Validation of a virtual reality risk assessment tool: the VR balloon analogue risk task

Yannick Oude Heuvel

University of Twente

Faculty of behavioral, management and social sciences

Human factors and engineering psychology

Master thesis

1st supervisor: Dr. Simone Borci

2nd supervisor: Jeanette Hadaschik MSc.

7th of June, 2023

Preface

Finishing my master thesis has been one of the most difficult, but also most rewarding things I have done in my life. Writing my bachelor thesis had been a very difficult experience for me (partly due to the complications of covid, which happened during that time), so it was incredibly important to me that my experience writing my master thesis would be better. I am happy to say that it was! Working on this thesis required a level of responsibility and initiative that I had not yet experienced in any other subject or school assignment, and it made me grow, not just professionally but also personally. I am proud of the result and happy that I can enter a new chapter in my life, looking back to the conclusion of my higher education with smile.

Thanks to everyone that helped me during this process, even if it was just by listening to me complain when I needed it. I want to give a special thanks though, to a few people in particular. I want to thank my supervisors Simone and Jeanette, for showing me how to think and act as a research professional. It has been a great pleasure to be taught, and treated as an equal by two people that I look up to, and consider great researchers. I want to thank Nesya, who has worked on this project before me, and was always willing to make time and help me whenever I needed it. I want to thank my girlfriend Eline, who was my emotional support and who kept me sane during this insane process. Finally, I want to thank you for taking the time to read my master thesis, I hope you'll enjoy it!

Abstract

In this study an attempt is made to validate a virtual reality (VR) version of the Balloon Analogue Risk Task (BART). The BART is a behavioral measure of a person's risk taking propensity (RTP). Using a virtual reality environment instead of a regular computer interface for this task can make the task more closely resemble a real situation, and thereby improve its validity.

The original BART validation study (Lejuez et al., 2002) is replicated, with a VR BART in addition to a regular computer-based (CB) BART. The results found in the original study, were not found with the VR BART. Therefore, the VR BART can not be validated as an assessment tool for RTP at this point. However, by comparing the CB and VR BARTs in the current sample, insights into the the use of VR for RTP assessment are gained. By comparing how people do on both tasks, this study shows that people act more risk-averse in a VR BART. Furthermore, by comparing subjective immersion in the CB and VR BARTs in the current sample, it is shown that people feel a higher sense of immersion when doing the task in VR. This study argues that the fact that people take less risks in the same task when they are more immersed in the task environment, indicates that a VR BART is better at capturing peoples natural behavior than a CB BART. More generally, this study argues that being able to understand risk-taking behavior on an individual level, will require the design of immersive and realistic test situations, which could be made with VR.

Table of contents

Introduction

Taking a risk is usually defined as engaging in behavior that involves some danger or chance of harm (Leigh, 1999; Trimpop, 1994). People accept these dangers because there is some reward that they might gain from the behavior. When people choose whether to take a risk, they compare the potential dangers of the behavior, and the chance of those dangers happening, with the potential benefit of the behavior (Vlek & Stallen, 1980).

Perceived risk can be defined as the reaction of an individual to this calculation (Vlek & Stallen, 1980). Two people might perceive a different level of risk for the same situation, as this is affected by many components other than the situation itself (Slovic, 1964). Several authors investigated the link between individual differences and risk-taking behavior. Previous studies suggested, for instance, that risk-taking behavior is related with aspects such as personality factors (Bogg and Roberts, 2004; Gullone & Moore, 2004), age (Jozef et al., 2016), and gender (Byrnes et al., 1999). The likelihood of an individual engaging in risk-taking behavior regardless of the situation can be summarized by the term risk-taking propensity (RTP) (Slovic, 1964).

It is important to be able to understand and predict risk-taking behavior. Risk-taking behavior, by definition, carries some sort of danger. Smoking, stealing, doing drugs, etc. all have the potential for substantial negative outcomes for the person engaging in them, and for society at large. Being able to assess RTP can lead to a better understanding of the causes of risk-taking behavior which could help prevent risky situations. Therefore, it is important to be able to accurately assess RTP.

Assessment of RTP can be done in different ways. RTP is often assessed using self-report measures for constructs that are thought to be related to risk-taking (e.g., venturesomeness (Eysenck et al., 1985), impulsivity (Patton et al., 1995) or sensation-seeking (Zuckerman et al., 1978)) or behaviors that are considered risk-taking (e.g., smoking, drug use, or alcohol consumption). There are disadvantages to using self-report measures to assess RTP though. Firstly, self-report measures can measure factors that are related to RTP, but due to the multifaceted nature of RTP, they cannot assess it in its entirety. This is especially clear with the assessment of behavioral outcomes of risk-taking: someone might smoke daily (associated with high RTP) but never drink (associated with low RTP). Secondly, questionnaires are self-report tools, and therefore respondents can be affected by two types of bias are likely to occur in self-report data. First, there is social desirability bias, which causes people that are filling in a questionnaires to answer in a way that makes them look more socially desirable, especially when answering questions about sensitive subjects like drug use (Althubaiti, 2016). Then, there is recall bias, which causes people to sometimes misremember or misattribute when answering questions about themselves and situations they have been in, causing them to give a false account of the situation (Althubaiti, 2016).

To counteract some of the disadvantages of self-report measures, and assess RTP more accurately, Lejuez et al (2002) developed the Balloon Analogue Risk Task (BART). The BART is a behavioral measure of RTP, usually done on a computer, where participants are tasked with filling up balloons with air. Specifically, participants are informed that they will receive an amount of money depending on how much air they pump into the balloon, but they will not get any money if the balloon explodes. Every balloon takes a different number of pumps to explode, and the participant can collect their reward whenever they choose. The goal of the task for the participant is to pump as much air into the balloon as possible, without it exploding.

While the assessment of RTP has come a long way with the BART, it can still be improved. Researchers have found that the BART has low convergent validity with other measures of RTP that might impede its use as a valid assessment tool (Duckworth & Kern, 2011; Hopko et al., 2005; Lauriola et al., 2014; Steiner & Frey, 2021). In particular, researchers have argued that by increasing the realism of the BART, participants' reactions will also be more realistic, and it will thereby increase the BART’s (ecological) validity (i.e. to what extent findings in a research setting can be generalized to a real world setting) (de-Juan-Ripoll et al., 2018; Slater, 2009). Recently, researchers have started using virtual reality (VR) in RTP assessment to achieve that end (Delanie et al., 2021; de-Juan-Ripoll et al., 2021). Other than the ability to instill more realism into RTP assessment, by creating immersive research environments and realistic stimuli, VR has a few additional practical advantages as a research tool. Using VR can make it easier to do stealth assessments, meaning the assessment is interwoven and integrated, for instance, in a gamified learning task in a way that is not perceived by the users as a separate form of evaluation (de-Juan-Ripoll et al., 2018). Stealth assessment has the potential to decrease noise from data, as participants will be less affected by response bias and test anxiety (Giglioni et al., 2017). Furthermore, VR makes it easier and more natural to monitor physiological data, as measurement tools can often be incorporated in the VR hardware (Bohil et al. 2011; Halbig & Latoschik, 2021). One disadvantage of using VR that should be taken into consideration is the relatively often occurring motion sickness for users of VR (Chatta et al., 2020), which could make it difficult to finish an assessment for some people. Despite the practical disadvantage of motion sickenss, VR can still offer multiple advantages to RTP assessment, including several practical advantages and most importantly increasing the immersion felt by the person tested. Therefore, the BART task can be improve by using a VR environment to display the task, as this will make it a better assessment tool, due to increasing its ecological validity

