IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 20 July 2022, accepted 11 August 2022, date of publication 16 August 2022, date of current version 9 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3198957

RESEARCH ARTICLE

Virtual Reality e-Commerce: Contextualization and Gender Impact on User Memory and User Perception of Functionalities and Size of Products

GUILHERME GONÇALVES[1,2], GALVÃO MEIRINHOS[3], VÍTOR FILIPE[1,2], MIGUEL MELO[1], AND MAXIMINO BESSA[1,2]

[1]Institute for Systems and Computer Engineering, Technology and Science, 4200-465 Porto, Portugal
[2]Department of Engineering, University of Trás-os-Montes e Alto Douro, 5001-801 Vila Real, Portugal
[3]Department of Letters, Arts and Communication, University of Trás-os-Montes e Alto Douro, 5001-801 Vila Real, Portugal

Corresponding author: Guilherme Gonçalves (guilhermeg@utad.pt)

This work was supported in part by the National and European Funds through the Portuguese Funding Agency FCT—Fundação para a Ciência e a Tecnologia under Project SFRH/BD/148714/2019 and in part by the National Funds through the Portuguese Funding Agency FCT—Fundação para a Ciência e a Tecnologia under Project UIDB/50014/2020.

ABSTRACT Virtual reality (VR) potential to isolate users from the real world while producing a rich virtual environment where users act similarly to how they would, in reality, is still being investigated in several fields. In this work, we investigated the effects of product contextualisation and gender under an immersive VR application where users can explore in-depth a commercial product with a "hands-on" experience. An experimental between-subjects study was performed with 38 participants between 18 and 28 years. The product tested consisted of a double-door refrigerator equipped with a touchscreen. Two independent variables were studied: Context (the refrigerator was filled with food products and placed in a kitchen), Neutral Context (empty refrigerator displayed in an empty white room), and Gender (Female and Male). As for the dependent variables, we considered how clarified users felt about the product functionalities, its size, the extent users remember details and characteristics of the refrigerator, and the user's subjective workload. The evidence shows that contextualisation and gender have no impact on any dependent variables. Therefore, we concluded that presenting a product in its context does not benefit significantly benefit it. Thus, opting for a neutral context would be preferable to save computational costs and human resources necessary to build and run the higher complexity environments required to contextualise the product.

INDEX TERMS Virtual reality, immersive, context, gender, user perception, product evaluation, memory.

I. INTRODUCTION

Online shopping revolutionised the way we purchase products. From the comfort of our homes, we can search endless marketplaces, verify reviews from previous buyers and get products delivered to one's door. However, online shopping also has its disadvantages: we only get to experience the product first-hand after purchase, namely when it is delivered. For example, a study performed by Liu *et al.* [1] concluded that physical store shoppers of luxury goods that are more averse to online risks consider valuable the possibility of checking the product personally before purchasing it. These consumers also consider the shopping experience and interactions necessary. Another study conducted by Kim *et al.* [2] revealed that online shoppers that feel a higher degree of product-level uncertainty are unlikely to buy expensive

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaogang Jin.

products online despite having experience with online shopping. The same authors also concluded that users are reluctant to purchase expensive products using only digital information, preferring only to purchase cheaper products online. Again, this uncertainty comes from the impossibility of experiencing the product first-hand.

Lately, virtual reality (VR) potential is being widely researched [3], more specifically, how can VR contribute to overcoming current limitations and constraints of the real world. Thus, VR could help mitigate some of the inherent constraints of buying online by allowing clients to test products before purchase. In this case, we specifically tested whether Product Contextualization (Subsection A) could help users to understand better the product characteristics (similarly to what is observed in the real world) or not.

Thus, in this experimental study, we studied the effects of Product Contextualization on the user's perception of a product (how well users remember details of the product and how well they feel clarified about its functionalities and size) under an immersive VR (IVR) setup. We also analyze the role of gender in these variables. This study's results contribute to the body of knowledge of VR and e-commerce by better understanding how Product Contextualization applies to users under an IVR setup and if gender differences are present or not. Such data can bolster new investigations in this field, help to improve strategies for online shoppers (and more specifically for each gender, if such is the case), enhance their experience and trust, and benefit practitioners in the field.

A. PRODUCT CONTEXTUALISATION

Product Contextualisation is a well-known concept in commerce that can help consumers better visualise how they would use the products. It consists of placing the product in question (focal product), together with other products (complementary products), creating an environment similar to the one the consumer would use.

There are three factors in product contextualisation [4], [5]: Functional - complementary products that can be consumed jointly with the focal product (e.g., users buying a new phone are also likely to buy other accessories such as phone covers, screen protectors or power banks). Aesthetic - complementary products that create a satisfying aesthetic relationship when used together with the focal product (e.g. using watch straps that go well with what the user is wearing). Sociocultural - complementary products with no connection to the focal product but related through sociocultural association (e.g. associating roses with romantic dinners).

Product Contextualization is easier to implement in 3D environments than in the real-world [6]. Under an IVR setup, context could be easily and readily changed [3]. Such could also allow faster iterations through user studies to understand which context a new product has more chances to elicit purchase intentions. Furthermore, users often prefer to inspect products within the context [6]. It can also influence the consumer's attitude towards the products [3], [5].

B. VIRTUAL REALITY

Virtual Reality (VR) is a technology capable of simulating virtual worlds in real-time, where users and 3D entities interact using sensorimotor channels [7]. In other words, it can completely isolate the users from the stimuli of the real world while replacing them with synthesised ones [8]. Its potential in several fields such as entertainment, video games, education [9], [10], medical [11], [12], marketing [13], rescuers [14], among others is being constantly studied.

Nowadays, IVR can be promptly experienced through several devices such as smartphones (e.g. using Google Cardboard), laptops and desktop computers (e.g. using VR headsets such as HTC Vive or Oculus Rift), or directly using standalone VR headsets such as the Oculus Quest without the need for other devices.

IVR also adds layers of safety because experiences simulating risky situations and environments can be performed in the safety of controlled locations where users could train or experience scenarios without endangering themselves or others [15], [16]. Furthermore, it could further reduce time constraints and logistics costs because users can experience scenarios without needing to travel to specific locations. It could also mitigate possible problems with the availability of products/materials. For example, users who want to check a specific product in-store but cannot, because the same is unavailable, could instead experience the same product (to a certain extent) using IVR.

VR can also allow collaboration in real-time or asynchronous, which increases the efficiency of the available time and funds for all parties, interacting through the use of different platforms while being in different places [17], [18]. For example, in the context of e-commerce, consumers could check online products in IVR with the help of store staff that could assist the users with the products in real-time.

In a study performed by Suh *et al.* [19], users experienced virtually high experiential (non-IVR featuring an interactable 3D model) and virtually low experiential (static images) products. They concluded that VR increases the overall consumer learning about products. Therefore, the use of VR, and particularly IVR, could further mitigate costs for online stores and sometimes consumers [20] by reducing the number of return items (because users were not fully informed before the purchase) [21], [22] by increasing the consumer trust.

Delarue *et al.* [3] discussed VR in contextualized product testing. They affirm that VR requires specific knowledge of artistic and computer skills and that developers that meet these requirements often prefer to work in entertainment rather than consumer research. Authors also state that realism in VR is still not natural enough. Developing IVR requires specific skills, but we argue that VR development platforms (such as Unity or Unreal Engine) are becoming more and more user-friendly and straightforward. Game engines are also constantly being optimised and improved to provide even more rich visuals, which, combined with the evolution of technology and helped by the substantial widespread of VR and its community, means that achieving more natural visuals in IVR is now easier than before and should continue to be so.

One example of an application that takes advantage of product contextualisation is IKEA. Users can place the store products in their home via augmented reality (AR) and visualise beforehand how they would look together with the rest of their furniture [23]. Another example is applications that display how clothes would fit users using AR [24], [25]. Although AR presents some advantages in this type of "hands-on" when compared to other media, several authors argue that VR is the most likely to have the most significant impact [26], [27], [28]. This is due to its inherent ability to isolate the user from real-world stimuli completely. With the users "shut off" from reality, the replication of multisensory stimuli in-product experiences can happen without the interference of the conditions of the real world where the experiment is taking place.

Two major concepts in the field of IVR that help understand the potential its potential are the sense of presence and immersion. Presence is "a state of consciousness" [29], usually described in the literature as the "sense of being there". The higher the sense of presence, the more likely users will behave in the virtual world the same way as they would in reality. Csikszentmihalyi's theory of flow [30] shows similarities to Presence. Flow is a mental state where users are so focused on the task that they lose the sense of time and awareness of their surroundings. The sense of Presence has been used throughout the literature as a metric to evaluate the effectiveness of a VR experience [31], [32]. The level of immersion (often confused with presence [33]) describes how well the VR system can isolate the users from the real world and how well it delivers the synthesised stimuli of the virtual environment [29]. This is considered an objective metric contrary to presence, which is of subjective nature. The higher the level of immersion, the more likely users are to feel a greater sense of presence [34], which improves the likelihood of users having similar personal and physiological responses to the ones they would have in the real world.

