular, we used a *glmm* with the Type of Design Error (AE vs. PE) and the Viewing Condition (Active vs. Passive) as fixed factors, with by subject and by error (i.e., the eight valid screenshots) random intercepts and by subject random slope for Type of Design Error (Baayen, Davidson & Bates, 2008). Importantly, this type of analysis has been proved to be optimal for a research in which the distribution of responses and participants across conditions is inevitably unbalanced. Two-tailed *p*-values were obtained from type 3 F-statistics with the denominator’s degrees of freedom calculated by subtracting from the total number of observations the model’s degrees of freedom minus one. As indices of effect size we reported the partial eta-square (η^2_p) , and the concordance correlation coefficient, *rc*. Following Vonesh, Chinchilli & Pu (1996, but see also Rigutti, Fantoni & Gerbino, 2015) such a latter index provides a reliable measure, in the −1 to 1 range, of the degree of agreement between observed values and values predicted by generalized linear-mixed effect models. Finally, we reported Cohen's *d* as a standardized measure of significant difference between means.

Disregarding catch trials, the data analysis was based on the 100% of active participants' responses to valid trials (*n*= 80 resulting from the combination of 10 active observers and 8 errors), and 97% of passive participants' responses (*n*= 680 out of the total of 704 responses, resulting from 85 - 3 excluded from the total of 88 - passive observers and 8 errors).

We removed three passive observers from the analysis because of the application of two exclusion criteria both aimed at minimizing any possible biasing effect of social desirability intrinsic in our natural setting:

(1) relative the sample mean deviation, one passive participant achieved an individual error-detection proportion that deviates from the one of the active observer of the group he/she belonged more than ± 2.5 *SD*;

(2) two passive participants provided a positive response (i.e., raised their hand) after the catch trial presentation.

After the application of the exclusion criteria our total sample size (*n* = 95) was large enough to rely on reliable statistical conclusions. Indeed, it was larger than the critical sample size of 90 calculated for an experimental design such as the one we implemented considering a power of about 0.80 and a moderate effect size (*f*=0.30).

The pattern of average error detection proportion as a function of the Type of Design Error (x-axis) and Viewing Condition (colour coding as by the legend) shown in Fig. 3A is only in part consistent with our expectations. Expectation 2 (not 2b) indeed holds for AE but not for PE.

Figure 3. Error detection proportion in Experiment 1 (natural setting). A: Mean and SE values of error detection proportion in active (green) vs. passive (blue) viewing conditions as a function of the Type of Error (*affordance* vs. *perceptual*) on the abscissa. B-C: Mean and SE values of error detection proportion in active (green) vs. passive (blue) viewing for the four types of affordance errors (in B) and the four types of perceptual errors (in C). The numbers along the abscissa indicate the relative ordering of error's appearance along the exploration path of the 3D ship corridor, from the starting position to the end (same encoding of Fig. 1).

This was confirmed by the results of the *glmm* analysis (*rc*= 0.52, 95% CI [0.48, 0.55]) revealing an unexpected trend toward significance of the interaction between Type of Design Error and Viewing Condition ($F_{1, 751}$ = 3.90, *p* = 0.048, η^2_p = 0.005, 95% CI [0.000, 0.017]). Such a trend was due to the fact that the likelihood of detecting AEs was larger for active (*M* = 0.70 ± 0.07, with *M* indicating average proportion of error detection plus/minus one standard error), rather than passive (*M* = 0.53 ± 0.02) observers (consistent with Expectation 2, not with Expectation 2b). This difference produced a marginally significant *glmm* estimated active vs. multi-passive error detection advantage of about (0.22 ± 0.23, *z* = 1.61, one-tailed *p* = 0.053, *d*= 0.34). Differently, the likelihood of detecting PEs was similar for active (*M* = 0.77 ± 0.07) vs. passive (*M* = 0.80 ± 0.02) conditions thus producing a non significant active vs. multi-passive advantage (-0.03 ± 0.15; *z* = 0.43, one-tailed *p* = 0.332).

The *glmm* analysis further revealed that also Expectation 1 was only in part validated by our pattern of data. The facilitation for PE over AE predicted on the basis of Expectation 1 was confirmed only for passive, but not for active viewing conditions. For the passive viewing condition the likelihood of detecting an error was indeed smaller for AEs (*M* = 0.53 ± 0.02), rather than PEs (*M* = 0.80 ± 0.02) Type of Errors (*z* = -2.38, one-tailed *p* = 0.009, *d* = 0.60). No such a difference was found for the active viewing condition in which the likelihood of detecting an error was instead similar for AEs (*M* = 0.70 ± 0.07) and PEs (*M* = 0.77 ± 0.07) Type of Errors (*z* = 0.47, one-tailed *p* = 0.318).

A further post-hoc *glmm* analysis was carried out on the distribution of the likelihood of errors detection amongst our two types of errors (Fig. 3B and 3C for affordance and perceptual, respectively). This analysis revealed the active vs. multi-passive advantage for AE Type of Error was mainly due to two out of the four AEs: AE3 and AE4 (Fig. 3B). We indeed found a significant active vs. multi-passive error detection advantage for AE3 (*z* = -2.10, one-tailed *p* = 0.018, *d* = 0.71) and AE4 (*z* = -1.84, one-tailed *p* =0 .033, *d* = 0.64), but not for AE1(*z* = -1.18, one-tailed *p* = 0.120) and AE2 (*z* = 0.92, one-tailed *p* = 0.178). No reliable differences were instead found for any one of the four PEs (PE1: *z* = -0.07, one-tailed *p* = 0.470; PE2: *z* = 1.24, one-tailed *p* = 0.107; PE3: *z* = 0.37, one-tailed *p* = 0.354; PE4: *z* = -0.21, one-tailed *p* = 0.415; Fig. 3C).

2.1.5. Discussion

In the current Experiment, we found a trend toward an active vs. multi-passive viewing advantage with AEs but not PEs. As we hypothesized, the failure to find a difference between active and multi-passive observers in PEs can be the result of an overall ceiling effect. According to Expectation 1, such a ceiling was more likely to occur on PEs rather than AEs, given that PEs was expected to be easier to be detected than AEs. Such a response ceiling is likely to be induced by the type of natural setting used in the current experiment favouring positive (hand raised) rather than negative (hand not raised) responses. Positive responses were indeed more socially desirable than negative responses in our task, thus producing a social desirability bias.

Given these mixed results, the natural experimental setting and the heterogeneous sample that we employed in Experiment 1, we ran a second experiment using a more controlled setting.

2.2. Experiment 2: Hide and seek in a controlled setting

Would results be similar (as those of Experiment 1) when responses at the basis of error detection are less biased by social factors? In order to answer such a question, we performed Experiment 2 that was a replication of Experiment 1 in a VR laboratory (not a natural setting), with a sample of participants more uniform in term of age than the one used in Experiment 1, and with smaller groups of active/passive observers. Furthermore, relative to Experiment 1, in Experiment 2 we further controlled for social desirability bias, introducing:

(1) a method to validate the truthfulness of the errors reported by the active observers signaled by his/her screenshots;

(2) an additional catch trial, now reaching a 25% of catches over the total number of valid trials (*n*= 8);

(3) the removal of the ludic dimension as a context within which the participants were asked to perform the task (i.e., the task was not described as a game and participants were not awarded with gadgets for their performance).