Hypotheses

The BART is a well-established tool for the assessment of RTP, but its ecological validity could potentially be improved by using a VR environment to display the task instead of a normal graphical user interface (GUI). In this research, the validity of a VR version of the BART, and the immersion felt by those that do the task in VR (subjective immersion), are investigated. The primary goal of this research is to validate the VR BART as a tool to measure RTP. To achieve this, the original BART validation study by Lejuez et al. (2002) is replicated. Furthermore, the original study is extended, by also investigating whether a VR BART would elicit similar results to the computer-based (CB) version of the task. Specifically, like in the original study, adjusted number of pumps (ANP) is calculated, by adding the number of times participants pumped for every balloon that did not end up exploding together. It is investigated whether there is any difference in the ANP of participants when using the CB BART and VR BART. Then, it is investigated whether correlations between ANP and scores on self-report measures for risk-constructs and risk behaviors that were found by Lejuez et al. (2002), can be replicated when participants perform the BART in a VR environment. Finally, to confirm the potential upside of using VR, it is investigated whether people indeed feel that they are more immersed in a VR environment when doing the BART. To achieve these goals, the following three hypotheses were investigated:

H1: There is no significant difference between ANP on the CB BART and the VR BART

H2: Risk-related constructs and risk behaviors that were correlated with ANP in the original research by Lejuez et al. (2002) are also correlated with ANP on the VR BART of the participants in the present study

H3: Subjective immersion is significantly higher for participants when they perform the VR BART as compared to the CB BART

Methods

Participants

Data from 20 people (11 men and 9 women) between the ages of 19 and 56 (m = 28.2, sd = 9.2) was used. Of these 20 participants, 17 were Dutch (85%), one was German (5%), one was Bulgarian (5%) and one was from Luxembourg (5%). Furthermore, only two participants (10 %) had no former experience with VR. Participants were recruited in two different ways. 18 participants (90%) were recruited by convenience sampling: through the personal network of the researcher, people were asked to participate in the study. Two participants (10%) were recruited by using the University of Twente's participant recruitment platform (SONA). Participants had to be over 18 years old. Of the 24 people that signed up, one did not show up, and three got sick during the VR portion of the experiment and had to quit prematurely.

Materials

The material used in the present experiment can be divided in two main components: *i) the battery of self-report measures for various risk-related constructs, risk behaviors, subjective immersion, and motion sickness;* and *ii) the CB and the VR BART.*

The battery of self-report measures

A battery of self-report measures similar to the one used by Lejuez et al. (2002) was used. For a list of changes between the original battery and the current battery see Appendix A.

Risk-related constructs

***Sensation seeking*.** Sensation seeking was measured using the 8-item Brief Sensation Seeking Scale (BSSS; Hoyle et al., 2002). It assesses four factors of sensation seeking (experience seeking, boredom susceptibility, thrill and adventure seeking) using a 5-point Likert scale. Hoyle et al. (2002) reported an internal consistency of .76.

***Impulsivity*.** Impulsivity was measured using the 15-item Baratt Impulsiveness Scale short form (BIS-15; Spinella, 2007). It assesses three factors of impulsivity (attention impulsivity, motor impulsivity and non-planning) on a 4-point Likert scale. Spinella (2007) reported an internal consistency of .79.

***Impulsivity/Venturesomeness*.** Impulsivity was also measured alongside venturesomeness using the impulsivity (I7-I) and venturesomeness (I7-V) subscales of the Eysenck Impulsiveness Scale (I7; Eysenck et al., 1985). Both are assessed using yes/no questions. Impulsivity was assessed using 19 items, and venturesomeness using 16 items. Eysenck et al. (1985) reported internal consistencies for both subscales controlled by gender. Impulsivity had internal consistencies of 0.84 (male) and 0.83 (female) and venturesomeness had internal consistencies of 0.85 (male) and 0.84 (female).

***Behavioral constraint*.** Behavioral constraint was measured using the 37-item behavioral constraint superfactor of the multidimensional personality questionnaire brief form (MPQ-BF; Patrick et al., 2002). The behavioral constraint superfactor consists of the control, harm avoidance and traditionalism, and is assessed with true/false questions. Patrick et al. (2002) reported internal consistencies of .74, .76 and .78 for the control, harm avoidance and traditionalism scales respectively.

Risk behaviors

***Alcohol consumption*.** Alcohol consumption was measured using the 10-item Alcohol Use Disorders Identification Test (AUDIT; Saunders et al., 1993), which assesses quantity and frequency of drinking, symptoms of dependence, adverse psychological reactions, with multiple choice items. Its reliability has since been tested, with a review from Reinert & Allen (2007) finding 18 studies from 2002 till 2005 which have reported internal consistencies ranging from .75 to .97 (median = .83).

***Drug use*.** Drug use was measured by asking the participant how much of several classes of drugs they had consumed over the past 12 months. The following classes of drugs are used: 1. cocaine, 2. heroin, 3. marijuana, 4. amphetamines, 5. inhalants, 6. tranquilizers, 7. hallucinogens, 8. ecstasy, 9. opioid analgesics, 10. synthetic drugs, 11. anabolic androgenic steroids, 12. other.

***Gambling behavior*.** Gambling behavior was measured using a 23-item version of the Gambling Attitudes and Beliefs Survey (GABS-23; Bouju et al., 2013). The GABS-23 assesses gambling behavior along five factors: attitudes, chasing, emotions, luck, and strategy, on a 4-point Likert scale. Bouju et al. (2013) validated the GABS-23 with 2 replicate studies, reporting an internal consistency of .93 for the first, and .89 for the second.

***Behaviors measured with a single item*.** Cigarette use, risky sexual behavior, stealing, and seatbelt use were measured using the same single items that were used by Lejuez et al (2002). Participants were asked the following questions. For cigarette use: “what is the average number of cigarettes you smoke daily?”. For sexual behavior: “over the past 12 months, with how many people have you had sexual intercourse without a condom?”. For stealing: “over the last 12 months, how many times have you stolen something from the store?”. For seatbelt use: “what percentage of time do you wear a seatbelt while driving in a car?”.

Additional self-report measures

In contrast to the original study, one measure was added to assess the subjective immersion felt by participant in both environments. Moreover, one measure was added to assess motion sickness during the VR-simulation.

Subjective Immersion**.** Sense of presence was measured using the 11-item I-group presence questionnaire (IPQ; Berkman & Çatak, 2021). The IPQ measures a users subjective sense of being in a virtual environment, and this was interpreted as a marker of subjective immersion. The IPQ assesses sense of presence in a virtual environment along three factors: spatial presence, involvement, and realism. Additionally, it contains one item that assesses 'general sense of being there'. The IPQ uses a 5-point Likert scale. Berkman and Çatak (2021) reported internal consistencies of 0.77, 0.68 and 0.65 for spatial presence, involvement, and realism respectively.

Motion sickness**.** The virtual reality sickness questionnaire (VRSQ, Kim et al., 2018) was used to assess motion sickness symptoms in participants. The VRSQ measures symptoms specifically related to motion sickness while in virtual reality. It assesses virtual sickness along two factors: oculomotor (4 items) and disorientation (5 items).