C. LIMITATIONS OF VIRTUAL REALITY IN E-COMMERCE

Although IVR's tremendous potential and technology and knowledge continuously improve, there are still technological limitations when using IVR. Regarding navigation through the environment for better exploration, several setups allow us to walk through the virtual environment physically; however, many setups are limited by a specific tracking area. The absence of a tracking area can also be concerning, as walking without limitations can be a hazard for some people as they can wander off a safety area, unaware of the real world. There are solutions and workaround for these issues, such as using a teleportation method, allowing users to be stationary and using the controller to indicate where they want to go, or using omnidirectional treadmills, which would imply higher costs and logistics. With a simple IVR setup, consisting of one VR headset and a pair of controllers, there is the impossibility to feel other stimuli than audiovisual (such as haptic, smells, taste, and others). This means that users can go through objects but cannot feel their weight, their texture, and, depending on the product, their smell, taste, and temperature. Some stimuli can be replicated in a multisensory setup, but even then, there would be technological limitations, cost constraints and business model incompatibilities. For example, testing a car in IVR could prove to be possible by using simulators comprised of force-feedback actuators. However, even though they could not simulate every force, a car and driver endure during real-world driving, which could mislead buyers. Another example is regarding smells. Although there are devices for VR headsets that can release smells [35], they need to be purchased and installed beforehand; therefore, the possibility of experimenting with a new perfume online in an IVR is infeasible because to test it, the consumer would need to recreate it at home, defeating the purpose of the application. The same reasoning could be applied to taste and temperature. Furthermore, suppose consumers are given the means to recreate products to preview them in IVR. In that case, businesses could risk consumers exploiting this possibility to their advantage, counteracting the intention to purchase.

Some products are also inherently easier and only possible to try in the real world than in IVR [3]. For example, the experience of trying a new pair of running shoes would depend not only on the feeling itself, the terrain they are tested but also on how long they are used. Another example would be swimming equipment that, although one could visually inspect it in IVR and check some parts of its functionality, could not actually be tested in IVR, as it would require the user to be submersed in actual water.

VR headsets can also distort the user's distance perception leading users to underestimate distances [36]. In commerce, this factor could mislead consumers into thinking that a product is smaller than it actually is, further diminishing trust in IVR to gather product information. Although, some methods have been studied and proved to improve the accuracy of users' distance perception. A cluttered environment can help users estimate distances in VR more correctly as they have other objects to give a sense of scale [37]. Indoor environments can further assist the user's distance perception by presenting the clutter closer to the user than an outdoor setup [37].

Another possible issue is the cybersickness symptoms (nausea, vertigo, disorientation, among others) that some users can experience during and after the experience. These symptoms usually appear when the visual and vestibular systems enter in conflict, in other words, when we see we are moving but feel we are stationary and vice-versa (like driving a car in IVR without any force feedback or trying a virtual rolling coaster) [38], [38]. Therefore, a simple consumer-level VR headset (to address as many possible consumers as possible) would limit the number of products that could be fully replicated and tested as if they were real without incurring cybersickness.

D. THE IMPACT OF VISUAL COMPLEXITY

Wedel *et al.* [26] brought to attention that previous studies seem to indicate that abstract representations may be more beneficial because of the reduced complexity they provide versus environments with higher levels of realism (by replicating the as best as possible the real world). This would indeed make specific tasks easier to perform as there are fewer distractions in the virtual world and improve the computational performance of the system [39]. For example, Mania *et al.* [40] studied how rendering quality (flat-shaded vs radiosity) affected location-based recognition memory and states of awareness. They concluded that a higher proportion of recognition was achieved in the less realistic condition. This is important because memory recall is essential in commerce, being the basis of advertising [41], and product placement (the placement of product and/or brand names or their logos in a scripted medium [42]) as it can impact purchasing decisions [43]. It is thus essential to understand how visual complexity can interfere with the consumers' ability to gather information and recall it afterward.

Immersive platforms are prone to possible readability issues [44] when they display an image fidelity below what the human perception can process, which is a common situation in today's VR headsets. In online shopping, IVR applications with small detailed products, this factor can influence (to an extent) how users behave and process the visual information. For example, if there are important fine details in the product, users might miss them, and active guidance could be needed to make sure they get closer to the product to perceive it. However, readability issues should not be a significant problem in products of bigger dimensions and without important small visual details.

Another factor influencing how users perceive the visual information presented to them is how the human brain processes information. For example, Itti and Koch [45] discussed that only a small part of the visual information is processed in such a way that it can influence human behaviour. There are sections in the virtual environments that are visually salient (dependent on the context of the experience) which are the ones targeted first. Driving attention toward other aspects of the scene requires a voluntary effort. Another aspect of human perception is that it may be incapable of perceiving differences in the quality of objects that are not related to the task at hand [46]. This means that if a task focuses on one aspect of the virtual object, other details might get through unnoticed. This is supported in the widely known study from Simons *et al.* [47], where 'change blindness' (blindness to large changes in the scene when we move our view from one place to another) and 'inattentional blindness' (blindness to objects outside our attention focus) can lead us to miss much visual information even if it presents right in front of us.

This inability to fully recognise and comprehend the information because users are focused on their tasks or the technological limitations of VR headsets can have implications for IVR online shopping applications. We cannot control technological limitations more than by using the best available equipment. However, we can circumvent them by making sure users can get close enough to virtual objects to fully perceive the details, even if this adds an extra layer of complexity that does not exist in the real world to the same extent.

E. USER's WORKLOAD IN IVR

Mental workload is influenced by the amount of working memory resources a task requires [48]. These memory resources are limited [49], only available for brief moments (from seconds to minutes, enough to execute the task) [50], and shared by several mental processes from sensory-level to meaning-level processing [51]. Mental overload occurs when a user is subjected to such an amount of stimuli that is no longer possible to maintain the performance [52]. Mental overload is dependent on factors such as the task complexity, the user's characteristics, feedback, and the strategies used to perform the task [53]. There is evidence in the literature that performing in VR requires a higher mental workload than performing the same task in the real-world [54], [55]. Aspects such as interface, interactions, text input postures, and movements can impact the mental workload [56], [57], [58], [59], [60], [61], [62]. The time available to execute tasks is an important driver of mental workload [63]. Time constraints can thus stress and affect the user's performance [64].

However, some types of tasks seem to require the same or less mental workload in VR and in the real world [65], [66], [67]. Furthermore, the mental workload seems to also be related to the user's attention [68]. Souchet *et al.* [53], in their review of IVR ergonomics and risks at the workplace, affirm that literature on mental workload in VR is still scarce.

F. GENDER DIFFERENCES

Gender is still insufficiently understood in an e-commerce context [69]. Studies indicate differences between genders in key aspects of online shopping. For example, Wolin *et al.* [70] found evidence that males have more positive attitudes toward online shopping than females. Garbarino *et al.* [71] concluded that women perceive a higher level of risk in online purchasing than men and that having a site recommended by friends reduces the perceived risk and increases willingness to purchase online to greater extents in females than males. Pascual-Miguel *et al.* [72] used the extended unified theory of acceptance and use of technology (UTAUT2) [73] to accommodate the variables "perceived risk" and "perceived trust". Evidence shows that gender influenced the relationships between effort expectancy (how the technology is perceived to be easy to use) and purchase intention and between social influence (how the user values the technology in the social network context) and purchase intention. Product type also influenced significantly the relationship between perceived risk and purchase intention only for females.

The gender differences can be partially explained through the Selectivity Hypothesis [74]. Briefly, it describes how genders process information in different ways. For example, males tend to have higher elaboration thresholds, indicating that they are more likely to purchase products without the

TABLE 1. List of the hypothesis being tested.

Hypothesis	Description
H1	Contextualisation will impact the extent users are clarified about the product functionalities.
H2	Contextualisation will impact the extent users are clarified about the size of the product.
H3	Contextualisation will impact the extent users remember details from their experience with the product.
H4	Contextualisation will impact the user's workload.
H5	Gender will impact how users are clarified about the product functionalities.
H6	Gender will impact the extent users are clarified about the size of the product.
H7	Gender will impact the extent users remember details from their experience with the product.
H8	Gender will impact the user's workload.

need to analyse all the available information. In contrast, females will inspect more deeply the available information before purchasing [74].

G. HYPOTHESIS

State of the art shows evidence of the importance of product contextualisation in commerce as well as the potential of IVR to address current gaps in increasing the users' trust (by experiencing the product "first hand") and reduce costs for both the users and sellers by eliminating the needs to travel to stores, having display product for consumers to experiment with (and possibly get damaged in the process) and reduce returned items due to lack of information or contextualisation on how consumers would use the product. In this work, we put forward an experimental study with users to address the potential of product contextualisation in an IVR setup.

Contextualization can help consumers to paint a clearer picture of how they would use the products. Therefore, in *H1*, we hypothesize whether experimenting with the product in the context it would be used could help users better clarify its functionalities.