Importantly, the new catch trial was intentionally extracted from an immersive virtual environment apparently different from the one tested during the exploration session. Such an apparent difference should implicitly inform the participants that the series of screenshots presented during the response-encoding phase did not necessarily include actual design errors. This should further reduce the tendency to respond according to social desirability.

2.2.1. Participants

One-hundred undergraduates of the University of Trieste (Mean age = 20.60, *SD* = 3.29, [18, 40] range, 77% female), all with normal or corrected to normal vision, participated in the experiment in return for course credits. According to our pre-experimental questioning (same as in Experiment 1) our participants did not suffer of any specific visual diseases relevant for our task (ocular dominance, colour blindness, and severe forms of eyestrain). Participants were subdivided in 17 groups with variable size, each of which ranged from three to eight. Groups were tested in individual sessions and were formed following the same procedure used in Experiment 1. Fourteen groups were composed by first year bachelor psychology students (*n*= 91), and three groups each composed by three participants included second year master psychology students (*n*= 9). Participants were assigned to active (*n* = 17) and passive (*n* = 83) conditions following the same selection criteria used in Experiment 1.

2.2.2. Apparatus and Stimuli

The experiment was conducted in a VR laboratory equipped with the same experimental apparatus of Experiment 1, with three major differences:

(1) The stereo projection system was driven by an MSI laptop instead of a PC, equipped with an Intel Core i7 7820HK 2.90 GHz processor with 32GB RAM and nVIDIA GeForce GTX 1070 graphics card;

(2) the LCD screens used to simultaneously display the dynamic cyclopean view of the 3D ship corridor explored by the active observer, were smaller (22" not 50", Samsung S22E450M LED set at a screen resolution of 1024 × 768);

(3) the overall viewing distance of the passive observers was settled at about 43 cm (not 100 cm) so to equate the displays in term of retinal size.

In order to provide a method that corroborates the reliability of the errors reported by the active observers, the IVR system was aided by a functionality that allows to store in real time the screenshots of the cyclopean views of the active observer. A screenshot was stored during the exploration as soon as the active participant pressed the R button of the Space Mouse. The spatial disposition of participants within the setting, and the immersive virtual environment explored during the task (the 3D L-shaped ship corridor, Fig. 1D) together with the design errors (Fig. 1A and 1B) were the same as in Experiment 1. The screenshots used during the response-encoding phase were the same as in Experiment 1 (8 valid screenshots, 4 AEs + 4 PEs, plus 1 catch trial screenshot), with 1 additional screenshot used as catch trial. This additional screenshot displayed a missing holder in a ship thruster (Fig. 1C, screenshot 2). This catch trial screenshot was apparently different from all of the other screenshots for colour properties, being reddish not greenish. Such a difference in colour purposely magnifies the un-relatedness of some trials with the 3D ship corridor thus reducing the tendency to favour positive over negative responses in our task. In Experiment 2, the same 2X2 mixed factorial design of Experiment 1 applied, with four experimental conditions resulting by the full factorial combination of 2 Types of Design Errors × 2 Viewing Conditions.

2.2.3. Procedure

The procedure was similar to the one used in Experiment 1 with the major difference that the ludic dimension was now intentionally removed from the task so to optimize the control in our experimental setting. The procedure included the same six phases described in subsection 2.1.3 and lasted almost the same time (10 minutes). Major differences regarded:

(1) the first phase, in which task was described as a game, was omitted from the current instruction;

(2) in the second phase, in which, although following the same selection criteria used in Experiment 1, albeit a relatively larger number of participants were assigned to the active (n = 17) and the passive (n = 83) conditions;

(3) the fifth experimental phase, that although requiring the participants to perform the same modified *hide and seek* task of Experiment 1 (lasting 4 minutes), it included an additional task for the active participant (to push the R button of the SpaceMouse as soon as he/she detects a design error during the exploration phase);

(4) the sixth response encoding phase that was conducted following the exact same procedure as in Experiment 1, but that included the serial presentation of 10 (not 9) screenshots.

At last, we collected demographic information (age and gender), and participants were thanked and debriefed following the same procedure used in Experiment 1.

The debriefing in Experiment 2 also served the purpose of gathering information from those active observers that during the exploration phase did take a number of screenshots that differs from those signalled by the raising of their hand. Ten observers out of the total of 17 took a screenshot of all of the errors they reported during the response-encoding phase. The remaining 7 observers either took no screenshot at all (n = 1), or took one (n = 5) to three (n = 1) screenshot less than the number of errors they reported during the response encoding phase. The average number of correspondence between screenshots and errors reported by the raising of hands was of about 82%. This showed that our response method was reliable. In the debriefing phase, all seven participants with a smaller correspondence between the screenshots they took and the errors they reported claimed that they might have: 1) forgotten about pushing the SpaceMouse button during the exploration phase; 2) pushed the SpaceMouse button accidentally or to signal relevant parts of the environment in addition to mere errors. In line with this latest claim, the analysis of the screenshots reveals that observers in general used the SpaceMouse button not only to signal a detection error but also to signal a salient part of the 3D environment. The average individual proportion of screenshots relative to the total number of actual errors (n = 8) was indeed reliably larger than the average individual error detection proportion (0.72 ± 0.09 *vs*. 0.48 ± 0.05, t = 3.91, df= 16, p = 0.001, d = 1.95). Consequently, we decided not to use the screenshots as an additional exclusion criterion for our participants, given that our result suggested that they were used for a mixture of purposes (accidental button press, to signal a detection error, or to signal a relevant part of the environment).

2.2.4. Results

As in Experiment 1, we analyzed the individual likelihood of detecting a design error as an index of performance in our modified *hide and seek* task by applying the same *glmm* model (with *probit* as a link function) to the whole set of binary responses on valid screenshots (1 = error detected; 0 = error not detected), with Type of Design Error (AE vs. PE) and Viewing Condition (Active vs. Passive) as fixed factors, and with by subject and by error random intercepts and by subject random slope for Type of Design Error. We used the same indices of effect size as in

Experiment 1 in order to support the reliability of our statistical effects. Disregarding catch trials, the data analysis was now based on the 100% of active participant's responses to valid trials (n = 136 resulting from the combination of 17 active observers and 8 errors), and on the 96% of passive participant's responses (n = 640 out of the total of 664 responses, resulting from 80 - 3 excluded from the total of 83 - passive observers and 8 errors). After the application of the same exclusion criteria used in Experiment 1, we removed from the analysis three passive observers who raised their hand when one of the two catch trials was presented. According to the power analysis, our total sample size of 97 resulted to be large enough for a moderate effect size and a power of 0.80. Figure 4A depicts the pattern of average error detection proportion as a function of the Type of Design Error (x-axis) and Viewing Condition (colour coding as by the legend). The pattern is apparently different from the one of Fig. 3A (Experiment 1), being now consistent with both Expectation 1 (main effect of the Type of Design Error), and Expectation 2 (main effect of Viewing Condition). Furthermore, differently from Experiment 1 the Type of Design Error × Viewing Condition is now absent. Hence the type of setting (natural vs. controlled) played a role in our task.