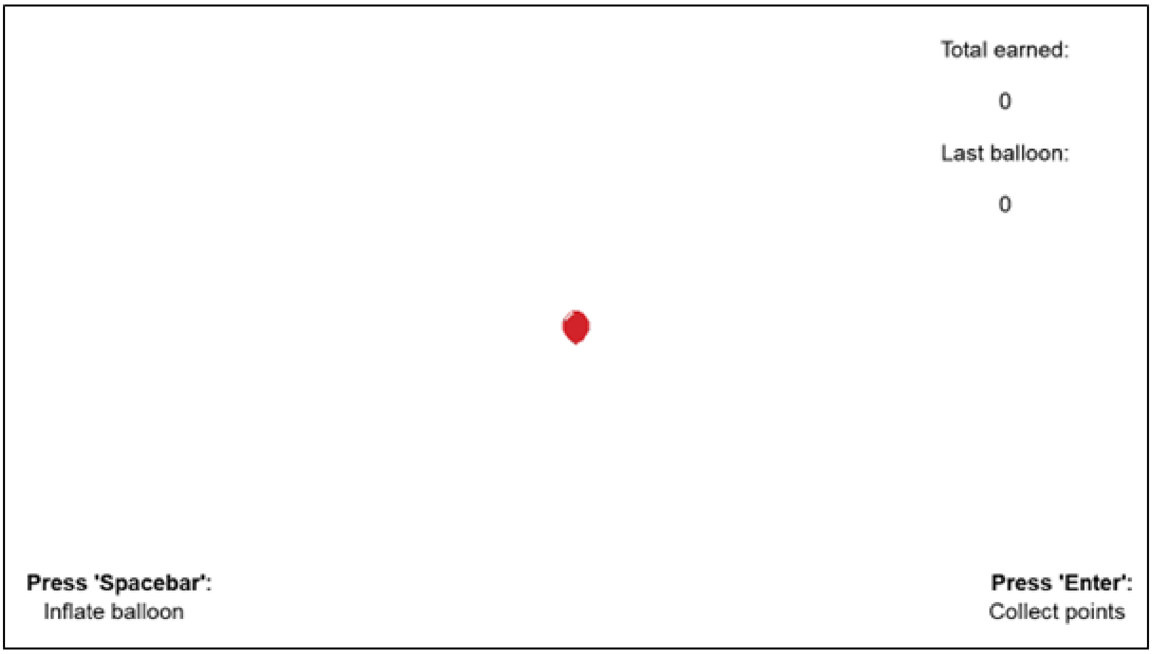
General properties of the BART

The BART is a task in which pariticipants pump air into balloons. Every time a participant pumps, the balloon gets bigger and they store points in a temporary bank. At any point, the participant can either choose to pump, or stop pumping and collect the points in their temporary bank, at which point these points are moved to a permanent bank and the participant moves on to the next balloon. The balloons all have 'breaking points'. When the number of pumps go past this breaking point, the balloon will explode, the participant loses any points in the temporary bank, and they move onto the next balloon.

Computer based BART. The computer-based BART GUI consisted of: a red balloon, a label that says "Press 'Spacebar': Inflate balloon", a label that says "Press 'Enter': Collect points", a label that says "Total earned:" and displays the total points earned for this round (every balloon added up), and a label that says ”Last balloon:” and displays the points earned on the last balloon (see figure 1). The computer-based BART was made using Opensesame (version 3.3.12), an open-source program based on python that is used to create behavioral experiments (Mathôt et al., 2012).

Figure 1

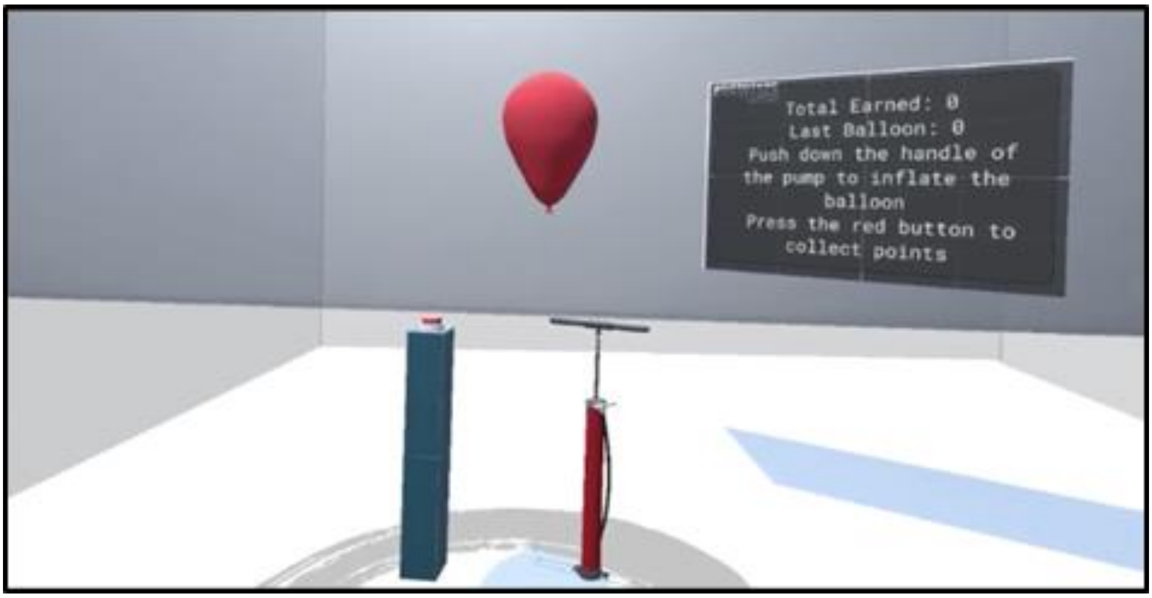
GUI for the computer-based BART



Virtual reality BART. The VR BART environment consisted of: a white floor and 4 glass walls closing it off, a red balloon, a red button, a red hand pump, and a large screen with the following text: “Total earned:” (displaying the total points earned this round), “Last balloon:” (displaying the points earned on the last balloon), “Push down the handle of the pump to inflate the balloon”, and “Press the red button to collect points” (see figure 2). The VR BART was made using Unity (version 2020.3.28f1), a tool for developing games and other 3d projects that based on c# (Haas, 2014).

Figure 2

Virtual environment for the virtual reality BART



Other materials

A online questionnaire made with qualtrics, an tool for creating questionnaires (*Qualtrics XM: The Leading Experience Management Software*, 2023), was used. This questionnaire included an informed consent form, all the different self-report measures, and directions to take part in the CB and VR BARTs.

A special area was prepared where the participant could sit during the VR BART. This area consisted of a chair with a lot of space around it. This was done to ensure the participant would not bump in to anything when they were doing the VR BART.

Lastly, the following hardwares were used: (i) an Oculus rift S VR headset, (ii) a laptop on which the participant filled in the self-report battery and did the CB BART, (iii) a desktop computer that was used to run the unity environment and the Oculus software.