Evidence points towards an underestimation of distances when using VR headsets. However, this can be counteracted by adding multiple familiar objects to participants, increasing the sense of space and improving distance perception [37]. Contextualization is then expected to increase this notion of space. This can happen externally (where the product is placed) and internally (what is placed in the product). For example, in products with an interior (such as refrigerators, microwaves, washing machines, dressers, or cabinets), the contextualization of that product both externally and internally (such as having clothes in a dresser or milk in a refrigerator). Therefore, we considered the hypothesis that (*H2*) context will influence the extent users are clarified about the size of the product.

Experimenting with a contextualised product should improve how much information about the product the users remember. However, previous works indicate that visual complexity can hinder the ability to focus on the task [39]. When adding context, we introduce visual complexity and thus hinder what users remember from their experience with the product [40], and by extension, affect the effectiveness of the application to influence purchase decisions [43]. This factor could counteract the positive effects of contextualisation in IVR. Therefore, the third hypothesis being tested is that (*H3*) the contextualised environment will impact the details (related to the product's characteristic) users remember from their experience. The extra visual complexity brought by contextualisation could increase the users' workload as some parts of the environment could become more visually salient than the product they are testing, thus requiring a voluntary effort to keep the focus on the product [45]. Furthermore, literature shows that users performing tasks in IVR suffer from a higher mental workload than when performing in the real world [54], [55]. Therefore, their ability to acknowledge the information provided by the IVR experience when performing the tasks could be hindered in the highest visual complexity condition (Context condition). However, because users are also guided through the experience with several tasks, the focus required to perform them can lead them to become unaware of their surroundings [30], and blind to objects outside their attention focus [47], which could counteract the effect of visual complexity. Thus, we included in this work the study of mental workload components and other workload perception variables (*H4*).

Several pieces of evidence in the state of the art show that genders differ in several aspects that can affect this study's variables. For example, the Selectivity Hypothesis shows how genders differ in the type and amount of information they process before making a purchase. Such could influence how clarified they could be about the product functionalities, their size, and how well they remember/acknowledge the product's details before purchasing it. Such could ultimately also lead to differences in user workloads. There is still no consensus on this matter, and gender is still being studied [69]. Therefore, the same hypothesis addressing contextualisation of the product was also tested for differences between genders (*H5*, *H6*, *H7*, *H8*).

Table 1 displays all the hypotheses being tested.

II. METHODOLOGY

The experimental study used a between-group design. Sample, variables, instruments, materials, procedures, and statistical procedures are described in detail below.

A. SAMPLE

We used a convenience non-probabilistic sampling method to gather 38 volunteers (21 male) with ages between 18 and 28 years (M=31.370, Std.Dev=2.541). These were recruited at the university.

The majority of the sample consisted of university students (86.8%), while the rest were workers (13.2%). More than half of the participants reported having previous experience with VR technologies (57.9%). When asked if they have ever brought a refrigerator (the product being tested in this study), the majority said no (73.7%). None of the participants withdrew from the study; thus, everyone finished the study successfully.

B. VARIABLES

Table 2 contains a description of all variables considered in this study and their respective sub-scales/levels. The dependent variables considered in this study were: Functionality Clarification, Size Clarification, Memory, and Perceived Workload. Perceived Workload has 6 subscales: Mental (the resulting workload from thinking, deciding, and processing what is needed to perform the task), Physical (how intense the physical activity performed to complete the task), Temporal (related to the time pressure of completing the task), Effort (the level of effort required from the user to maintain the level of performance), Performance (how successful the user was on completing the task), Frustration (the insecurity, dissatisfaction, or discouragement the user felt during the task).

The independent variables considered were product contextualisation (Context and Neutral Context) and Gender (Male and Female).

C. INSTRUMENTS

A sociodemographic questionnaire was used to gather information about previous experience with VR by asking when was the last time subjects experienced VR and their level of satisfaction with those VR experiences on a 5-point scale. It also addressed whether subjects had previously purchased a refrigerator.

Another questionnaire (after the experiment) addressed whether users felt clarified about the product functionalities and the product size, all using a 5-point Likert scale.

Memory questions were based on Mania *et al.* [40] study, which includes questions about specific details about the refrigerator: "What was the brand of the refrigerator?", "How many years of warranty had?", "How many drawers had?", "How many compartments did the freezer have?" and "What was the lowest possible temperature for the freezer?". Each has four possible choices, whereas 1 of them is the correct one. For each question, users reported their confidence level in their answer using a 5-point Likert scale from "No confidence" to "Certain".

The perceived workload was addressed using the NASA-TLX [75] questionnaire that assesses workload on six questions, using 7-point scales, with the following scales: Mental Demand, Physical Demand, Temporal Demand, Overall Performance, Effort, Frustration Level. This questionnaire has been used in multiple studies and was proven to be a reliable and robust method to evaluate the users' perceived mental workload in tasks [76].

D. MATERIALS

The computer running the VR experiment was equipped with an Intel Core i7-8700K processor, 32 GB RAM, and an NVIDIA GeForce RTX 2080ti under a Windows 10 operating system. The head-mounted display was an HTC Vive Pro equipped with a wireless module and two HTC Vive controllers for interaction with 1440 × 1600@90hz per-eye resolution.

Unity 2022 high-definition render pipeline was used to build the application. The product considered for this study was a double-door refrigerator with a freezer (Fig. 1 and 2) with built-in ice and water dispenser. On one of the doors, there was an interactive touchscreen (Fig. 3) where users could check and modify the freezer and refrigerator temperatures, current date and time, alarm settings for open doors, and choose between water and ice dispenser, and lock/unlock the touch input. Participants were allowed to move physically in 4m x 4m, so teleportation was unnecessary, and navigation was done solely by physical walking.

In the Context condition, the refrigerator was filled with food products, some of the standard sizes, such as milk or wine, to give a sense of the size of the interior of the product. The refrigerator was also presented in a home kitchen and other contextual appliances and furniture. In addition, there was an interactable instruction manual and a cup laid down on the counter near the refrigerator to aid the users during the interaction tasks with the refrigerator. In the Neutral condition, the interior of the refrigerator was empty. It was presented in an empty white room (the same room as Context condition but without any furniture and appliances). The only exception was a pedestal that was added by the side of the refrigerator where the instruction manual and a cup rested.

E. PROCEDURES

Before the experiment took place, a consent form and sociodemographic questionnaire were filled out by participants, and they were explained that the data collected was anonymous.There was no danger to the participant in performing the experiment, and they were free to withdraw at any time without any consequences. Participants were further explained how to use the controllers to interact with the environment. Then a researcher assisted the participants in equipping the VR headset. Half of the participants experienced the Contextualized condition, while the other half experienced the Neutral condition. Each participant only experimented one of the conditions. Participants were balanced by gender (8 females and 11 males in the Context condition and 9 females and 10 males in the Neutral condition).

Participants were placed in the virtual environment in front of the refrigerator in both conditions. Then they were given

TABLE 2. Description of the dependent and independent variables in the study and their sub-scales.

Dependent Variables	**Description**	**Sub-Scales**
Functionality clarification	The extent to which users were clarified about the functions of the product.	None
Size clarification	The extent to which users were clarified about the size of the product.	None
Memory	The extent users were aware of details about their experience with the product.	Correct Answers, Level of confidence
Workload	Evaluates how difficult users perceived the tasks.	Mental, Physical, Temporal, Performance, Effort, Frustration
Independent Variables	**Description**	**Levels**
Contextualisation	Whether the refrigerator had food products inside and shown within the context of a kitchen or empty and presented in a simple white room.	Context and Neutral
Gender	Biological gender of participants.	Male and Female

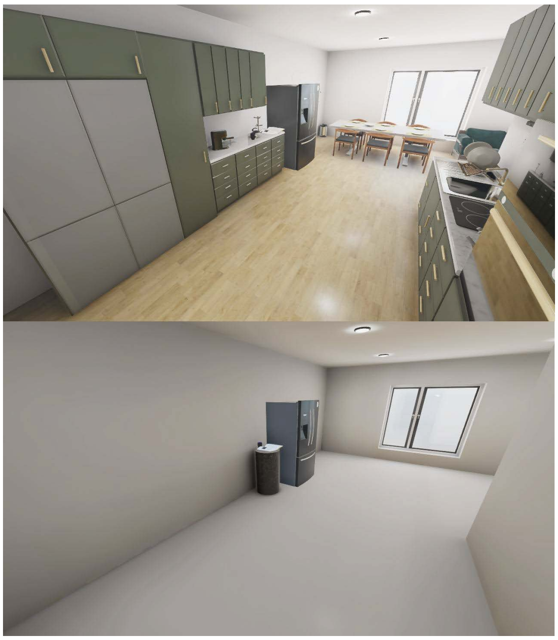


FIGURE 1. Top: Context condition. Bottom: Neutral context condition.

a series of tasks aimed at exploring the characteristics of the refrigerator:

1) Inspect the exterior of the refrigerator.
2) Open the doors and check the insides.
3) Open and close all drawers inside the refrigerator.
4) Close all drawers and doors.
5) Open the freezer drawer and check its interior.
6) Close the freezer.
7) Pick up the instructions manual.
8) Decrease the temperature of the refrigerator to 3°C.
9) Decrease the temperature of the freezer to its minimum.
10) Update the time and date to the correct ones (given by the researcher).
11) Set up the open door alarm to 1m30s.
12) Deactivate the open door alarm.