Figure 4. Error detection proportion in Experiment 2 (controlled setting). A: Mean and SE values of error detection proportion in active (green) vs. passive (blue) viewing conditions as a function of the Type of Error (*affordance* vs. *perceptual*) on the abscissa. B-C: Mean and SE values of error detection proportion in active (green) vs. passive (blue) viewing for the four types of affordance errors (in B) and the four types of perceptual errors (in C). The numbers along the abscissa indicate the relative ordering of error's appearance along the exploration path of the 3D ship corridor, from the start position to the end (same encoding of Fig. 1).

These observations were confirmed by the statistical *glmm* analyses (*rc* = 0.58, 95% CI [0.54, 0.61]). Indeed, as consistent with Expectation 2 (not 2b), *glmm* results showed a significant main effect of the Viewing Condition ($F_{1, 767}$ = 5.02, p = 0.025, η^2_p = 0.007, 95% CI [0.000, 0.019]), with a larger likelihood of detecting an error in active (M = 0.48 $\pm$ 0.04), rather than in passive viewing condition (M = 0.38 $\pm$ 0.02). This effect was quantified by an active vs. multi-passive error detection advantage of about 0.15 $\pm$ 0.13 (z= 2.15, one-tailed p= 0.016, d = 0.20). Expectation 1 was supported by the significant main effect of Type of Design Error ($F_{1, 767}$ = 8.23, p = 0.004, η^2_p = 0.011, 95% CI [0.002, 0.026]), with a larger likelihood of detecting an error for PE (M = 0.56 $\pm$ 0.03), rather than for AE (M = 0.23 $\pm$ 0.02). Unlike Experiment 1, the interaction between Type of Design Error and Viewing Condition was not statistically significant ($F_{1, 767}$ = 0.12, p = 0.727, η^2_p = 0.000, 95% CI [0.000, 0.005]).

As in Experiment 1 we further performed a post-hoc *glmm* analyses on the way in which the likelihood of errors detection was distributed amongst our two Types of Errors (Fig. 4B and 4C for affordance and perceptual, respectively). This analysis revealed that the origin of the active vs. multi-passive advantage was likely to be qualified by two out of four AEs (Fig. 4B, AE1, z = -1.52, one-tailed p = 0.064, d = 0.38; and AE2, z = -1,43, one-tailed p = 0.077, d = 0.38), and two out of four PEs (Fig. 4C, PE1, z = -1.45, one-tailed p = 0.073, d = 0.37; and PE3, z = -1.39, one-tailed p = 0.083, d = 0.39), all of which showed a trend toward significance concerning the active vs. passive error detection advantage.

2.2.5. Discussion

Experiment 2 revealed that both the active vs. multi-passive viewing advantage predicted by Expectation 2 (not 2b), and the effect of facilitation due to the type of error predicted by Expectation 1, did occur under a controlled setting. Joining data from both Experiments revealed an overall reduction of the tendency to provide a positive response in Experiment 2 relative to Experiment 1. The likelihood of detecting an error strongly decreased in the controlled setting of Experiment 2 (M = 0.40 $\pm$ 0.02), relative to the natural setting of

Experiment 1 (*M*= 0.67 ± 0.02). Such a difference was confirmed by the significant main effect of the Experiment ($F_{1, 1523}$ = 92.53, p < .001, η^2_p = 0.057, 95% CI [0.040, 0.077]), when it was included as an additional fixed factor in the *glmm* analysis. The direction of this effect was consistent with an effect of social desirability, which more likely biases responses towards positive detection in natural rather than in controlled settings.

3. Conclusions

We reported two experiments on the link between the type of vision one might experience in a collaborative virtual environment (active vs. multiple passive), the type of error one might look for during a cooperative multi-user exploration of a complex design project (affordance vs. perceptual violations), and the type of setting - manipulated through Experiments - within which a multi-user activity is performed (natural in Experiment 1 vs. controlled in Experiment 2). Our two experiments demonstrated that the likelihood of error detection within a complex 3D immersive virtual environment is characterized by an active vs. multi-passive viewing advantage (consistent with our Expectation 2). In particular, we found that such an advantage depends on multiple sources of information, like:

(1) The degree of knowledge dependence of the Type of Error the passive/active users were looking for (*low* for perceptual vs. *high* for affordance violations), as the advantage tended to manifest itself irrespectively from the setting for affordance, but not for perceptual violations. This was suggested by the main effect of the Type of Viewing in Experiment 2 vs. the Type of Viewing × Type of Error interaction in Experiment 1. This difference was characterized by an anisotropy: a facilitation in the detection of PEs over AEs (consistent with our Expectation 1) which is particularly evident under the

controlled settings (Experiment 2), given that it also emerges in the active vision conditions. We interpreted such facilitation as a by-product of the relative complexity of the encoding process supporting the detection of the error. AEs would involve a more complex encoding process than PEs to be detected, as their encoding depends on previous knowledge about the structure of the object, while PEs being based on mere violation of perceptual organization principles as good continuation and colour similarity does not requires the access to knowledge (Norman, 1988; MacKay & James, 2009). Furthermore, the overall facilitation of PEs over AEs is consistent with a strand of evidence suggesting that in tasks involving objects' recognition, artifact recognition is slowed down given that they automatically activate multiple levels of information, from manipulative to functional (Gerlach, 2009; Anelli, Nicoletti & Borghi, 2010; Costantini et al., 2011; Fantoni et al., 2016).

(2) The degree of social desirability induced by the setting in which the task was performed, as the active vs. multi-passive advantage occurred irrespectively from the Type of Error in the controlled (Experiment 2) but not in the natural (Experiment 1) setting. This anisotropy was qualified by an overall enhancement of the likelihood of reporting an error detection inducing a response ceiling in the natural rather than in the controlled setting. This was consistent with the fact that social desirability biases occur more often in natural than controlled settings (Crowne and Marlowe 1964; Fisher, 1993; Shadish, Cook & Campbell, 2002).

Taken together, these results can be reconciled with the somewhat mixed findings stemming from literature on active/passive exploration/observation for error detection and way finding

(Liu, Ward & Markall, 2007; Chrastil & Warren, 2012; Bülthoff, Mohler & Thornton, 2018). The peculiar information to which active but not passive viewers have access during exploration/observation like stereopsis (monoscopic cyclopean view in passive viewers), extra-retinal and proprioceptive information from ego-motion (absent in passive viewers), and agency/intentionality (absent in passive viewers), would have contributed to determinate a superiority of active vs. passive vision in the detection of errors in our complex 3D environment. However, this superiority might result both from the correspondence between retinal and extraretinal egomotion signals, and from the correspondence between retinal and proprioceptive signals from hand movement used to control the viewpoint motion within the immersive virtual environment. This latest information component linked to manual control is indeed known to positively affect perceptual performance in both 3D (Harman, Humphrey & Goodale, 1999; James et al., 2002) and 2D space (Ichikawa & Masakura, 2006; Scocchia, et al., 2009).

Notably, following Liu, Ward & Markall (2007), a superiority of active over passive vision is only apparently in contrast with the lower cognitive load to which our passive, rather than active, observers were subjected to during our task (with only the active observer being involved during the error detection task also in complex activities required by the active exploration of the environment, like controlling/moving the point of view, deciding where to look for etc.). It was indeed possible that in our task such an active vs. passive disadvantage was overshadowed by an active vs. passive advantage produced by the compatibility between visual and non-visual information, together with agency and intentionality, that characterize active (not passive) vision. These peculiar components might have enhanced the allocation of visual spatial attention towards relevant features of the complex 3D spatial layout we used in our Experiments. According to this interpretation, a perspective point for future studies would be to test whether this overshadowing could be minimized by reducing the complexity of the 3D environment. This reduction of complexity could lead to a reversed pattern of advantage in line with the one we found in the current study: an active vs. multi-passive disadvantage.