Procedure

A within subjects design was used, with all participants being subjected to the same self-report battery, and both the CB and VR BARTs. However, the order in which the participants performed the CB and VR BART was randomized for each participant. When participants were recruited, they were informed that they would have to fill in several questionnaires, partake in two 'games' involving balloons, and that the experiment would take between one and two hours. They were also informed that the two people with the top combined scores in the games would receive a goodie bag. When participants arrived at the experiment location, this information was restated to them. Participants were then asked to sit in front of a computer, where a Qualtrics questionnaire was opened. The first part of the questionnaire was an informed consent form. When the participant had given informed consent, the Qualtrics questionnaire continued and they had to complete the BSSS, the BIS-15, the I7-I, the I7-V, the behavioral constraint questions of the MPQ-BF, the AUDIT, the GABS, the drug use questions, and the single items for cigarette use, sexual behavior, stealing, and seatbelt use. To limit order effects, participants completed these questionnaires in a random order. After they completed the self-report measures, they had to complete the CB BART and the VR BART, and the IPQ after both BARTs respectively. The order of the BART tasks was also randomized. This was done to limit the effect of learning between both conditions on the data.

The BART procedure

During the BART task, participants were first given a set of instructions, explaining the task and how to operate it (see Appendix B). These instructions also informed participants that the next five balloons would be for practice, and would not count towards their ability to win the prize. They proceeded to complete five practice balloons, after which they received an additional set of instructions (see Appendix B). The participants then completed three rounds of 10 balloons each (30 in total). The balloons all had predetermined breaking points that were kept the same for each participant, to avoid noise due to variance in the sequences of breaking points. The breaking points of the 30 'real' balloons were the same ones that were used by Lejuez et al. (2002). For the breaking points of the 5 practice balloons a list of five numbers with a mean of 64 was used (see Appendix C for a list of all breaking points). Participants would get 50 points per pump in their temporary bank, but were unable to see the number of points in their temporary bank. They were able to see the number of points in their permanent bank throughout the task. When a balloon exploded, participants would receive auditory feedback (a popping sound) in addition to the visual feedback of the exploding balloon. At the end of both BARTs*,* participants filled in the IPQ.

Specific procedure of the CB BART. For the CB BART, participants could remain seated behind the computer. The researcher would come up to the computer and start the BART in OpenSesame. Participants received instructions through the computer screen. Participants used the keyboard to control the experiment, using spacebar to pump the balloon, enter to collect points, and any key on the keyboard to go through the instructions when they were given. Between the 10th and 11th balloon, and between the 20th and 21st balloon, participants would receive a message on the computer screen that they had completed the round, and they could move onto the next round when they were ready.

Specific procedure of the VR BART. For the VR BART participants were first asked to fill in the VRSQ, and then they were instructed to sit in the VR area. The researcher then started the scene in unity. The researcher would help the participant put on the VR headset and controllers in a safe way. Participants received instructions through the screen in the VR environment. To pump participants needed to use their hands to push down the handle of the pump in VR. To collect points, and go through the instructions as they were given, participants needed to push down the red button. Between the 10th and 11th balloon, and the 20th and 21st balloon participants were asked to take of their headset and controllers and take a small break. During these breaks participants filled in the VRSQ again. If participants' motion sickness symptoms seemed to increase between rounds, they were asked if they still wanted to continue with the experiment. When they were done with all 30 balloons they filled in the VRSQ for a final time.

When the participants were done with both BARTs, they were thanked for their contribution and the experiment concluded. For changes in the procedure made between the original study and the current study see Appendix A.

Data analysis

Data analysis was performed using r. In order to be as close to the original study as possible when analyzing the data, a classical inferential approach was used. Adjusted number of pumps (ANP) on the BART (i.e. the mean pumps of any balloons that did not explode) was used as an indication of BART performance. Data from the five practice balloons was not used. First, the difference in BART performance between the current CB and VR BARTs was investigated. A paired t-test was performed to test whether the ANP was significantly different between the CB and VR BARTs. To investigate whether learning between rounds affected the VR BART differently than the CB BART, differences between individual rounds were also tested (e.g. difference between CB and VR BART for round 1). Then, the relationships between ANP on the CB and VR BART and participants' age, gender, scores for the different risk-related constructs (sensation seeking, impulsivity on the BIS-15, impulsivity on the I7, venturesomeness and behavioral constraint) and risk behaviors (alcohol consumption, drug use, gambling behavior, cigarette use, sexual behavior, stealing and seatbelt use) were investigated. Spearman rank correlations were calculated to find whether there were significant correlations between these variables. Pearson’s r was calculated to find wether there was a significant correlation between the CB and VR BART ANP. Finally, the difference between subjective immersion in the current CB and VR BARTs was investigated. A paired t-test was done to test whether subjective immersion was higher for the VR BART as compared to the CB BART.

Results

It was hypothesized (H1) that participants will have similar performance when using the CB BART or VR BART. ANP was used as a measure of performance on the BART for any data analysis. ANP was calculated by taking the average number of time the participant pumped for all balloons that did not explode.

A paired samples t-test was conducted to compare ANP on the CB and VR BARTs. Results show that ANP on the CB BART (m = 40.91, sd = 14.33) was significantly higher than ANP on the VR BART (m = 31.51, sd = 14.68); t(19) = 3.19, $p < .05$.

Furthermore, to investigate whether ANP differed for both versions per round of 10 balloons, paired samples t-tests were conducted to compare ANP on the CB and VR BARTs for all three rounds of balloons separately as shown in Table 1. ANP on the CB BART for round 1 (m = 37.57, sd = 14.27) was significantly higher than ANP on the VR BART for round 1 (m = 27.59, sd = 14.88); t(19) = 2.66. $p < .05$. ANP on the CB BART for round 2 (m = 43.11, sd = 17.18) was also significantly higher than ANP on the VR BART for round 2 (m = 32.85, sd = 16.95); t(19) = 2.65, $p < .05$. Lastly, ANP on the CB BART for round 3 (m = 43.36, sd = 15.17) was also significantly higher than ANP on the VR BART for round 3 (m = 35.08, sd = 14.54); t(19) = 3.42, $p < .05$. Figure 1 shows the overall differences between the conditions per round.