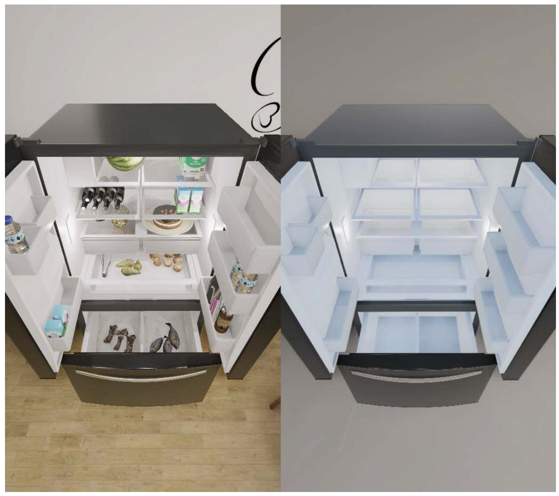


FIGURE 2. Left: Refrigerator filled with food products in Context condition. Right: Empty refrigerator in neutral context condition.

13) Lock and unlock the touchscreen.
14) Set dispenser mode to ice.
15) Pick up a cup to take some ice cubes from the ice dispenser.

The researcher then helped the participant to unequip, who then filled out the rest of the questionnaires. In the end, they were asked to give their opinion of the experiment (optionally).

F. STATISTICAL PROCEDURES

In one-way ANOVAs, the existence of outliers was verified by visual inspection boxplots. The Shapiro-Wilk test verified the assumptions of data normality, and homogeneity of variances was verified by Levene's test of homogeneity of variances. Because non-normality does not affect Type I error rate considerably in one-way ANOVAs when sample sizes are equal [77], in the cases that data was not normally distributed. Sample sizes were the same between groups (which was the case for the Neutral and Context conditions), a one-way

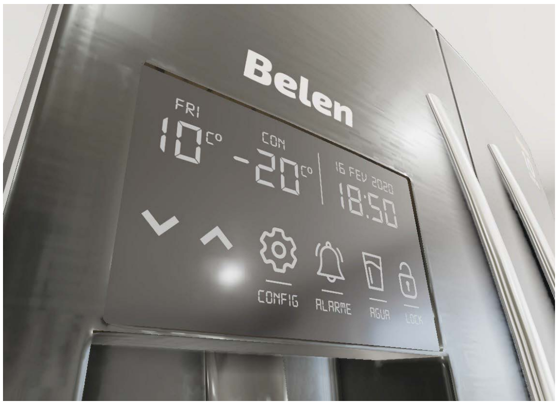


FIGURE 3. The touch screen UI of the refrigerator.

ANOVA would still be conducted. A Welch ANOVA was performed for the cases where homogeneity of variances was not met.

A Kruskal-Wallis H would be used when one-way ANOVAs could not be performed. The distributions of scores were verified by visual inspection of boxplots. For similar score distributions, judgements about medians' differences would be made. For non-similar score distributions, judgements based on differences in mean ranks would be done.

III. RESULTS

Means, standard deviations, medians, mean ranks and significance levels for all dependent variables between Context conditions, and genders, can be visualised in Tables 3 and 4.

A. FUNCTIONALITY CLARIFICATION

To investigate whether the context a product is presented influences the extent to which users were clarified about the product functionalities, we conducted a one-way ANOVA. No outliers were found, and data were not normally distributed. The homogeneity of variances assumption was violated. Instead, a Kruskal-Wallis H test was used. Distributions were similar between conditions. Size clarification medians were not statistically different between conditions ($\chi^2(1) = 0.419, p = 0.518$).

A one-way ANOVA was conducted to investigate possible differences between genders in Functionality Clarification. Four outliers were found in the female group and data was not normally distributed. Kruskal-Wallis H test was instead done. Distributions were not similar between conditions. Size clarification mean ranks were not statistically different between genders ($\chi^2(1) = 0.935, p = 0.334$).

As for the groups with and without past VR experience, we found no outliers, but data was not normally distributed for either group. A Kruskal-Wallis H test was done. Distributions were not similar between conditions. Results show that mean ranks were not significantly different ($\chi^2(1) = 0.948$, $p = 0.330$).

B. SIZE CLARIFICATION

A one-way ANOVA was executed to investigate if the context where a product is presented influences how users were clarified about the size of the product. No outliers were found, and data were not normally distributed. Variances were homogeneous. Results indicated no statistically significant differences in Size Clarification between Neutral and Context conditions ($F(1, 36) = 1.087, p = 0.304$).

A one-way ANOVA was done to investigate possible differences between genders. No outliers were found and data were not normally distributed. Kruskal-Wallis H test was instead performed. Distributions were not similar between conditions. Size clarification mean ranks were not statistically different between genders ($\chi^2(1) = 2.840, p = 0.092$).

No outliers were found in the Size Clarification variable between users with and without past VR experience. Still, data was not normally distributed, resulting in a Kruskal-Wallis H test being conducted instead. Distributions were similar between conditions. Results demonstrate that medians were not significantly different ($\chi^2(1) = 1.220, p = 0.269$).

C. MEMORY

Memory scores were divided into two for this analysis. We first investigated whether the user responded correctly to questions related to the product, and then analysed their overall confidence level.

1) CORRECT ANSWERS

Each correct answer is valued at one point, with a max of five points (all right answers) and zero (all wrong answers). A one-way ANOVA was conducted to evaluate if context influenced users' ability to remember the product characteristics. No outliers were found. Data were not normally distributed for the Context condition. Results indicated that context did not influence significantly the extent users remember the characteristic of the product ($F(1, 36) = 2.586, p = 0.117$).

As for gender, no outliers were detected, but data was only normally distributed in the female group. A Kruskal-Wallis H test was conducted. Distributions were similar between conditions. The correct answers score medians were not statistically different between genders ($\chi^2(1) = 0.553$, $p = 0.457$).

No outliers were found for neither correct answers nor confidence scores between groups with and without previous VR experience. Data were not normally distributed for the correct answers score in the group with past VR experience. Therefore, a Kruskal-Wallis H test was performed. Distributions were similar between groups. Results from the Kruskal-Wallis H test show that Correct Answers medians were not significantly different ($\chi^2(1) = 2.387, p = 0.122$).

2) CONFIDENCE

The confidence score consisted of the mean score of all of the five confidence questions that accompanied each question about the product's characteristics. A one-way ANOVA was conducted to evaluate if context influenced how confident users were about questions related to the product characteristics. No outliers were found. Data were normally distributed

TABLE 3. Descriptive statistics of dependent variables and their sub-scales between Context (n=19) and Neutral (n=19) conditions and significance level (η_p^2 for ANOVAs and Cohen's D for Kruskal-Wallis).

Dependent Variables	Mean		Std.Dev		Median		Mean Ranks		Sig.	Eff. Size
	Neut.	Cont.	Neut.	Cont.	Neut.	Cont.	Neut.	Cont.		
Functionality Clarification	6.470	6.680	0.772	0.478	7.000	7.000	18.530	20.470	0.518	$\eta_p^2 = 0.028$
Size Clarification	6.370	6.05	0.895	0.970	7.000	6.000	21.450	17.550	0.304	$\eta_p^2 = 0.029$
Correct Answers	3.160	3.680	1.015	1.003	3.000	4.000	16.530	22.470	0.117	$\eta_p^2 = 0.067$
Confidence	3.747	4.042	0.646	0.568	4.000	4.200	16.680	22.320	0.144	$\eta_p^2 = 0.058$
Workload Mental	4.260	5.470	3.998	3.549	3.000	6.000	17.660	21.580	0.330	$\eta_p^2 = 0.026$
Workload Physical	2.950	1.95	3.504	2.877	2.000	1.000	21.610	17.390	0.343	$\eta_p^2 = 0.025$
Workload Temporal	7.210	8.000	5.213	4.110	8.000	9.000	18.370	20.630	0.607	$\eta_p^2 = 0.007$
Workload Performance	4.840	3.420	4.180	3.185	3.000	3.000	21.610	17.390	0.246	$\eta_p^2 = 0.037$
Workload Effort	10.680	8.680	6.201	5.218	11.000	9.000	21.530	17.470	0.289	$\eta_p^2 = 0.031$
Workload Frustration	1.210	1.110	1.813	2.904	0.000	0.000	21.290	17.710	0.228	$d = 0.041$

TABLE 4. Descriptive statistics of dependent variables and their sub-scales between male (n=21) and female (n=17) groups and significance level ($\eta_p^2 = 0.008$ for ANOVAs and Cohen's D for Kruskal-Wallis).