In general, our findings may make a useful contribution to the literature on error detection within 3D environments, technology, working sciences and methodology. Our results, indeed allowed us to provide a first tentative response to a relevant though still debated research question: *How the effectiveness of the interaction with a complex 3D environment is affected by the different types of devices (HMD vs. screen) mediating the immersive experience offered by VR technologies in collaborative contexts?.*

In the present study we empirically answered to this question that poses a novel research problem, namely the *multi-user vision problem*. This problem involves the understanding of whether a self-generated stereoscopic and immersive view of a complex layout leads to a more effective representation of the 3D scene compared to the passive reply of the same optic information simultaneously displayed on flat screens to multiple passive observers. This problem is rooted into actual working collaborative contexts, such as design review sessions in working domains (e.g., architecture, engineering and shipbuilding) in which the co-presence of more users, which share through cooperation the same project, typically occurs. These design review sessions nowadays are supported by standard collaborative IVR systems based on a mixed usage of passive and active viewing combining HMD and projection screens (Bayon, Griffiths & Wilson, 2006). In general, previous studies have not yet provided conclusive evidence on the superior advantage of a type of viewing over another during 3D interaction in collaborative immersive virtual environments. Active and passive viewing, although being based on the same visual input, have indeed access to substantially different sources of information (Wexler, Lamouret & Droulez, 2001; Wexler, 2003; Fantoni, Caudek & Domini, 2010; Caudek, Fantoni & Domini, 2011; Fantoni, Caudek & Domini 2012). Importantly, here we approached for the first time the *multi-user vision problem* using a novel adaptation of the yoked paradigm (Rogers & Rogers, 1992; Fantoni, Caudek & Domini, 2014). Our adapted yoked paradigm reproduced the simultaneous active/multi-passive conditions occurring in standard collaborative IVR systems: observers subdivided into groups, each composed by one active and multiple passive observers, were indeed asked to simultaneously search for and find (a modified *hide and seek task*) design errors within a complex 3D ship layout.

As a perspective point for the future development of cooperative software based on immersive virtual environments, we believe that our study might provide a relevant hint. A multi-user design review experience in which designers, engineers and end-users, all cooperate actively within the IVR wearing their own HMD, seems more suitable for the detection of relevant errors, than standard systems characterized by a mixed usage of active and passive viewing. This point is particularly relevant for the implementation of future technological solution based on IVR systems for the support of design review that, according to Fernández & Alonso's (2015) claims, should minimize implementation errors within projects.

4. Acknowledgements

We would like to thank all participants and Valentina Piccoli for helping with participants during NEXT 2017.

5. References

Anelli F, Nicoletti R, Borghi AM. 2010. Categorization and action: What about object consistence? *Acta Psychologica* 133:203-211. DOI: 10.1016/j.actpsy.2009.11.009

Baayen RH, Davidson DJ, Bates DM. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59:390-412. DOI: 10.1016/j.jml.2007.12.005

Bassanino M, Wu KC, Yao J, Khosrowshahi F, Fernando T, Skjærbæk J. 2010. The impact of immersive virtual reality on visualisation for a design review in construction. In *Proceedings of 2010 14th International Conference Information Visualisation (IV 2010). London: IEEE, 585-589. DOI:* 10.1109/IV.2010.85

Bayon V, Griffiths G, Wilson JR. 2006. Multiple decoupled interaction: An interaction design approach for groupware interaction in co-located virtual environments. *International Journal of Human-Computer Studies* 64:192-206. DOI: 10.1016/j.ijhcs.2005.08.005

Bowman DA, McMahan RP. 2007. Virtual reality: how much immersion is enough?. *Computer* 40:36-43. DOI: 10.1109/MC.2007.257

Braunstein ML, Tittle JS. 1988. The observer-relative velocity field as the basis for effective motion parallax. *Journal of Experimental Psychology: Human Perception and Performance* 14:582-590. DOI: 10.1037//0096-1523.14.4.582

Brooks F. 1999. What's Real about Virtual Reality? *IEEE Computer Graphics & Applications* 19:16-27. DOI: 10.1109/38.799723

Bülthoff I, Mohler BJ, Thornton IM. 2018. Face recognition of full-bodied avatars by active observers in a virtual environment. *Vision Research*. DOI: 10.1016/j.visres.2017.12.001

Caudek C, Fantoni C, Domini F. 2011. Bayesian modeling of perceived surface slant from actively-generated and passively-observed optic flow. *PloS One* 6:e18731. DOI: 10.1371/journal.pone.0018731

Chance SS, Gaunet F, Beall AC, Loomis JM. 1998. Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *Presence: Teleoperators and Virtual Environments* 7:168–178. DOI: 10.1162/105474698565659

Chandler M, Fritz AS, Hala S. 1989. Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child development* 60:1263-1277. DOI: 10.2307/1130919

Chen H, Lee AS, Swift M, Tang JC. 2015. 3D collaboration method over HoloLens™ and Skype™ end points. In *Proceedings of the 3rd International Workshop on Immersive Media Experiences*. Brisbane: ACM, 27-30. DOI: 10.1145/2814347.2814350

Chen N. 2014. Inside leap motion: 5 hands-on tips for developing in VR. Leap motion [blog]. Available at: http://blog.leapmotion.com/inside-leap-motion-5-hands-on-tips-for-developing-in-virtual-reality/ (Accessed 23 February 2018)

Chrastil ER, Warren, WH. 2012. Active and passive contributions to spatial learning. *Psychonomic Bulletin Review* 19:1-23. DOI: 10.3758/s13423-011-0182-x

Christou CG, Bülthoff HH. 1999. View dependence in scene recognition after active learning. *Memory & Cognition* 27:996-1007. DOI: 10.3758/BF03201230

Colas F, Droulez J, Wexler M, Bessiere P. 2007. A unified probabilistic model of the perception of three-dimensional structure from optic flow. *Biological Cybernetics* 97:461–477. DOI: 10.1007/s00422-007-0183-z

Costantini M, Ambrosini E, Scorolli C., Borghi AM. 2011. When objects are close to me: Affordances in the peripersonal space. *Psychonomic Bulletin & Review* 18:302–308. DOI: 10.3758/s13423-011-0054-4

Crowne D, Marlowe D. 1964. *The approval motive*. New York: Wiley.