Table 1

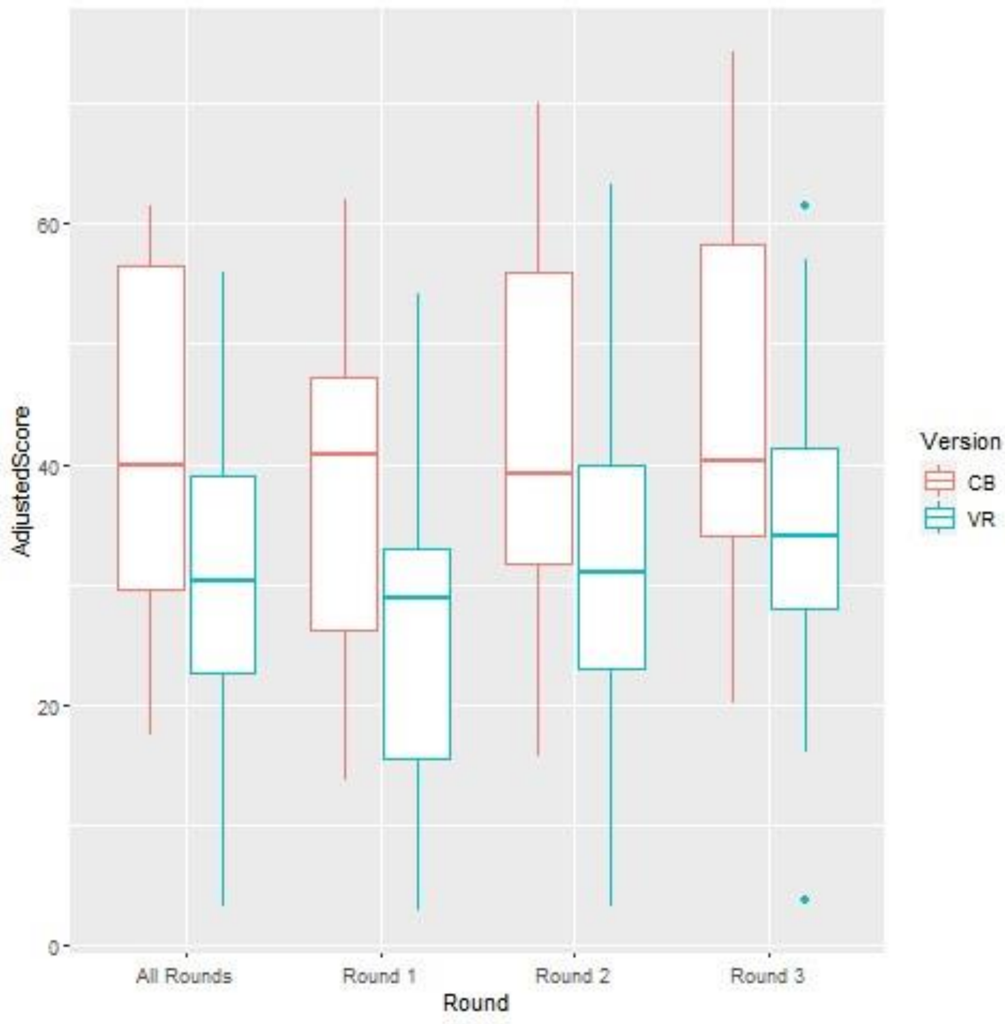
Differences in adjusted number of pumps (ANP) between the CB and VR conditions, for all 30 balloons, as well as for the first, second and third rounds of 10 balloons separately

	CB		VR				
	m	sd	m	sd	t	p	df
All balloons	40.91	14.33	31.51	14.68	3.19	.005*	19
Round 1	37.57	14.27	27.59	14.88	2.66	.02*	19
Round 2	43.11	17.18	32.85	16.95	2.65	.02*	19
Round 3	43.36	15.17	35.08	14.54	3.42	.003*	19

Note. p* < .05. p** < .001.

Figure 3

Boxplot comparing differences in ANP for both conditions, for all rounds, and every round separately.



Note: ‘AdjustedScore’ represents ANP, ‘All Rounds’ represents all 30 balloons, ‘Round 1’ represents balloons 1 -10, ‘Round 2’ represents balloons 11 – 20, ‘Round 3’ represents balloons 21-30.

Moreover, it was hypothesized (H2) that self-report questionnaires that were significanty correlated with ANP on the original BART (Lejuez, 2002) will also be correlated with the VR BART. Firstly internal consistencies for the current sample were calculated (see table 2). Then spearman rank correlations between ANP on both the CB and VR BARTs and scores on the self-report battery as well as age and gender were calculated (see table 3). For the CB BART, a significant positive correlation was found between ANP and score on the I7-V ($r(18) = .49$, $p = .03$). For the VR BART a significant positive correlation was found between ANP and age ($r(18) = .46$, $p = .04$). These results are different from Lejuez et al., who found significant correlations with all self-report measures for risk-related constructs that they used except for the I7-V, and all self-report measures for risk behaviors that they used. Lastly, Pearsons r between ANP on the CB BART and ANP on the VR BART was calculated. A significant positive correlation was found ($r(18) = .59$, $p = .006$).

Table 2

Internal consistencies of the questionnaires in the current sample

Questionnaire	Cronbach alpha
BSSS	.82
BIS-15	.89
I7-I	.88
I7-V	.83
BC (MPQ-BF)	.86
AUDIT	.87
GABS-23	.88
IPQ (CB-BART)	.44
IPQ (VR-BART)	.74

Table 3

Spearman rank correlations between BART performance and demographics, risk-related constructs, and risk behaviors.

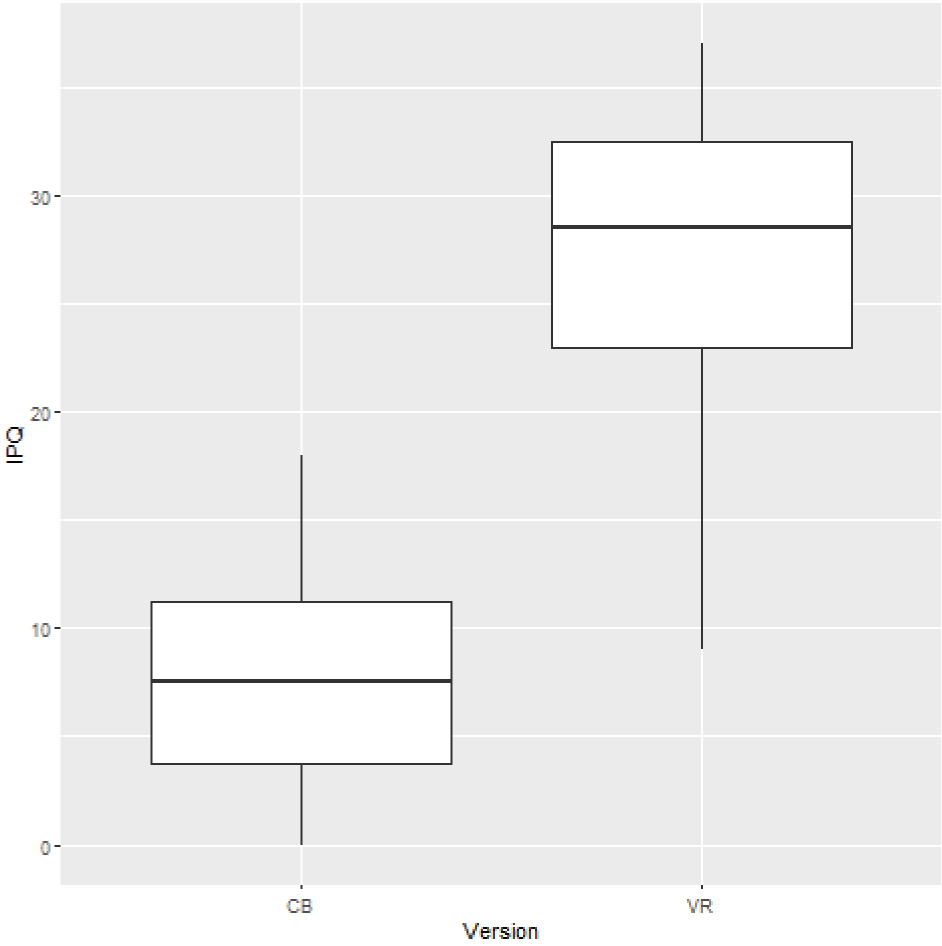
	Df	CB BART (current)	VR BART (current)	Original BART (Lejuez et al., 2002) (df = 84)
Age	18	.24	.46*	.07
Gender	18	-.17	-.18	.30*
BSSS	18	.30	.06	.35**
BIS-15	17	.19	.09	.28*
I7-I	18	.01	-.07	.20
I7-V	18	.49*	.19	.24*
BC (MPQ-BF)	17	-.34	-.20	.36**
AUDIT	18	.12	.06	.28*
Drugs	18	.22	-.22	.28*
GABS-23	18	.09	-.12	.44**
Cigarettes	18	-.13	-.13	.36**
Sex	18	.11	.24	.25*
Stealing	17	<.001	-.30	.25*
Seatbelt	16	.42	.31	.25*
VR BART (current)	18	.59*	-	-

Note: gender responses were coded as male: 1 and female: 2; degrees of freedom are shown for each different correlation (due to missing data not every correlation had the same degrees of freedom); Some of the questionnaires Lejuez et al. (2002) used were different from the ones shown in this table (see Appendix A); *p < .05. **p < .001.