Dependent Variables	Mean		Std.Dev		Median		Mean Ranks		Sig.	Effect Size
	Male	Female	Male	Female	Male	Female	Male	Female		
Functionality Clarification	6.520	6.650	0.602	0.702	7.000	7.000	18.190	21.120	0.334	$d = 0.199$
Size Clarification	5.950	6.530	1.071	0.624	6.000	7.000	16.980	22.620	0.092	$d = 0.662$
Correct Answers	3.570	3.240	0.926	1.147	3.000	3.000	20.640	18.090	0.457	$d = 0.317$
Confidence	3.943	3.835	0.559	0.697	4.000	3.600	20.240	18.090	0.144	$\eta_p^2 = 0.008$
Workload Mental	4.670	5.120	3.719	3.951	3.000	4.000	18.810	20.350	0.668	$d = 0.117$
Workload Physical	3.050	1.710	3.840	2.054	1.000	1.000	20.980	17.680	0.348	$d = 0.435$
Workload Temporal	8.050	7.060	4.544	4.854	9.000	9.000	20.620	18.120	0.522	$\eta_p^2 = 0.011$
Workload Performance	3.570	4.820	4.202	3.046	3.000	5.000	16.670	23.000	0.077	$d = 0.340$
Workload Effort	9.670	9.710	5.370	6.342	9.000	10.000	19.260	19.790	0.984	$\eta_p^2 < 0.000$
Workload Frustration	1.100	1.240	2.071	2.796	0.000	0.000	19.740	19.210	0.859	$d = 0.056$

for both conditions. There was homogeneity of variances. Results indicated that context did not influence significantly the users confidence ($F(1, 36) = 2.231, p = 0.144$).

A one-way ANOVA was also performed to verify gender differences in their confidence scores. No outliers were detected and data was normally distributed for both groups. There was homogeneity of variances. Results indicated that gender did not influence significantly confidence scores ($F(1, 36) = 0.279, p = 0.601$).

No outliers were found for confidence scores between groups with and without past VR experience. Data were normally distributed. The homogeneity of variances assumption was not violated in the one-way ANOVA. Results indicated that there are no significant differences in confidence levels between conditions ($F(1, 36) = 0.232, p = 0.633$).

D. PERCEIVED WORKLOAD

A one-way ANOVA was conducted to determine differences in the different Workload types (Mental, Physical, Temporal, Performance, Effort, and Frustration) between Neutral and Context conditions. Four outliers were found in the Frustration variable in the context condition. Thus, for this variable, a Kruskal-Wallis H test was performed.

No outliers were found in Temporal and Effort variables between conditions. Mental workload data was not normally distributed in the Neutral condition as well as Physical and Performance variables in both conditions. Nevertheless, a one-way ANOVA was still performed as the sample size was equal between groups. There was homogeneity of variances for all the workload related variables between conditions for this one-way ANOVA.

Mental scores ($F(1, 36) = 0.974, p = 0.330$), Physical scores ($F(1, 36) = 0.925, p = 0.343$), Temporal scores ($F(1, 36) = 0.269, p = 0.607$), Performance scores ($F(1, 36) = 1.389, p = 0.246$), and Effort scores ($F(1, 36) = 1.157, p = 0.289$) were not significantly different between Neutral and Context conditions.

Distributions of user Frustration scores were not similar for both groups. Kruskal-Wallis H test revealed that distributions of scores were not statistically significantly different between conditions ($\chi^2(1) = 1.451, p = 0.228$).

Due to no statistical differences between conditions, we considered all the female sample vs. male samples to

TABLE 5. Results for each hypothesis. R = Rejected, A = Accepted, P = Partially Accepted.

H1	H2	H3	H4	H5	H6	H7	H8
R	R	R	R	R	R	R	R

identify possible differences between genders for all workload variables. Two outliers were found in Performance scores in the male group. Other three outliers were detected in Frustration scores, two in the male group and one in the female group. Thus, we performed a Kruskal-Wallis H test for these two variables. No other outliers were found.

Mental scores in the female group were not normally distributed as well as Physical scores in both groups, Performance scores in the male sample, and Frustration scores in both groups. The rest of the data was normally distributed. Thus, a Kruskal-Wallis H test was performed for Mental, Physical, Performance, and Frustration scores, and one-way ANOVA for Temporal and Effort scores.

There was homogeneity of variances for both Temporal and Effort scores. Temporal scores ($F(1, 36) = 0.419$, $p = 0.522$), Effort scores ($F(1, 36) < 0.000, p = 0.984$) were not significantly different between genders.

Distributions of Mental, Physical, and Frustration scores were similar for all groups while Performance distributions were not. Median score were not statistically significantly different between genders for Mental $\chi^2(1) = 0.184$, $p = 0.668$, Physical $\chi^2(1) = 0.880, p = 0.348$, and Frustration scores $\chi^2(1) = 0.032, p = 0.859$. Performance mean ranks were not statistically significant between genders $\chi^2(1) = 3.132, p = 0.077$.

Comparing users with and without past VR experience, in all the workload scales, outliers were only found in the Frustration scale, one in each group. Data were not normally distributed in Physical, Performance (the group with past VR experience), and Frustration. Therefore, a Kruskal-Wallis H test was performed for these scales, and for the others, a one-way ANOVA.

Distributions were similar between conditions. Medians were not significantly different between conditions for Physical ($\chi^2(1) = 0.011, p = 0.915$), Performance ($\chi^2(1) = 0.073, p = 0.787$) and Frustration ($\chi^2(1) = 0.021$, $p = 0.886$) scales.

All the scales considered for the one-way ANOVA, met the homogeneity of variances assumption. Results indicated that Mental ($F(1, 36) < 0.000, p = 0.993$), Temporal ($F(1, 36) = 0.002, p = 0.962$), and Effort ($F(1, 36) = 0.117, p = 0.734$) were not significantly different between groups.

E. PARTICIPANT's FEEDBACK

Participants were asked to provide further details by sharing their thoughts about the experiment. Of the 38 participants, 18 shared their opinion. One subject suggested adding more information about the product, specifically the price and how eco-friendly the product was (power efficiency). One of the participants that performed the Context condition shared that, compared with his experience of a real environment, the virtual experience did not have some details. More specifically, the participant mentioned the hands (they were floating and not connected to a virtual body), and some reflections seemed erratic. Another participant mentioned that the hands would pass through the virtual objects. Similarly, one reported that the only technological problems were virtual hands and screen resolution. Adding to the screen resolution issue, another participant mentioned difficulties in reading the text in the manual because it was a bit blurry. Regarding the proportions and dimensions of the virtual space, in the context condition, one participant mentioned that the virtual world felt bigger when crouching to pick up a cup that fell to the ground.

We identified two contrasting opinions regarding the realism of the experience between the Context and Neutral conditions: one of the participants that performed the context condition stated that it felt like the tasks were being conducted in a real kitchen like in the real world while another participant that performed the neutral condition said that although everything looked normal and made sense, it felt like it was not real.

Overall, the majority of the participants that commented on their experience stated that the environment made sense to them and there was nothing out of the regular, and the product functionalities did what they were supposed to and in helpful time. Some mentioned that it would help them in their purchase decisions, and it was an interesting application to be applied in the field.

IV. DISCUSSION

This discussion is divided into subsections, one for each hypothesis. Results are synthesised in Table 5.

A. FUNCTIONALITY CLARIFICATION

We speculated that contextualisation would help users better clarify the product functionalities. Results indicated no differences between a product being presented in a contextualised environment or a neutral one led us to reject *H1*. The base functionalities of the product might have been well known by the participants because it should be a mundane and usual appliance in the participants' lives. If it was an unknown product with functionalities that users had never experienced and were unsure how they would help them, then displaying and trying such a product in its usual context could have made a difference. We further speculate that because users were focused on the task (reducing awareness of their surroundings and therefore mitigating possible effects of visual complexity) and because the tasks were the same between conditions, the processed information should have been the same in this regard.

We hypothesised that gender would have an impact on how clarified users were about the product functionalities. Results indicated no significant differences, thus rejecting *H4*. We suspect that participants were not willing to buy a refrigerator at that time and knowing the experience was only for research purposes might have led participants to ''simulate'' as if they were going to buy a refrigerator which is different from really looking for one as there would be money and logistics involved which would make them take the experience more seriously in this regard (thus showing the effects of the Selectivity Hypothesis [74]).