Durlach N, Mavor A. 1995. *Virtual Reality: Scientific and Technological Challenges*, Washington, DC: National Academy Press. DOI: 10.17226/4761

Fantoni C, Caudek C, Domini F. 2010. Systematic distortions of perceived planar surface motion in active vision. *Journal of Vision* 10:12. DOI: 10.1167/10.5.12

Fantoni C, Caudek C, Domini F. 2012. Perceived surface slant is systematically biased in the actively-generated optic flow. *PloS One* 7:e33911. DOI: 10.1371/journal.pone.0033911

Fantoni C, Caudek C, Domini F. 2014. Misperception of rigidity from actively generated optic flow. *Journal of Vision* 14:10. DOI: 10.1167/14.3.10

Fantoni C, Rigutti S, Piccoli V, Sommacal E, Carnaghi A. 2016. Faster but less careful prehension in presence of high, rather than low, social status attendees. *PloS One* 11:e0158095. DOI: 10.1371/journal.pone.0158095

Fernández RP, Alonso V. 2015. Virtual Reality in a shipbuilding environment. *Advances in Engineering Software* 81:30-40. DOI: 10.1016/j.advengsoft.2014.11.001

Fetsch CR, Wang S, Gu Y, DeAngelis GC, Angelaki DE. 2007. Spatial reference frames of visual, vestibular, and multimodal heading signals in the dorsal subdivision of the medial superior temporal area. *Journal of Neuroscience*, 27: 700-712.

Fisher RJ. 1993. Social desirability bias and the validity of indirect questioning. *Journal of consumer research* 20:303-315. DOI: 10.1086/209351

Fleury C, Ferey N, Vézien JM, Bourdot P. 2015. Remote collaboration across heterogeneous large interactive spaces. In *2015 IEEE Second VR International Workshop on Collaborative Virtual Environments (3DCVE)*. Arles: IEEE, 9-10. DOI: 10.1109/3DCVE.2015.7153591

Gerlach C. 2009. Category-specificity in visual object recognition. *Cognition* 111:281-301. DOI: 10.1016/j.cognition.2009.02.005

Harman KL, Humphrey GK, Goodale MA. 1999. Active manual control of object views facilitates visual recognition. *Current Biology* 9:1315-1318. DOI: 10.1016/S0960-9822(00)80053-6

Jaekl PM, Jenkin MR, Harris LR. 2005. Perceiving a stable world during active rotational and translational head movements*. Experimental Brain Research* 163:388-399. DOI: 10.1007/s00221-004-2191-8

James KH, Humphrey GK, Vilis T, Corrie B, Baddour R, Goodale MA. 2002. “Active” and “passive” learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, & Computers* 34:383-390. DOI: 10.3758/BF03195466

Jez M, Fantoni C, Keber M, Rigutti S, Stragà M, Miceu P, Ambrosio L, Carnaghi A, Gerbino W. 2018. A Shared Immersive Virtual Environment for Improving Ship Design Review.

Technology and Science for the Ships of the Future: 770 - 777. DOI: 10.3233/978-1-61499-870-9-770.

Jones BS, Naef M, McLundie M. 2006. Interactive 3D environments for ship design review and simulation. In *Proceedings of the 5th International Conference on Computer Applications and Information Technology in the Maritime Industries*. Oegstgeest: IMarEST Benelux Branch, 409-418.

Kennedy RS, Lane, NE, Berbaum KS, & Lilienthal MG. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The InternationalJournal of Aviation Psychology* 3:203-220. DOI: 10.1207/s15327108ijap0303_3

Knoblauch K, Maloney LT. 2012. *Modeling psychophysical data in R* (Vol. 32). New York: Springer Science & Business Media.

Ichikawa M, Masakura Y. 2006. Manual control of the visual stimulus reduces the flash-lag effect. *Vision Research* 46: 2192-2203.

Liu CH, Ward J, Markall H. 2007. The role of active exploration of 3D face stimuli on recognition memory of facial information. *Journal of Experimental Psychology: Human Perception and Performance* 33:895-904. DOI: 10.1037/0096-1523.33.4.895

MacKay DG, James, LE. 2009. Visual cognition in amnesic HM: selective deficits on the What's-Wrong-Here and Hidden-Figure tasks. *Journal of Clinical and Experimental Neuropsychology* 31:769-789. DOI: 10.1080/13803390802502606

Meijer F, Van der Lubbe, RH. 2011. Active exploration improves perceptual sensitivity for virtual 3D objects in visual recognition tasks. *Vision Research* 51:2431-2439. DOI: 10.1016/j.visres.2011.09.013

Naji JJ, Freeman TC. 2004. Perceiving depth order during pursuit eye movement. *Vision Research* 44:3025-3034. DOI: 10.1016/j.visres.2004.07.007

Norman DA. 1988. The psychology of everyday things. New York: Basic Books.Ono H, Steinbach MJ. 1990. Monocular without stereopsis with and head movement. *Perception & Psychophysics* 48:179-187. DOI: 10.3758/BF03207085

Ono H, Ujike H. 2005. Motion parallax driven by head movements: Conditions for visual stability, perceived depth, and perceived concomitant motion. *Perception* 34:477-490. DOI: 10.1068/p5221

Peh CH, Panerai F, Droulez J, Cornilleau-Pérès V, Cheong LF. 2002. Absolute distance perception during in-depth head movement: Calibrating optic flow with extra-retinal information. *Vision Research* 42:1991-2003. DOI: 10.1016/S0042-6989(02)00120-7

Riecke BE, Bodenheimer B, McNamara TP, Williams B, Peng P, Feuereissen D. 2010. Do we need to walk for effective virtual reality navigation? Physical rotations alone may suffice. In Holscher C, Shipley T, Olivetti Belardinelli M, Bateman J, Newcombe N, eds. *Spatial cognition VII: Lecture notes in computer science*. Berlin: Springer, 234–247. DOI: 10.1007/978-3-642-14749-4_21

Rigutti S, Fantoni C, Gerbino W. 2015. Web party effect: a cocktail party effect in the web environment. *PeerJ* 3:e828. DOI: 10.7717/peerj.828

Rogers S, Rogers BJ. 1992. Visual and nonvisual information disambiguate surfaces specified by motion parallax. *Perception & Psychophysics* 52:446-452. DOI: 10.3758/BF03206704

Ruddle RA, Volkova E, Bülthoff HH. 2011. Walking improves your cognitive map in environments that are large-scale and large in extent. *ACM Transactions on Computer–Human Interaction*, 18 [Article No. 10]:1-20. DOI: 10.1145/1970378.1970384

Ruddle RA, Volkova E, Mohler B, Bülthoff HH. 2011. The effect of landmark and body-based sensory information on route knowledge. *Memory & Cognition* 39:686-699. DOI: 10.3758/s13421-010-0054-z

Schell J, Shochet J. 2001. Designing Interactive Theme Park Rides. *IEEE Computer Graphics & Applications* 21:11-13. DOI: 10.1109/38.933519

Scocchia L, Grosso RA, de'Sperati C, Stucchi N, Baud-Bovy G. 2009. Observer's control of the moving stimulus increases the flash-lag effect. *Vision research*, 49: 2363-2370.

Shadish W, Cook TD, Campbell DT. 2002. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

Shao F, Robotham AJ, Hon KK. 2012. Development of a 1: 1 Scale True Perception Virtual Reality System for design review in automotive industry. In *Advances in Manufacturing Technology – XXVI: Proceedings of the 10th International Conference on Manufacturing Research (ICMR2012)* 2:468-473.

Sherrington C. 1906. *The Integrative Action of the Nervous System*. New York: Charles Scribners Sons.