Lastly, differences in subjective immersion between the VR and CB BARTs were investigated. A paired samples t-test was conducted between IPQ results for the CB and VR BARTs. Results show that IPQ score was significantly higher for the CB BART (m = 27.3, sd = 7.15) as compared to the VR BART (m = 8.1, sd = 5.5); t(19) = -12.278, p < .001 see figure 2).

Figure 4

Boxplot showing difference in IPQ score between the CB and VR BARTs



Discussion

The goals of this study were to validate a VR version of the BART, and investigate whether people feel more immersed during a VR BART as compared to a CB BART. Firstly, it was hypothesized that ANP on the CB BART would be similar to ANP on the VR BART. In contrast to the expected results, ANP on the CB BART was significantly higher than ANP on the VR BART. The results also show that ANP increases every round, as the participant is learning the game, but ANP increases at a similar rate for both conditions. Even when looking at each individual round of 10 balloons, ANP on the CB BART is always significantly higher than ANP on the VR BART. With the current results, there is not enough evidence to reject the null hypothesis for H1.

Although this result was not expected, it can be explained. As stated earlier in this report, risk is a personal calculation between the perceived potentials for rewards and dangers, and the size of those rewards and severity of those dangers (Vlek & Stallen, 1980). The fact that ANP is consistently lower on the VR BART indicates that participants were more careful when doing the task in VR. The size of the rewards, and potential for rewards were constant, and known to the participants. However, the balloon game was a black box for the participants, meaning they did not explicitly know the potential for danger (i.e. the balloon exploding), and had to guess this for every balloon. It therefore seems likely that a lower ANP during the VR BART is due to the participant perceiving the potential of the balloon exploding as higher. This seems to be further supported by the fact that several participants noted after the experiment that they felt more anxious about the balloon exploding during the VR BART, because the balloon was more realistic and bigger, and because it would expand in very close proximity to them.

Secondly, it was hypothesized (H2) that ANP on the VR BART would correlated with the same variables as ANP on the original BART during the study by Lejuez et al. (2002). The hypothesized similarity was not found. Out of all the tested variables, VR BART performance only significantly correlated with age (which was one of the few variables that Lejuez et al. (2002) tested which was not correlated with ANP). There is therefore also not enough evidence to reject the null hypothesis for H2.

ANP on the CB BART used in the current study was only correlated to the I7-V, and therefore did not correlate in to the same variables as the original BART either. It should be noted that other researchers have had mixed success with replicating the results found by Lejuez et al. (2002), with two meta-analyses finding that the high degree of convergent validity of the BART found in the original study was often not found (Duckworth & Kern, 2011; Lauriola et al., 2014). There was a correlation between ANP on the current CB BART and ANP on the VR BART, which shows that, even though there is a significant difference in ANP, both BARTs used in the current study measure something related. Even though the current CB and VR BARTs measure something related, both are not correlated to the same risk-related constructs and risk behaviors as the original BART. Consequently, the cause of the different results seems to be a difference between both studies, or a lack of convergent validity of the BART in general, rather than a difference between CB and VR .

Lastly, it was hypothesized (H3) that subjective immersion would be higher for the VR BART than for the CB BART. The results seem to support this, as they indeed show participants feeling a significantly higher level of immersion during the VR BART compared to the CB BART. Therefore, the null hypothesis for H3 is rejected.

Limitations

This study had several limitations. A first limitation is the fact that several changes had to be made between the BART used in this study and the one developed by Lejuez et al (2002). As indicated earlier in the discussion, this difference in study design could have contributed to the different correlations that were found in the current study. Firstly, participants were given only 30 balloons per condition (totaling 60 balloons), whereas in the original study they were given a total of 90 balloons, consisting of 3 types of balloons with different average breaking points. In the original study these balloons were all mixed, which would have made it harder to learn the patterns in the explosion points for the participants in the study.

Secondly, participants in the original study won money depending on their score. Because this was not financially viable, this was not possible in this study. Instead, participants could win a goodie bag if they were one of the top scorers. This might have affected the data, as people act more impulsive and take more risks when only a hypothetical reward is at stake instead of a real monetary reward (Xu et al., 2016).

Moreover, data was collected in an inconsistent way. While data collection was already happening, changes were made to the instructions of some of the questionnaires and both BARTs. Because of this, some participants received slightly different instructions that other participants. Furthermore, the order in which participants filled in the questionnaires and did the tasks was also changed, which caused a few participants to complete the BARTs before filling in the questionnaires.

The difference found between the CB and VR BART should have been further investigated with qualitative data. Some interesting insights were gained through talking with participants about their experience with both BARTs after the experiment, but this qualitative data was not collected in a systematic way, so it could not be used for the data analysis.

Furthermore, a low internal consistency was found for the IPQ when taken after the CB BART. This could indicate that the IPQ is not a valid measure for the CB BART. The IPQ is meant to be used for measuring sense of presence only in virtual environments (Igroup, n.d.), which might have caused this low internal consistency.

Finally, a relatively small sample size was used. Data from 20 participants was used in this study, which is low compared to the 86 in the original study. Earlier research using the BART has had mixed results with replicating the strong correlations found by Lejuez et al. (2002). Meta analyses on the relationship between BART score and risk-related constructs have found only weak correlations (Duckworth & Kern, 2011; Hopko et al., 2005; Lauriola et al., 2014). Having a large sample is therefore even more important, to be able to accurately gauge whether the potentially weak correlations are significant or not.

Conclusion

In this study, it is shown that changing the BART interface from computer to VR changes the results of the measure. People feel more immersed in a VR BART, and it seems that this causes them to act more carefully. Participants changed their behavior (pumped less), because the environment made the risk of the balloon exploding feel more real. This indicates that, when comparing a VR and CB BART, the behavior that a person shows in a VR BART more closely resembles the behavior they would show in a real situationl. These results support the idea that VR can improve RTP assessment, especially when the goal is to understand an individuals behavior. More generally, this study shows that researchers should be mindful of the fact that behavioral assessment tools should not be treated as unbiased, as they do not necessarily elicit natural behavior. Furthermore, when designing or improving behavioral assessment tools, media like VR should be used to increase the tools immersion as they will then evoke more natural behavior from the people they are used on. Although the results of this study show the merits of VR in RTP assessment, this study was unsuccessful in validating the VR BART. Further research should continue to expand on this, if possible by organizing a study with a larger sample size, as it is clear that a validated VR BART will be of great benefit to the assessment of RTP.

References

Althubaiti, A. (2016). Information bias in health research: definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare*, 211. https://doi.org/10.2147/jmdh.s104807

Bartz, W. R., & Loy, D. L. (1970). The Shipley-Hartford as a brief I.Q. screening device. *Journal of Clinical Psychology*, *26*(1).