B. SIZE CLARIFICATION

The refrigerator is an equipment that requires a considerable amount of space in someone's kitchen. Its size should be an important characteristic to be taken into account. The size can sometimes be difficult to measure in VR without having other objects with known sizes for a mental comparison to exist. We hypothesised that presenting the refrigerator next to kitchen appliances and furniture should improve how users are clarified about the product size. The refrigerator's interior filled with food products, some of them in standard sizes (such as milk or wine bottles), should have further improved the user's notion of the interior size because they would serve as a comparison basis. However, the analysis revealed that contextualisation did not have an effect. Thus, users' levels of clarification about the product size were similar whether presented near other objects (kitchen furniture and appliances) and filled with food products, thus rejecting *H2*. It could be that users might have used their height and the size of the virtual hands as a comparison basis more than the food products inside and the furniture around it. The depth perception that IVR allowed could also have cued users enough to have a more precise notion about the size of the product. It could also be that users not knowing that there would be a question regarding the size of the product together with not looking to buy such a product and knowing it was only an experiment could have led them just not to pay attention to the product size.

Gender did not affect size clarification, which led us to reject *H6*. Once again, it could be that users were not simply looking for such a product to buy at the time of the experiment and thus were not actively analysing how important the size would be if they were to buy that product for their homes. This could lead to gender differences, if any, to be mitigated.

C. MEMORY

Memory was defined as the extent users can remember details from the product they experienced and their level of confidence about how certain they remember those characteristics correctly. Contextualisation implies more visual complexity. As indicated by state-of-the-art, such complexity may influence the capability of users to focus and remember correctly details of the experience [26], [40]. This analysis shows that this was not the case as contextualisation did not affect Memory, both in how correctly users remembered the refrigerator characteristics and how confident they were in those responses. As such, *H4* was rejected.

Mania *et al.* [40] study, which compared rendering quality (flat-shaded vs. radiosity), concluded that users performed better in the less realistic condition. Thus, a possible justification for this result is that, although both environments differed in contextualisation. Therefore, the level of visual complexity and the level of rendering quality did not change. Also, both situations of being in an empty white room or a kitchen filled with appliances and furniture, as well as seeing an empty or filled refrigerator, are coherent and possible to happen in reality. Therefore, if the realism level did not change between contexts then user performance (regarding Memory) would be similar.

Another possible justification was that participants performed the same tasks between conditions, which could reduce the effect ''change blindness'' and ''inattentional blindness'' of information that is outside their attention focus would have. Furthermore, Csikszentmihalyi's theory of flow [30] explains that users focused on a task can lose awareness of their surroundings. Therefore, although there was more visual complexity in the environment in the contextualised condition, the fact that users were focused on their tasks could have led them not to be aware of their surroundings and ultimately ''ignore'' the visual complexity outside their focus span.

Due to the Selectivity Hypothesis [74] and how genders pay attention to different characteristics in a product before purchasing it, we hypothesised that gender would influence the user's Memory. *H8* was rejected, and no differences were found both in how correctly users remembered specific characteristics of the product as well as their level of confidence in their responses. The same possible justification discussed in *H1* (functionality clarification) could also be applied here. If participants were not looking for such a product at that time while also knowing it was a simulation of a possible e-commerce application and not a real one could have led to the mitigation of gender differences.

D. PERCEIVED WORKLOAD

We expected that the increase in visual complexity caused by product contextualisation would influence the users' workload. Although, we also speculated that because users were guided by the researcher in doing a sequence of simple tasks, they would be more focused on what they were doing and not distracted by the surrounding environment, counteracting the effect of the distraction of the visual complexity. From the analysis done, we concluded that contextualisation did not influence the perceived workload scales, thus rejecting the *H4* hypothesis.

Further work is necessary to understand if this lack of workload differences could have been caused by the visual complexity effect on increasing the workload being counteracted by the fact that tasks were relatively simple and with no limit or scoring together with the user's focus on the tasks leading them to become less aware and thus less distracted by the surrounding environment, or if none of the above verified.

This presents evidence supporting that product contextualisation will not increase the user workload as long as they are guided through a sequence of tasks with no time constraints, keeping the users focused on the product yet aware of where they are.

We speculated that the user's workload would reflect this evidence of how males and females process information differently [74] when presented with a product to purchase. Results led us to reject the *H8* hypothesis. A possible justification for this result was already discussed: the fact that users knew it was only an experiment and that no actual product could be purchased at the end could have impaired the differences reported in the literature. It is also likely that participants were not actually looking for such a product to begin with, not triggering the processes that distinguish genders in the amount of information needed before purchase.

Another possibility is that users were guided equally through the same tasks to get to know the product. Thus, users did not get the freedom to explore the product on their own terms, which could lead to one gender exploring the product more in-depth and having a different workload type.

It is also possible that due to the tasks being simple and not time limited and because there was no evaluation of performance, participants' workload was not high enough to trigger differences between genders (if any).

V. CONCLUSION, LIMITATIONS AND FUTURE WORK

In this study, we investigated the concept of product contextualisation under an IVE setup as well as differences between gender. None of the hypotheses (clarification about product functionalities, its size, how well users remember the product characteristics, and the user's subjective workload) was accepted. Such put forward evidence that it would be preferable to present products in a neutral context, reducing costs related to human resources and computational needs to create a contextual environment for the displayed product. Further costs would also be spared as different context environments would need to be done for different product types.

Regarding the study limitations, participants were likely not looking to buy the product in the study (refrigerator) when they performed the experiment. They were also aware that it was solely an experiment and no product could be purchased at the end. These factors could have led participants to act and process information differently than what they would if it was an actual purchase experience. We tried to use a product that could have a use for everyone doing the experiment while also being suitable to be tested in IVR. Although a refrigerator is a product that is of use for everyone at home, the sample used (the majority being university students) might have not been that interested in the product either because they find it an ordinary everyday appliance. Another possibility is that they already have them in their homes, or if they live in university dormitories, then it would be irrelevant for them because they are not allowed to have such appliances in the dorms and also because the buildings already have several refrigerators to serve students.

When purchasing large and expensive items, people might invite family members or friends to help them through the item's search, purchase, and transportation process. Therefore, it would be valuable to check the impact of adding a multi-user functionality in this type of IVR application when analysing these types of items vs a single-user mode. However, the type of item should consider the cultural element of the sample being considered, as some might prefer purchasing the product alone while others might want company during the process. There are also products that, due to cultural reasons, people might feel uncomfortable checking and purchasing in the presence of other people (such as researchers, store employees, or friends either in physical form or as virtual avatars in an IVR environment.

Considering the participants' opinions of the experiment, further work should address the full information/specs about the product that would be available in a real purchase event. The hands/virtual body could also be improved and researched its impact. However, to keep the IVR application feasible to be used in real-case scenarios, at the users' homes or physical stores, using trackers to properly replicate the body movement could be a drawback in terms of costs and logistics. Some participants mentioned the text in the manual to be hard to read. Although participants could pick up the manual and read it at the distance they liked, the text size and available resolution might have made some participants very close to the manual to read it clearly, which some might have found challenging to do because of their eye convergence limits.

In future work, a comparison between analysing the same product in IVR or physically with different contexts (neutral, within context) would be valuable. To further test the thresholds of what consumers can ignore/accept in these types of experiences, an incoherent context could also be added to the investigation (such as showcasing a brand new car in the middle of a laundry room). Further work could also focus on leaving users to experiment with the product freely (without a sequence of tasks being asked by the researchers).

REFERENCES

[1] X. Liu, A. C. Burns, and Y. Hou, "Comparing online and in-store shopping behavior towards luxury goods," *Int. J. Retail Distrib. Manage.*, vol. 41, pp. 885–900, Nov. 2013.

[2] Y. Kim and R. Krishnan, "On product-level uncertainty and online purchase behavior: An empirical analysis," *Manage. Sci.*, vol. 61, no. 10, pp. 2449–2467, 2015.

[3] J. Delarue and T. Lageat, "Conducting contextualized and real-life product tests: Benefits and experimental challenges," in *Context*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 457–473.

[4] B. G. Englis and M. R. Solomon, "Using consumption constellations to develop integrated communications strategies," *J. Bus. Res.*, vol. 37, no. 3, pp. 183–191, Nov. 1996.

[5] W. Zhu and C. B. Owen, "Design of the PromoPad: An automated augmented-reality shopping assistant," *J. Organizational End User Comput.*, vol. 20, no. 3, pp. 41–56, Jul. 2008.

[6] H. Li, T. Daugherty, and F. Biocca, "Characteristics of virtual experience in electronic commerce: A protocol analysis," *J. Interact. Marketing*, vol. 15, no. 3, pp. 13–30, 2001.

[7] P. Fuchs, G. Moreau, and P. Guitton, *Virtual Reality: Concepts and Technologies*. Boca Raton, FL, USA: CRC Press, 2011.

[8] P. Milgram and F. Kishino, "Taxonomy of mixed reality visual displays," *IEICE Trans. Inf. Syst.*, vol. 77, no. 12, pp. 1321–1329, Dec. 1994.

[9] T. Monahan, G. McArdle, and M. Bertolotto, "Virtual reality for collaborative e-learning," *Comput. Educ.*, vol. 50, no. 4, pp. 1339–1353, 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0360131506001989

[10] L. Freina and M. Ott, "A literature review on immersive virtual reality in education: State of the art and perspectives," in *Proc. Int. Sci. Conf. eLearn. Softw. Educ.*, vol. 1. Bucharest, Romania: 'CAROL I' National Defence Univ. Publishing House, 2015, p. 133.