Tanaya M, Yang K, Christensen T, Li S, O'Keefe M, Fridley J, Sung K. 2017. A Framework for analyzing AR/VR Collaborations: An initial result. In *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)* Annecy: IEEE, 111-116. DOI: 10.1109/CIVEMSA.2017.7995311

Teramoto W, Riecke BE. 2010. Dynamic visual information facilitates object recognition from novel viewpoints. *Journal of Vision* 10:11. DOI: 10.1167/10.13.11

Ujike H, Ono H. 2001. Depth thresholds of motion parallax as a function of head movement velocity. *Vision Research* 41:2835-2843. DOI: 10.1016/S0042-6989(01)00164-X

Vonesh EF, Chinchilli VM, Pu K. 1996. Goodness-of-fit in generalized nonlinear mixed-effects models*. Biometrics* 52(2):572-587. DOI 10.2307/2532896

von Helmholtz H. 2002. *Handbook of physiological optics.* New York: Dover.

Wallach H. 1987. Perceiving a stable environment when one moves. *Annual review of psychology* 38:1-29. DOI: 10.1146/annurev.ps.38.020187.000245

Wallach H, Stanton L, Becker D. 1974. The compensation for movement-produced changes in object orientation. *Perception & Psychophysics* 15:339-343. DOI: 10.3758/BF03213955

Waller D, Greenauer N. 2007. The role of body-based sensory information in the acquisition of enduring spatial representations. *Psychological Research* 71:322-332. DOI: 10.1007/s00426-006-0087-x

Waller D, Loomis JM, Haun DB. 2004. Body-based senses enhance knowledge of directions in large-scale environments*. Psychonomic Bulletin & Review* 11:157-163. DOI: 10.3758/BF03206476

Wang RF, Simons DJ. 1999. Active and passive scene recognition across views. *Cognition* 70:191-210. DOI: 10.1016/S0010-0277(99)00012-8

Wexler M. 2003. Voluntary head movement and allocentric perception of space. *Psychological Science* 14:340–346. DOI: 10.1111/1467-9280.14491

Wexler M, Lamouret I, Droulez J. 2001. The stationarity hypothesis: An allocentric criterion in visual perception. *Vision Research* 41:3023–3037. DOI: 10.1016/S0042-6989(01)00190-0

Wexler M, van Boxtel JJ. 2005. Depth perception by the active observer. *Trends in Cognitive Sciences* 9:431-438. DOI: 10.1016/j.tics.2005.06.018

Wilson PN, Foreman N, Gillett R, Stanton D. 1997. Active versus passive processing of spatial information in a computer-simulated environment. *Ecological Psychology* 9:207-222. DOI: 10.1207/s15326969eco0903_3

Wilson PN, Péruch P. 2002. The influence of interactivity and attention on spatial learning in a desk-top virtual environment. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition* 21:601-633.

Figure 1

Screenshots of design errors and the immersive virtual environment used throughout Experiment 1 and Experiment 2.

A and B: screenshots of the 8 actual design errors implemented in the digital mock-up of the corridor as presented to our participants during the response encoding phase, subdivided into the two types: affordance violations in A, and perceptual violations in B. Numbers indicate the relative ordering of appearance of violations along the immersive virtual environment explored during the task from the starting position to the corridor's end (in D). C: the two screenshots used as catch trials shown to the participants in an intermixed and randomized order together with the screenshots of the actual design errors used during the response-encoding phase of the experiments. Experiment 1 included the presentation of the only catch trial screenshot 1, Experiment 2 included the presentation of the both catch trials' screenshots (1 & 2). D: a bird eye view of the immersive virtual environment, from the starting position (coded by the blue star) to the corridor's end (orange circles stand for design errors). The immersive virtual environment was the rendering of a digital mock-up of an L-shaped 3D ship corridor along which the 8 design errors were sequentially implemented along the pathway the observer was required to travel (the numbering corresponds to their relative ordering of appearance along the pathway).

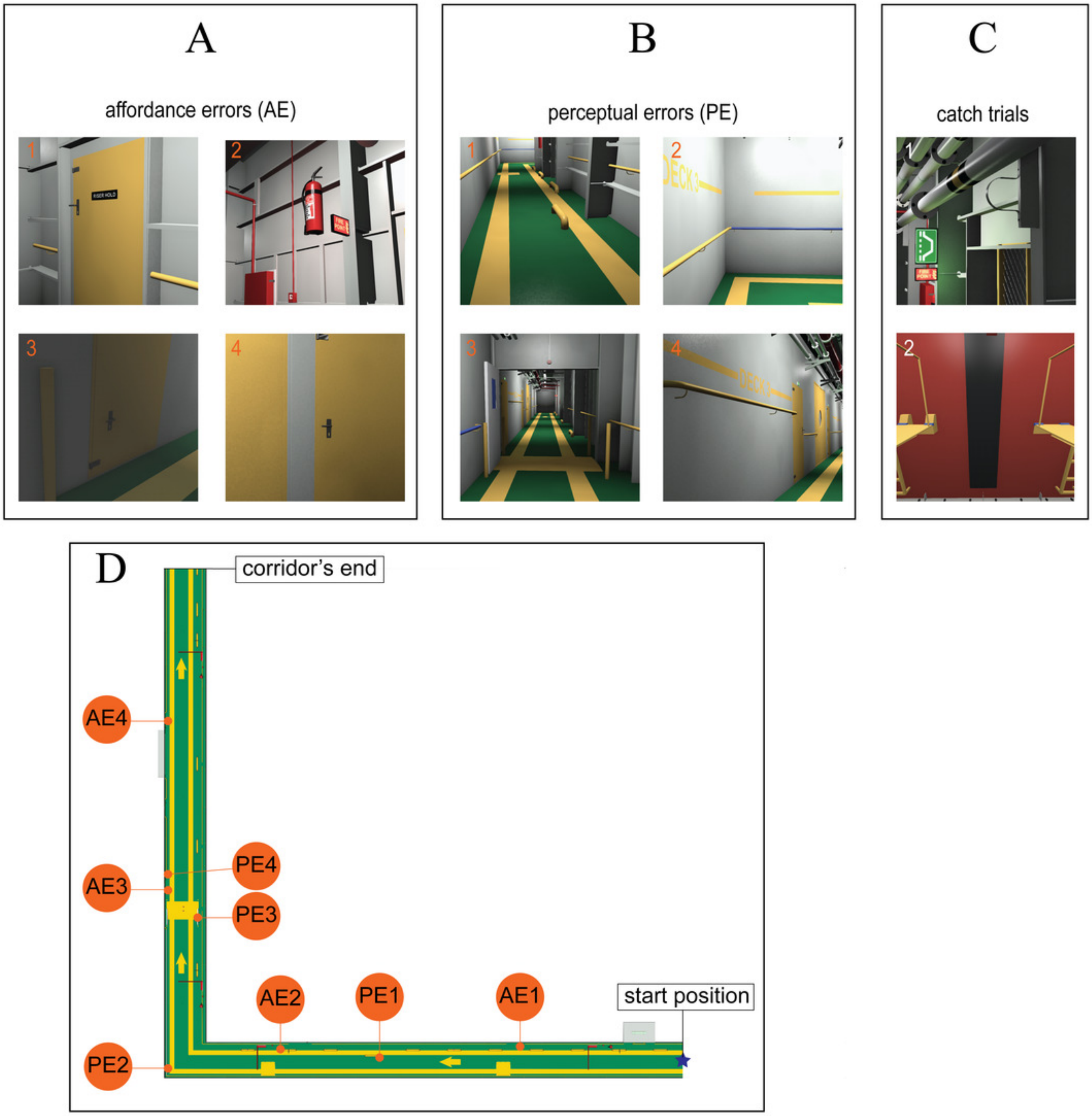

A
affordance errors (AE)
1
2
3
4
B
perceptual errors (PE)
1
2
3
4
C
catch trials
1
2
D
corridor's end
AE4
AE3
PE4
PE3
AE2
PE1
AE1
start position
PE2

Figure 2

Experimental setting of Experiment 1.