BERKMAN, M. L., & ÇATAK, G. (2021). I-group Presence Questionnaire: Psychometrically Revised English Version. *Mugla Journal of Science and Technology*. https://doi.org/10.22531/muglajsci.882271

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and Health-Related Behaviors: A Meta-Analysis of the Leading Behavioral Contributors to Mortality. *Psychological Bulletin*, *130*(6), 887–919. https://doi.org/10.1037/0033-2909.130.6.887

Bohil, C., Alicea, B. & Biocca, F. Virtual reality in neuroscience research and therapy. *Nat Rev Neurosci* **12**, 752–762 (2011). https://doi.org/10.1038/nrn3122

Bouju, G., Hardouin, J. B., Boutin, C., Gorwood, P., Le Bourvellec, J. D., Feuillet, F., Venisse, J. L., & Grall-Bronnec, M. (2013). A Shorter and Multidimensional Version of the Gambling Attitudes and Beliefs Survey (GABS-23). *Journal of Gambling Studies*. https://doi.org/10.1007/s10899-012-9356-3

Breen, R., & Zuckerman, M. (1999). `Chasing' in gambling behavior: personality and cognitive determinants. *Personality and Individual Differences*, *27*(6), 1097–1111. https://doi.org/10.1016/s0191-8869(99)00052-5

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, *125*(3), 367–383. https://doi.org/10.1037/0033-2909.125.3.367

Chattha, U. A., Janjua, U. I., Anwar, F., Madni, T. M., Cheema, M. A., & Janjua, S. I. (2020). Motion Sickness in Virtual Reality: An Empirical Evaluation. *IEEE Access*, *8*, 130486–130499. https://doi.org/10.1109/access.2020.3007076

Chicchi Giglioli, I.A., Parra, E., Cardenas-Lopez, G., Riva, G., Alcañiz Raya, M. (2017). Virtual Stealth Assessment: A New Methodological Approach for Assessing Psychological Needs. In: Alcañiz, M., Göbel, S., Ma, M., Fradinho Oliveira, M., Baalsrud Hauge, J., Marsh, T. (eds) Serious Games. JCSG 2017. Lecture Notes in Computer Science(), vol 10622. Springer, Cham. https://doi.org/10.1007/978-3-319-70111-0_1

de-Juan-Ripoll, C., Llanes-Jurado, J., Giglioli, I. a. C., Marín-Morales, J., & Alcañiz, M. (2021). An Immersive Virtual Reality Game for Predicting Risk Taking through the Use of Implicit Measures. *Applied Sciences*, *11*(2), 825. https://doi.org/10.3390/app11020825

de-Juan-Ripoll, C., Soler-Dominguez, J. L., Guixeres, J., Contero, M., Gutiérrez, N. R., & Alcañiz, M. (2018). Virtual Reality as a New Approach for Risk Taking Assessment. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02532

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*(3), 259–268. https://doi.org/10.1016/j.jrp.2011.02.004

Eysenck, S. B., Pearson, P., Easting, G., & Allsopp, J. (1985). Age norms for impulsiveness, venturesomeness and empathy in adults. *Personality and Individual Differences*, *6*(5), 613–619. https://doi.org/10.1016/0191-8869(85)90011-x

GULLONE, E., & MOORE, S. (2000). Adolescent risk-taking and the five-factor model of personality. *Journal of Adolescence*, *23*(4), 393–407. https://doi.org/10.1006/jado.2000.0327

Haas, J. K. (2014). *A History of the Unity Game Engine*. WORCESTER POLYTECHNIC INSTITUTE.

Halbig, A., & Latoschik, M. E. (2021). A Systematic Review of Physiological Measurements, Factors, Methods, and Applications in Virtual Reality. *Frontiers in Virtual Reality*, *2*. https://doi.org/10.3389/frvir.2021.694567

Hopko, D. R., Lejuez, C. W., Daughters, S. B., Aklin, W. M., Osborne, A., Simmons, B. L., & Strong, D. R. (2006). Construct Validity of the Balloon Analogue Risk Task (BART): Relationship with MDMA Use by Inner-City Drug Users in Residential Treatment. *Journal of Psychopathology and Behavioral Assessment*, *28*(2), 95–101. https://doi.org/10.1007/s10862-006-7487-5

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., & Donohew, R. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences*, *32*(3), 401–414. https://doi.org/10.1016/s0191-8869(01)00032-0

Igroup (n.d.). *igroup presence questionnaire (IPQ) overview*. Retrieved June 29, 2023, from http://www.igroup.org/pq/ipq/index.php

Josef, A. K., Richter, D., Samanez-Larkin, G. R., Wagner, G. G., Hertwig, R., & Mata, R. (2016). Stability and change in risk-taking propensity across the adult life span. *Journal of Personality and Social Psychology*, *111*(3), 430–450. https://doi.org/10.1037/pspp0000090

Kim, H. J., Park, J., Choi, Y., & Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, *69*, 66–73. https://doi.org/10.1016/j.apergo.2017.12.016

Lauriola, M., Panno, A., Levin, I. P., & Lejuez, C. W. (2014). Individual Differences in Risky Decision Making: A Meta-analysis of Sensation Seeking and Impulsivity with the Balloon Analogue Risk Task. *Journal of Behavioral Decision Making*, *27*(1), 20–36. https://doi.org/10.1002/bdm.1784

Leigh, B. C. (1999). Peril, chance, adventure: concepts of risk, alcohol use and risky behavior in young adults. *Addiction*, *94*(3), 371–383. https://doi.org/10.1046/j.1360-0443.1999.9433717.x

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue

Risk Task (BART). *Journal of Experimental Psychology: Applied*, *8*(2), 75–84. https://doi.org/10.1037/1076-898x.8.2.75

Lesieur HR, Blume SB (1987). The South Oaks Gambling Screen (SOGS): a new instrument for the identification of pathological gamblers. *Am J Psychiatry*. 1987 Sep;144(9):1184-8. doi: 10.1176/ajp.144.9.1184. PMID: 3631315.

Mathôt, S., Schreij, D., & Theeuwes, J. (2011). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment*, *14*(2), 150–163. https://doi.org/10.1037/1040-3590.14.2.150

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6

Qualtrics XM: The Leading Experience Management Software. (2023, February 7). Qualtrics. https://www.qualtrics.com/

Radloff, L. S. (1977). The CES-D Scale. *Applied Psychological Measurement*, *1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Reinert, D. F., & Allen, J. F. (2007). The Alcohol Use Disorders Identification Test: An Update of Research Findings. *Alcoholism: Clinical and Experimental Research*, *31*(2), 185–199. https://doi.org/10.1111/j.1530-0277.2006.00295.x

Reiss, S. P., Peterson, R. O., Gursky, D., & McNally, R. J. (1986). Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. *Behaviour Research and Therapy*, *24*(1), 1–8. https://doi.org/10.1016/0005-7967(86)90143-9

Roberts, D. K., Alderson, R. M., Betancourt, J. L., & Bullard, C. C. (2021). Attention-deficit/hyperactivity disorder and risk-taking: A three-level meta-analytic review of behavioral, self-report, and virtual reality metrics. *Clinical Psychology Review*, *87*, 102039. https://doi.org/10.1016/j.cpr.2021.102039

Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*, *88*(6), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Shipley, W. C. (1940). A Self-Administering Scale for Measuring Intellectual Impairment and Deterioration. *The Journal of Psychology*, *9*(2), 371–377. https://doi.org/10.1080/00223980.1940.9917704

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slovic, P. (1964). Assessment of risk taking behavior. *Psychological Bulletin*, *61*(3), 220–233. https://doi.org/10.1037/h0043608