[11] T. P. Grantcharov, V. B. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, and P. Funch-Jensen, "Randomized clinical trial of virtual reality simulation for laparoscopic skills training," *Brit. J. Surgery*, vol. 91, no. 2, pp. 146–150, Jan. 2004, doi: 10.1002/bjs.4407.

[12] F. Aïm, G. Lonjon, D. Hannouche, and R. Nizard, "Effectiveness of virtual reality training in orthopaedic surgery," *Arthroscopy, J. Arthroscopic Rel. Surg.*, vol. 32, no. 1, pp. 224–232, Jan. 2016. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0749806315006489

[13] T. H. Kim and H. J. Choo, "Augmented reality as a product presentation tool: Focusing on the role of product information and presence in AR," *Fashion Textiles*, vol. 8, no. 1, pp. 1–23, Dec. 2021.

[14] D. Narciso, M. Melo, J. V. Raposo, J. Cunha, and M. Bessa, "Virtual reality in training: An experimental study with firefighters," *Multimedia Tools Appl.*, vol. 79, no. 9, pp. 6227–6245, 2020.

[15] P. Monteiro, M. Melo, A. Valente, J. Vasconcelos-Raposo, and M. Bessa, "Delivering critical stimuli for decision making in VR training: Evaluation study of a firefighter training scenario," *IEEE Trans. Hum.-Mach. Syst.*, vol. 51, no. 2, pp. 65–74, Apr. 2020.

[16] D. Narciso, M. Melo, S. Rodrigues, J. P. S. Cunha, and M. Bessa, "Impact of different stimuli on user stress during a virtual firefighting training exercise," in *Proc. IEEE 20th Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2020, pp. 813–818.

[17] H. Y. Kan, V. G. Duffy, and C.-J. Su, "An internet virtual reality collaborative environment for effective product design," *Comput. Ind.*, vol. 45, no. 2, pp. 197–213, Jun. 2001.

[18] T.-J. Nam and K. Sakong, "Collaborative 3d workspace and interaction techniques for synchronous distributed product design reviews," *Int. J. Design*, vol. 3, no. 1, 2009.

[19] K.-S. Suh and Y. E. Lee, "The effects of virtual reality on consumer learning: An empirical investigation," *Mis Quart.*, vol. 29, no. 4, pp. 673–697, Dec. 2005.

[20] S. Rao, E. Rabinovich, and D. Raju, "The role of physical distribution services as determinants of product returns in Internet retailing," *J. Operations Manage.*, vol. 32, no. 6, pp. 295–312, Sep. 2014.

[21] J. D. Shulman, M. Cunha, and J. K. Saint Clair, "Consumer uncertainty and purchase decision reversals: Theory and evidence," *Marketing Sci.*, vol. 34, no. 4, pp. 590–605, Jul. 2015.

[22] T. Foscht, K. Ernstreiter, C. Maloles, I. Sinha, and B. Swoboda, "Retaining or returning? Some insights for a better understanding of return behaviour," *Int. J. Retail Distrib. Manage.*, vol. 41, no. 2, pp. 113–134, Feb. 2013.

[23] *IKEA App Page*. Accessed: Feb. 10, 2022. [Online]. Available: https://www.ikea.com/au/en/customer-service/mobile-apps/say-hej-to-ikea-place-pub1f8af050

[24] N. Lobo, "Intelli-mirror: An augmented reality based IoT system for clothing and accessory display," in *Proc. Int. Conf. Internet Things Appl. (IOTA)*, Jan. 2016, pp. 95–100.

[25] M. Yuan, I. R. Khan, F. Farbiz, S. Yao, A. Niswar, and M.-H. Foo, "A mixed reality virtual clothes try-on system," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1958–1968, Dec. 2013.

[26] M. Wedel, E. Bigné, and J. Zhang, "Virtual and augmented reality: Advancing research in consumer marketing," *Int. J. Res. Marketing*, vol. 37, no. 3, pp. 443–465, Sep. 2020.

[27] M. V. Sanchez-Vives and M. Slater, "From presence to consciousness through virtual reality," *Nature Rev. Neurosci.*, vol. 6, no. 4, pp. 332–339, Apr. 2005.

[28] W. R. Sherman and A. B. Craig, *Understanding Virtual Reality: Interface, Application, and Design*. San Mateo, CA, USA: Morgan Kaufmann, 2018.

[29] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments," *Presence, Teleoperators Virtual Environ.*, vol. 6, no. 6, pp. 603–616, Dec. 1997.

[30] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. New York, NY, USA: Harper & Row, 1990.

[31] R. R. Burke, "Virtual reality for marketing research," in *Innovative Research Methodologies in Management*. London, U.K.: Palgrave Macmillan, 2018, pp. 63–82.

[32] J. Marín-Morales, J. L. Higuera-Trujillo, A. Greco, J. Guixeres, C. Llinares, E. P. Scilingo, M. Alcañiz, and G. Valenza, "Affective computing in virtual reality: Emotion recognition from brain and heartbeat dynamics using wearable sensors," *Sci. Rep.*, vol. 8, no. 1, pp. 1–15, Dec. 2018.

[33] M. I. Berkman and E. Akan, *Presence Immersion Virtual Reality*. Cham, Switzerland: Springer, 2019, pp. 1–10.

[34] J. J. Cummings and J. N. Bailenson, "How immersive is enough? A meta-analysis of the effect of immersive technology on user presence," *Media Psychol.*, vol. 19, no. 2, pp. 272–309, Apr. 2016.

[35] *Feelreal Multisensory VR Mask*. Accessed: Feb. 16, 2022. [Online]. Available: https://feelreal.com/

[36] F. E. Jamiy and R. Marsh, "Distance estimation in virtual reality and augmented reality: A survey," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, May 2019, pp. 063–068.

[37] S. Masnadi, K. Pfeil, J.-V.-T. Sera-Josef, and J. LaViola, "Effects of field of view on egocentric distance perception in virtual reality," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2022, pp. 1–10.

[38] S. Davis, K. Nesbitt, and E. Nalivaiko, "A systematic review of cybersickness," in *Proc. Conf. Interact. Entertainment*, Dec. 2014, pp. 1–9.

[39] D. A. Bowman and R. P. McMahan, "Virtual reality: How much immersion is enough?" *Computer*, vol. 40, no. 7, pp. 36–43, Jul. 2007.

[40] K. Mania, D. Wooldridge, M. Coxon, and A. Robinson, "The effect of visual and interaction fidelity on spatial cognition in immersive virtual environments," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 3, pp. 396–404, May 2006.

[41] A. M. Barreto, "Do users look at banner ads on Facebook?" *J. Res. Interact. Marketing*, vol. 7, no. 2, pp. 119–139, May 2013.

[42] L.-P. Schneider, B. Systems, and T. B. Cornwell, "Cashing in on crashes via brand placement in computer games," *Int. J. Advertising*, vol. 24, no. 3, pp. 321–343, 2005. [Online]. Available: https://doi.org/10.1080/02650487.2005.11072928

[43] H. S. Krishnan and D. Chakravarti, "Memory measures for pretesting advertisements: An integrative conceptual framework and a diagnostic template," *J. Consum. Psychol.*, vol. 8, no. 1, pp. 1–37, Jan. 1999.

[44] C. Peukert, J. Pfeiffer, M. Meißner, T. Pfeiffer, and C. Weinhardt, "Shopping in virtual reality stores: The influence of immersion on system adoption," *J. Manage. Inf. Syst.*, vol. 36, no. 3, pp. 755–788, 2019.

[45] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vis. Res.*, vol. 40, nos. 10–12, pp. 1489–1506, Jun. 2000.

[46] V. Sundstedt, K. Debattista, P. Longhurst, A. Chalmers, and T. Troscianko, "Visual attention for efficient high-fidelity graphics," in *Proc. 21st Spring Conf. Comput. Graph. (SCCG)*. New York, NY, USA: Association for Computing Machinery, 2005, pp. 169–175.

[47] D. J. Simons and C. F. Chabris, "Gorillas in our midst: Sustained inattentional blindness for dynamic events," *Perception*, vol. 28, no. 9, pp. 1059–1074, 1999.

[48] J. Leppink, "Cognitive load theory: Practical implications and an important challenge," *J. Taibah Univ. Med. Sci.*, vol. 12, no. 5, pp. 385–391, Oct. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1658361217300835

[49] E. J. Adams, A. T. Nguyen, and N. Cowan, "Theories of working memory: Differences in definition, degree of modularity, role of attention, and purpose," *Lang., Speech, Hearing Services Schools*, vol. 49, no. 3, pp. 340–355, 2018.