A: schematic view of the experimental setting with three observers (one active, green arrow, and two passives, blue arrows) implemented in the natural condition during the science and technology exhibit of Experiment 1, with superposed the minimal distance the passive observers stand (as constrained by the 90 cm desk size), and their average viewing distance (100 cm when sitting with their elbows on the margin of the desk). This distance is taken from the 2 large LCD screens (127 cm diagonal), displaying in real time the exact same 3D, though monoscopic, view the active observer self-generated combining head movements and SpaceMouse control. In this example, the view produced by the active observer that the 3 passive observers are looking at in real time is consistent with PE1. B: a photograph of the setting during the training session with the digital mock-up (a ship thruster) used to familiarize the active observer with the SpaceMouse and the passive observers with the 3D graphic. Notice that the setting of Experiment 2 reproduced in smaller scale the one shown in the current scheme, including smaller screens though smaller viewing distances in order to equate the two Experiments for the size of the passive displays in term of retinal sizes (57.9° × 34.5°). B photographed by Carlo Fantoni.

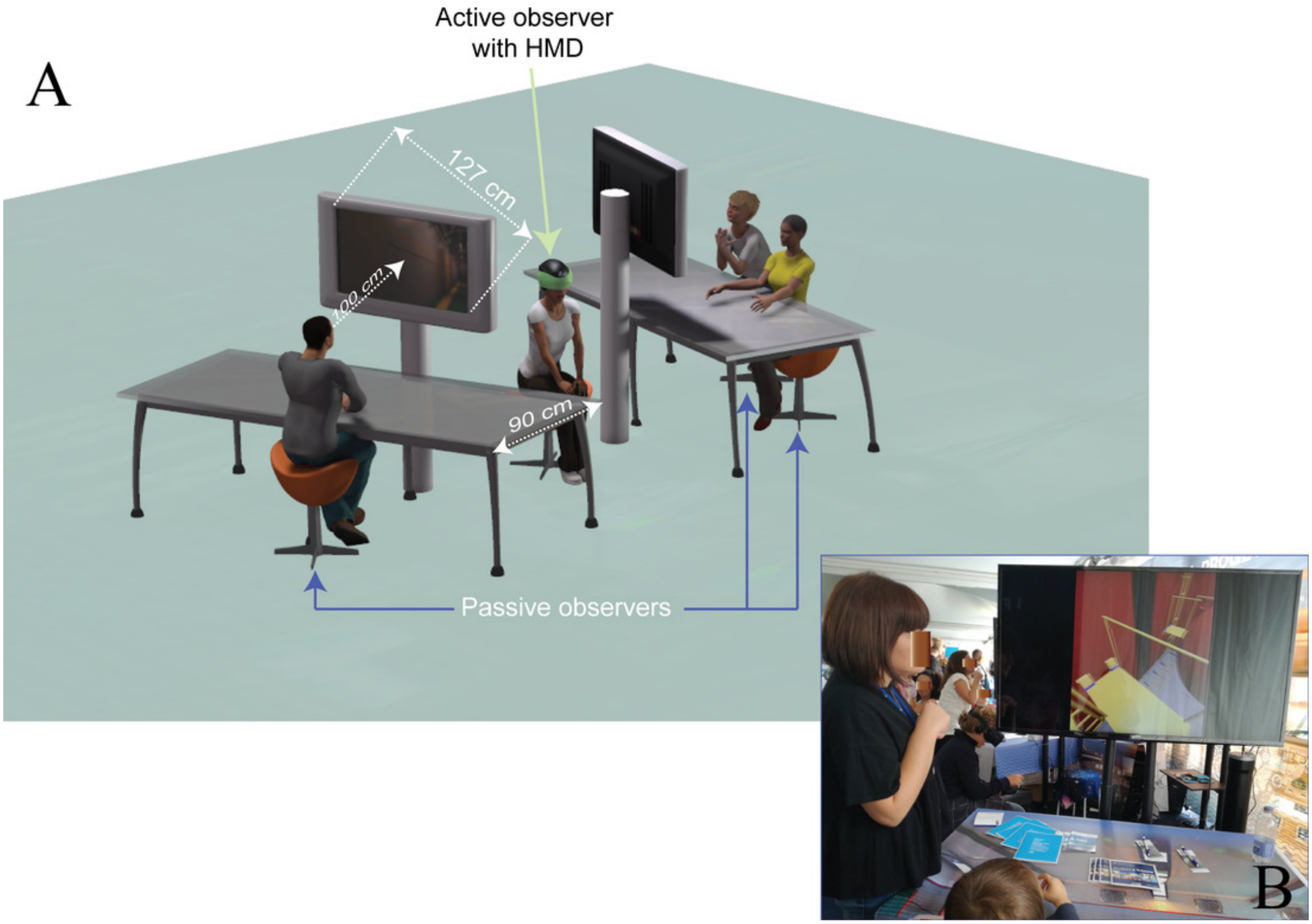

A
Active observer
with HMD
127 cm
100 cm
90 cm
Passive observers
B

Figure 3(on next page)

Error detection proportion in Experiment 1 (natural setting).

A: Mean and SE values of error detection proportion in active (green) vs. passive (blue) viewing conditions as a function of the Type of Error (*affordance* vs. *perceptual*) on the abscissa. B-C: Mean and SE values of error detection proportion in active (green) vs. passive (blue) viewing for the four types of affordance errors (in B) and the four types of perceptual errors (in C). The numbers along the abscissa indicate the relative ordering of error's appearance along the exploration path of the 3D ship corridor, from the starting position to the end (same encoding of Fig. 1).