SPINELLA, M. (2007). NORMATIVE DATA AND A SHORT FORM OF THE BARRATT IMPULSIVENESS SCALE. *International Journal of Neuroscience*, *117*(3), 359–368. https://doi.org/10.1080/00207450600588881

Steiner, M. J., & Frey, R. (2021). Representative design in psychological assessment: A case study using the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: General*, *150*(10), 2117–2136. https://doi.org/10.1037/xge0001036

The South Oaks Gambling Screen (SOGS): a new instrument for the identification of pathological gamblers. (1987). *American Journal of Psychiatry*, *144*(9), 1184–1188. https://doi.org/10.1176/ajp.144.9.1184

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment, Vol. 2.*

Personality measurement and testing (pp. 261–292). Sage Publications, Inc. https://doi.org/10.4135/9781849200479.n13

Trimpop, R. (1994). The Psychology of Risk Taking Behavior. *Advances in Psychology*. https://doi.org/10.1016/s0166-4115(08)x6053-4

Vlek, C., & Stallen, P. J. (1980). Rational and personal aspects of risk. *Acta Psychologica*, *45*(1–3), 273–300. https://doi.org/10.1016/0001-6918(80)90038-4

Xu, S., Pan, Y., Wang, Y., Spaeth, A. M., Qu, Z., & Rao, H. (2016). Real and hypothetical monetary rewards modulate risk taking in the brain. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep29520

Zuckerman, M., Eysenck, S. B., & Eysenck, H. J. (1978). Sensation seeking in England and America: Cross-cultural, age, and sex comparisons. *Journal of Consulting and Clinical Psychology*, *46*(1), 139–149. https://doi.org/10.1037/0022-006x.46.1.139

Appendix A

Changes made between the original and current studies

	Lejuez et al. (2002)	Current study	Explanation for the change
(i)	Sensation seeking was measured using the Sensation Seeking Scale (Zuckermann et al., 1978)	Sensation seeking was measured using the Brief Sensation Seeking Scale (Hoyle et al. 2002)	A shorter version was used to safe time on the assessment
(ii)	Impulsivity was measured using the Baratt Impulsiveness Scale – version 1.0 (Patton et al., 1995)	Impulsivity was measured using the Baratt Impulsiveness Scale short form (Spinella, 2007)	A shorter version was used to safe time on the assessment
(iii)	The behavioral constraint superfactor of the Multidimensional Personality Questionnaire (Tellegen & Waller, 2008) was used to measure a combination of risk-related constructs	The behavioral constraint superfactor of the Multidimensional Personality Questionnaire brief form (Patrick et al., 2002) was used to measure a combination of risk-related constructs	A shorter version was used to safe time on the assessment
(iv)	Gambling was measured using the South Oaks	Gambling was measured using only Gambling	A shorter version of the Gambling Attitudes and

	Gambling Screen (Lesieur & Blume, 1987) and Gambling Attitudes and Beliefs Scale (Breen & Zuckermann, 1999)	Attitudes and Beliefs Scale 23-items version (Bouju et al., 2013)	Beliefs Scale was used to safe time on the assessment. The South Oaks Gambling Screen was left out, because it was not used in the results of Lejuez et al. (2002).
(v)	Use of the following drug classes was assessed: (i) marijuana, (ii) stimulants, (iii) cocaine, (iv) hallucinogens, (v) opiates, (vi) sedatives, (vii) other	Use of the following drug classes was assessed: (i) cocaine, (ii) heroin, (iii) marijuana, (iv) amphetamines, (v) inhalants, (vi) tranquilizers, (vii) hallucinogens, (viii) ecstasy, (ix) opioid analgesics, (x) synthetic drugs, (xi) anabolic androgenic steroids, (xii) other	The number of drug classes assessed was changed to be more representative of the current drug use as compared to when the original study took place
(vi)	IQ was measured using the vocabulary section of the Shipley Institute	IQ was not measured	The measure of IQ was removed because no correlation was found between the BART and

	of Living Scale (Shipley, 1940)		IQ by Lejuez et al. (2002)
(vii)	Discriminant validity was measured using the empathy subscale of the Eysenck Impulsiveness Scale (Eysenck et al. 1985), Center of Epidemiological Studies Depression Scale (Radloff, 1977), and Anxiety Sensitivity Index (Reiss et al., 1986).	Discriminant validity was not measured	It was unnecessary to measure Discriminant validity of the BART in this study as it was already determined to be good by Lejuez et al. (2002).

Appendix B

List of explosion points used during the experiment

Balloon nr.	Practice round	1st round	2nd round	3rd round
1	64	64	9	93
2	6	96	64	45
3	106	32	104	64
4	60	17	37	22
5	84	85	77	85
6	-	64	84	76
7	-	105	94	110
8	-	44	64	53
9	-	55	49	28
10	-	78	58	64

Appendix C

Instructions given during the BART

1st instructions CB BART

"Welcome to the Balloon Game! In this game, you will be presented with 30 balloons, one at a time. For each balloon you can use the space bar to increase the size of the balloon. For each pump, you will gain 50 points in a temporary bank. It is your choice to determine how much to pump up the balloon, but at some point the balloon will explode. The explosion point varies across balloons, ranging from the first pump to enough pumps to make the balloon very big. At any point, you can stop pumping up the balloon and press the Enter key to collect your points. The points will be saved in your permanent bank, labelled Total Earned. The amount you earned on the previous balloon is labelled Last Balloon. If the balloon explodes before you press the Enter key then you move on to the next balloon and all money in your temporary bank is lost. Exploded balloons do not affect the points stored in your permanent bank. How may points you win matters because the two participants with the highest amounts of points will each win a goodie bag! Press any key to continue. The next 5 balloons are practice trials to get the hang of it. The points you win or loose during the practice don't count. If you have any questions, please make sure to ask them during the practice. Press any key to start practicing."

2nd instructions CB BART

"The next 30 balloons will count towards your total points and your chance to win a prize. If you have any questions, please ask the researcher now. Press any key to start playing and winning points!"

1st instructions VR BART

"Welcome to the Balloon Game! In this game, you will be presented with 30 balloons, one at a time. For each balloon you can use the pump to increase the size of the balloon. For each pump, you will gain 50 points in a temporary bank. It is your choice to determine how much to pump up the balloon, but at some point the balloon will explode. The explosion point varies across balloons, ranging from the first pump to enough pumps to make the balloon very big. Press the red button to continue with the instructions. At any point, you can stop pumping up the balloon and press the red button to collect your points. Pressing the button a second time will start the next balloon. The points will be saved in your permanent bank, labelled "Total Earned." The amount you earned on the previous balloon is labelled "Last Balloon." If the balloon explodes before you press the red button then you move on to the next balloon and all money in your temporary bank is lost. Exploded balloons do not affect the points stored in your permanent bank. How many points you win matters because the 2 participants with the highest amounts of points will each win a prize! The next 5 balloons are practice trials to get the hang of it. The points you win or loose during the practice don't count. If you have any questions about the game, please make sure to ask them during the practice. Press the red button to continue. The next 5 balloons are practice trials to get the hang of it. The points you win or loose during the practice don't count. If you have any questions about the game, please make sure to ask them during the practice. Press the red button to continue."

2nd instructions VR BART

"The next 30 balloons will count towards your total points and your chance to win a prize. After each round of 10 balloons, the researcher will ask you to take a break and fill in some questionnaires. If you have any questions, please ask the researcher now. Press the red button to start playing and winning points!"