[50] J. Eriksson, E. K. Vogel, A. Lansner, F. Bergström, and L. Nyberg, "Neurocognitive architecture of working memory," *Neuron*, vol. 88, no. 1, pp. 33–46, Oct. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0896627315007771

[51] M. D. Basil, *Multiple Resource Theory*. Boston, MA, USA: Springer, 2012, pp. 2384–2385. [Online]. Available: https://doi.org/10.1007/978-1-4419-1428-6_25

[52] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, "State of science: Mental workload in ergonomics," *Ergonomics*, vol. 58, no. 1, pp. 1–17, Jan. 2015. [Online]. Available: https://doi.org/10.1080/00140139.2014.956151

[53] A. D. Souchet, D. Lourdeaux, A. Pagani, and L. Rebenitsch, "A narrative review of immersive virtual reality's ergonomics and risks at the workplace: Cybersickness, visual fatigue, muscular fatigue, acute stress, and mental overload," *Virtual Reality*, pp. 1–32, Jul. 2022, doi: 10.1007/s10055-022-00672-0.

[54] A. Negu, S.-A. Matu, F. A. Sava, and D. David, "Task difficulty of virtual reality-based assessment tools compared to classical paper-and-pencil or computerized measures: A meta-analytic approach," *Comput. Hum. Behav.*, vol. 54, pp. 414–424, Jan. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563215301059

[55] R. Shen, D. Weng, S. Chen, J. Guo, and H. Fang, "Mental fatigue of long-term office tasks in virtual environment," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, 2019, pp. 124–127.

[56] A. Geiger, I. Bewersdorf, E. Brandenburg, and R. Stark, "Visual feedback for grasping in virtual reality environments for an interface to instruct digital human models," in *Proc. Int. Conf. Appl. Hum. Factors Ergonom.* Cham, Switzerland: Springer, 2017, pp. 228–239.

[57] H. Wu, Y. Deng, J. Pan, T. Han, Y. Hu, K. Huang, and X. L. Zhang, "User capabilities in eyes-free spatial target acquisition in immersive virtual reality environments," *Appl. Ergonom.*, vol. 94, Jul. 2021, Art. no. 103400.

[58] B. Gao, Z. Chen, X. Chen, H. Tu, and F. Huang, "The effects of audio-visual landmarks on spatial learning and recalling for image browsing interface in virtual environments," *J. Syst. Archit.*, vol. 117, Aug. 2021, Art. no. 102096.

[59] V. Biener, D. Schneider, T. Gesslein, A. Otte, B. Kuth, P. O. Kristensson, E. Ofek, M. Pahud, and J. Grubert, "Breaking the screen: Interaction across touchscreen boundaries in virtual reality for mobile knowledge workers," 2020, *arXiv:2008.04559*.

[60] D. Zielasko, B. Weyers, and T. W. Kuhlen, "A non-stationary office Des. substitution for Des.-based and HMD-projected virtual reality," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2019, pp. 1884–1889.

[61] M. Speicher, P. Hell, F. Daiber, A. Simeone, and A. Krüger, "A virtual reality shopping experience using the apartment metaphor," in *Proc. Int. Conf. Adv. Vis. Interfaces*, 2018, pp. 1–9.

[62] P. Knierim, V. Schwind, A. M. Feit, F. Nieuwenhuizen, and N. Henze, "Physical keyboards in virtual reality: Analysis of typing performance and effects of avatar hands," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–9.

[63] D. D. Heikoop, J. C. de Winter, B. van Arem, and N. A. Stanton, "Effects of platooning on signal-detection performance, workload, and stress: A driving simulator study," *Appl. Ergonom.*, vol. 60, pp. 116–127, Apr. 2017.

[64] K. Prasad, S. Poplau, R. Brown, S. Yale, E. Grossman, A. B. Varkey, E. Williams, H. Neprash, and M. Linzer, "Time pressure during primary care office visits: A prospective evaluation of data from the healthy work place study," *J. Gen. Internal Med.*, vol. 35, no. 2, pp. 465–472, Feb. 2020.

[65] J. A. W. Filho, W. Stuerzlinger, and L. Nedel, "Evaluating an immersive space-time cube geovisualization for intuitive trajectory data exploration," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 1, pp. 514–524, Jan. 2019.

[66] J. A. Wagner Filho, C. M. Freitas, and L. Nedel, "Comfortable immersive Analytics with the VirtualDesk metaphor," *IEEE Comput. Graph. Appl.*, vol. 39, no. 3, pp. 41–53, Feb. 2019.

[67] J. A. Wagner Filho, C. M. D. S. Freitas, and L. Nedel, "VirtualDesk: A comfortable and efficient immersive information visualization approach," *Comput. Graph. Forum*, vol. 37, no. 3, pp. 415–426, Jun. 2018.

[68] S. Sepp, S. J. Howard, S. Tindall-Ford, S. Agostinho, and F. Paas, "Cognitive load theory and human movement: Towards an integrated model of working memory," *Educ. Psychol. Rev.*, vol. 31, no. 2, pp. 293–317, Jun. 2019.

[69] X. Lin, M. Featherman, S. L. Brooks, and N. Hajli, "Exploring gender differences in online consumer purchase decision making: An online product presentation perspective," *Inf. Syst. Frontiers*, vol. 21, no. 5, pp. 1187–1201, Oct. 2019.

[70] L. D. Wolin and P. Korgaonkar, "Web advertising: Gender differences in beliefs, attitudes and behavior," *Internet Res.*, vol. 13, no. 5, pp. 375–385, Dec. 2003.

[71] E. Garbarino and M. Strahilevitz, "Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation," *J. Bus. Res.*, vol. 57, no. 7, pp. 768–775, 2004.

[72] F. J. Pascual-Miguel, A. F. Agudo-Peregrina, and J. Chaparro-Peláez, "Influences of gender and product type on online purchasing," *J. Bus. Res.*, vol. 68, no. 7, pp. 1550–1556, 2015.

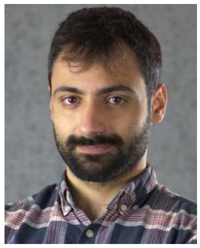
[73] V. Venkatesh, J. Y. Thong, and X. Xu, "Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology," *MIS Quart.*, vol. 36, no. 1, pp. 157–178, Mar. 2012.

[74] J. Meyers-Levy and D. Maheswaran, "Exploring differences in males' and females' processing strategies," *J. Consum. Res.*, vol. 18, no. 1, pp. 63–70, 1991.

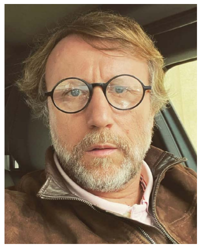
[75] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, Apr. 1988.

[76] S. G. Hart, "Nasa-task load index (NASA-TLX); 20 years later," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*. Los Angeles, CA, USA: Sage publications, vol. 50, no. 9, 2006, pp. 904–908.

[77] L. M. Lix, J. C. Keselman, and H. J. Keselman, "Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance f test," *Rev. Educ. Res.*, vol. 66, no. 4, pp. 579–619, Dec. 1996.

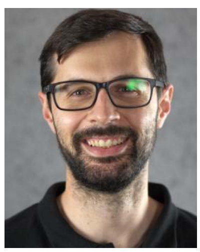


GUILHERME GONÇALVES received the bachelor's degree from the University of Trás-os-Montes e Alto Douro (UTAD), Portugal, in 2015, where he is currently pursuing the M.Sc. degree in multimedia. Since 2018, he has been a Research Fellow with the INESC TEC, Porto, Portugal. His research interest includes multisensory virtual reality.



GALVÃO MEIRINHOS received the B.A. degree in advertising and public relations, the B.A. degree in marketing, and the M.Sc. and Ph.D. degrees in information and communication sciences from the Universitat Autònoma de Barcelona. He is currently a Professor with the University of Trás-os-Montes and Alto Douro, Portugal, and the Department of Letters, Arts and Communication.

VÍTOR FILIPE received the M.S. degree in informatics from the University of Minho, Portugal, in 1997, and the Ph.D. degree in electrical engineering from the University of Trás-os-Montes e Alto Douro (UTAD), Portugal, in 2003. Since 2015, he has been a Senior Researcher with the INESC TEC. He is currently an Associate Professor with Habilitation in electrical engineering with the School of Science and Technology, UTAD. His research interests include computer vision, machine learning, and gait analysis.



MIGUEL MELO is currently an Assistant Researcher at INESC TEC, specialized in computer graphics. He is also the Manager of the Multisensory Virtual Reality Laboratory. His research interests include computer graphics, HDR, and multisensory virtual reality. He is also a member of the Executive Committee of Eurographics.

MAXIMINO BESSA is currently an Associate Professor with Habilitation at the Department of Engineering, University of Trás-os-Montes and Alto Douro, Portugal. He has been a Senior Researcher with the INESC TEC, since 2009, and the Director of the Multisensory Virtual Reality Laboratory MASSIVE. He has been a member of the Eurographics Association, since 2003, and the President of the Portuguese Computer Graphics Chapter.