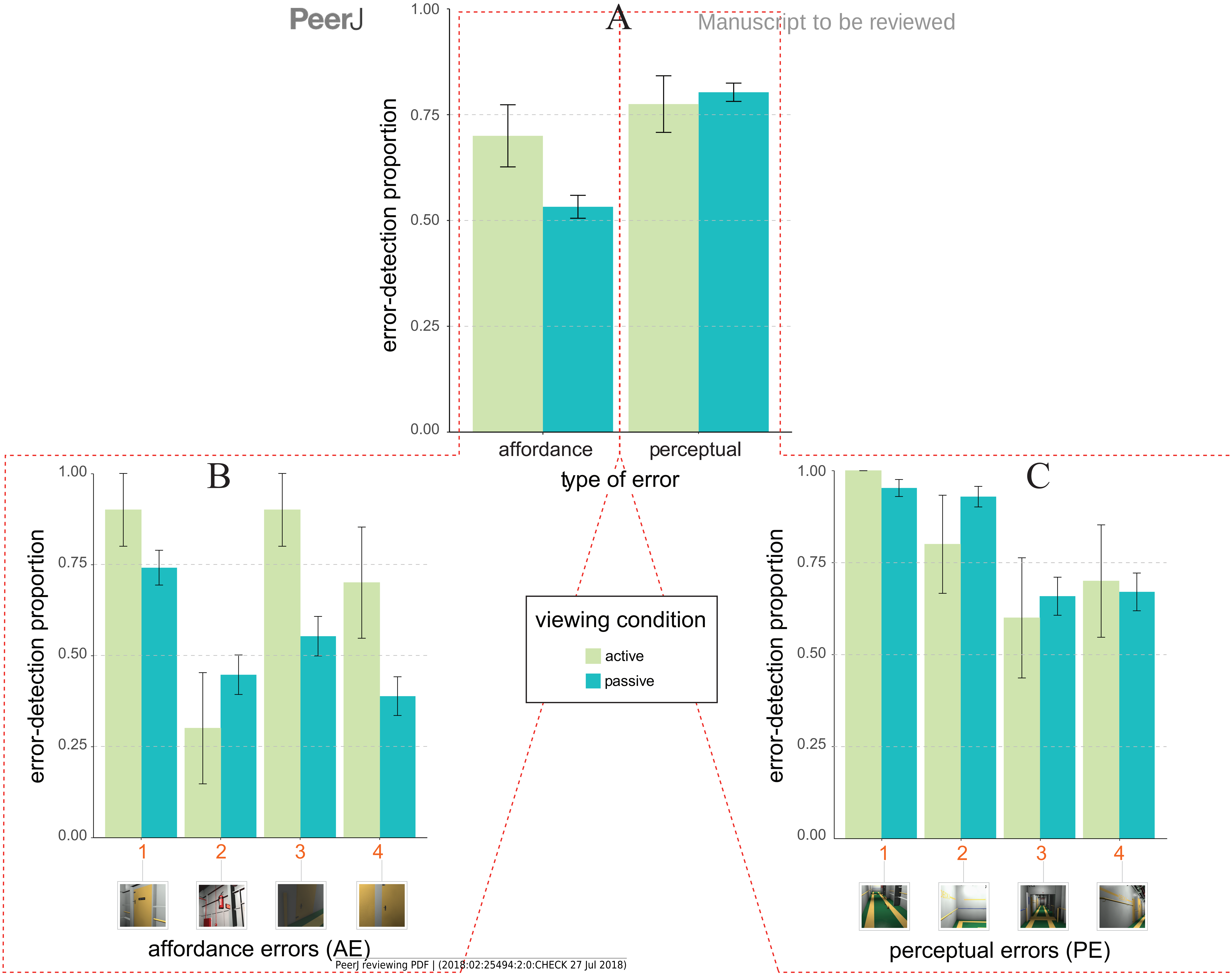
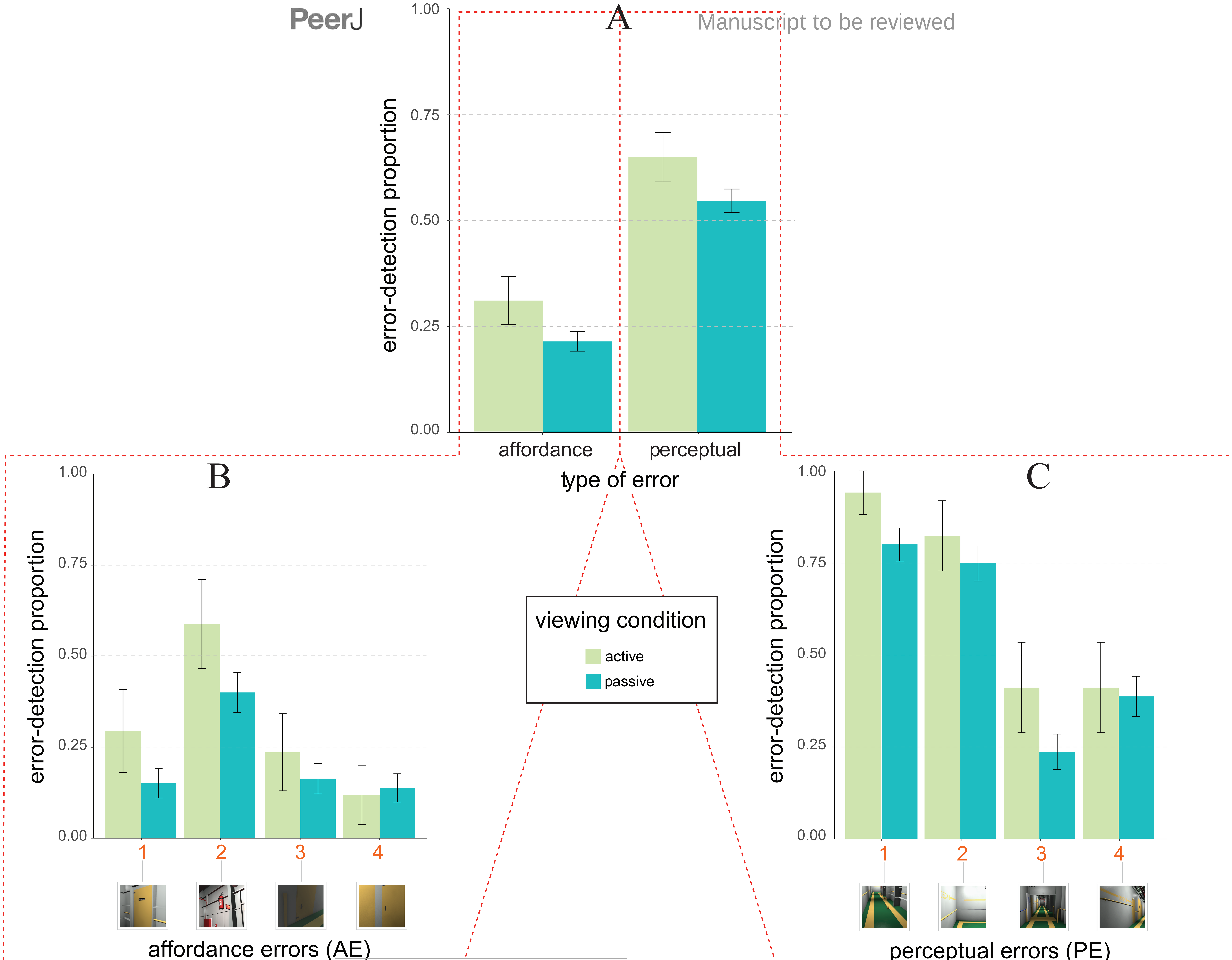

A
B
C
error-detection proportion
1.00
0.75
0.50
0.25
0.00
affordance
perceptual
type of error
viewing condition
active
passive
1
2
3
4
affordance errors (AE)
perceptual errors (PE)

Figure 4(on next page)

Error detection proportion in Experiment 2 (controlled setting).

A: Mean and SE values of error detection proportion in active (green) vs. passive (blue) viewing conditions as a function of the Type of Error (*affordance* vs. *perceptual*) on the abscissa. B-C: Mean and SE values of error detection proportion in active (green) vs. passive (blue) viewing for the four types of affordance errors (in B) and the four types of perceptual errors (in C). The numbers along the abscissa indicate the relative ordering of error's appearance along the exploration path of the 3D ship corridor, from the start position to the end (same encoding of Fig. 1).


A
B
C
1.00
0.75
0.50
0.25
0.00
error-detection proportion
affordance
perceptual
type of error
viewing condition
active
passive
1
2
3
4
affordance errors (AE)
perceptual errors (PE)