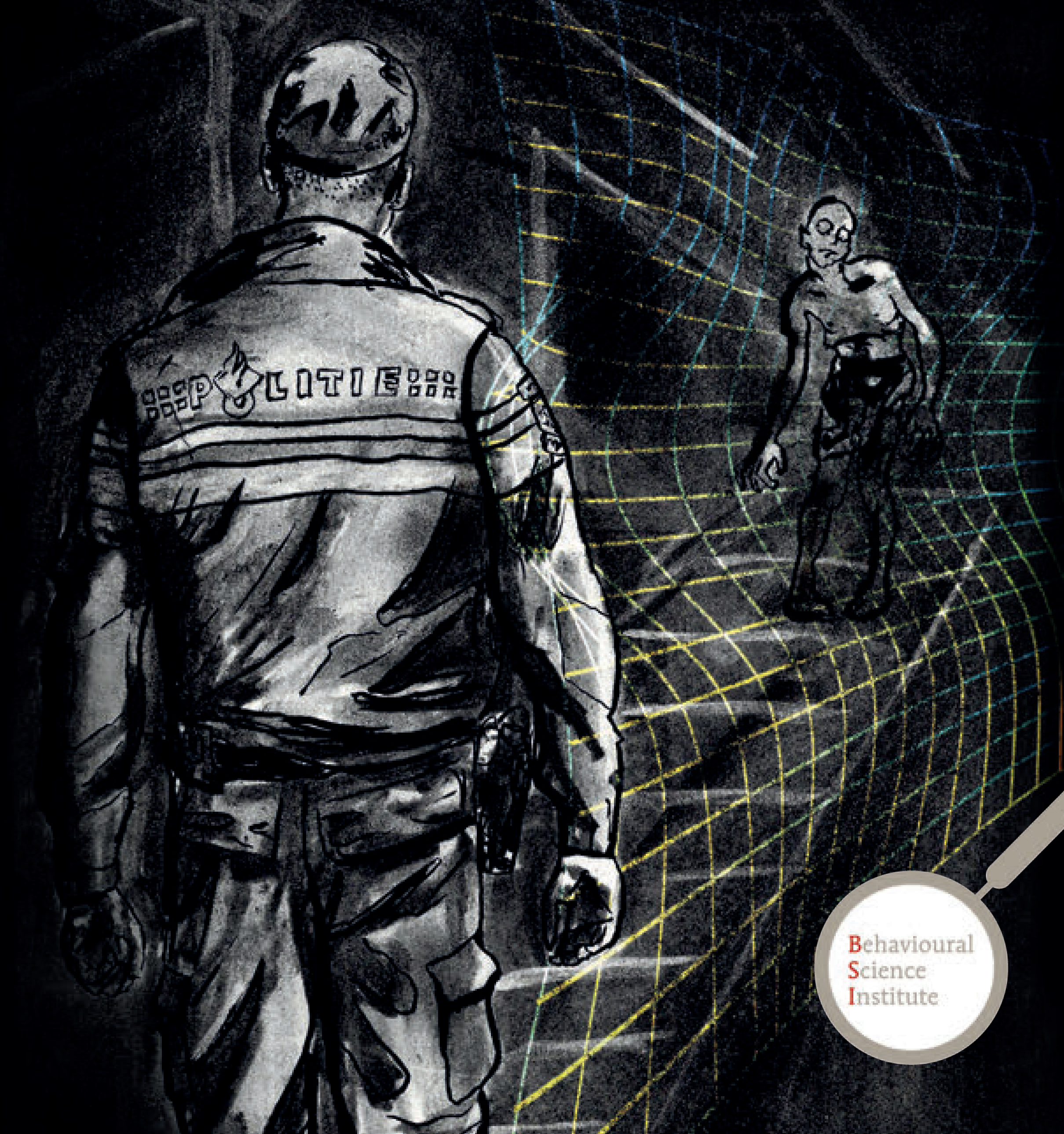
THE HEART ON DUTY

Training police officers in action using a biofeedback virtual-reality game

Abele Michela



Behavioural Science Institute

THE HEART ON DUTY

Training police officers in action using a biofeedback virtual-reality game

Abele Michela

Funding:

This work was supported by the Netherlands Organization of Scientific Research (NOW) Creative Industry grant # 055.16.139 2795 awarded to KR, IG, and FK also supporting AM. Additional funds were provided by CbusineZ after approval by the PolitieZorgPolis (PZP).

A 6 months extension of AM's PhD contract was financed by the Behavioural Science Institute of the Radboud University Nijmegen to compensate for COVID-19 related delays.

Cover design: Abele Michela & Nausicaa Planche
Lay-out & Print: Ridderprint | www.ridderprint.nl

THE HEART ON DUTY

Training police officers in action using a biofeedback virtual-reality game

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 22 januari 2024
om 16.30 uur precies

door

Abele Michela
geboren op 24 mei 1988
te Sierre (Zwitserland)

Promotor(en):

Prof. dr. K. Roelofs
Prof. dr. Isabela Granic (McMaster University Hamilton, Ontario, Canada)

Copromotor(en):

Dr. Floris Klumpers
Dr. Jacobien van Peer

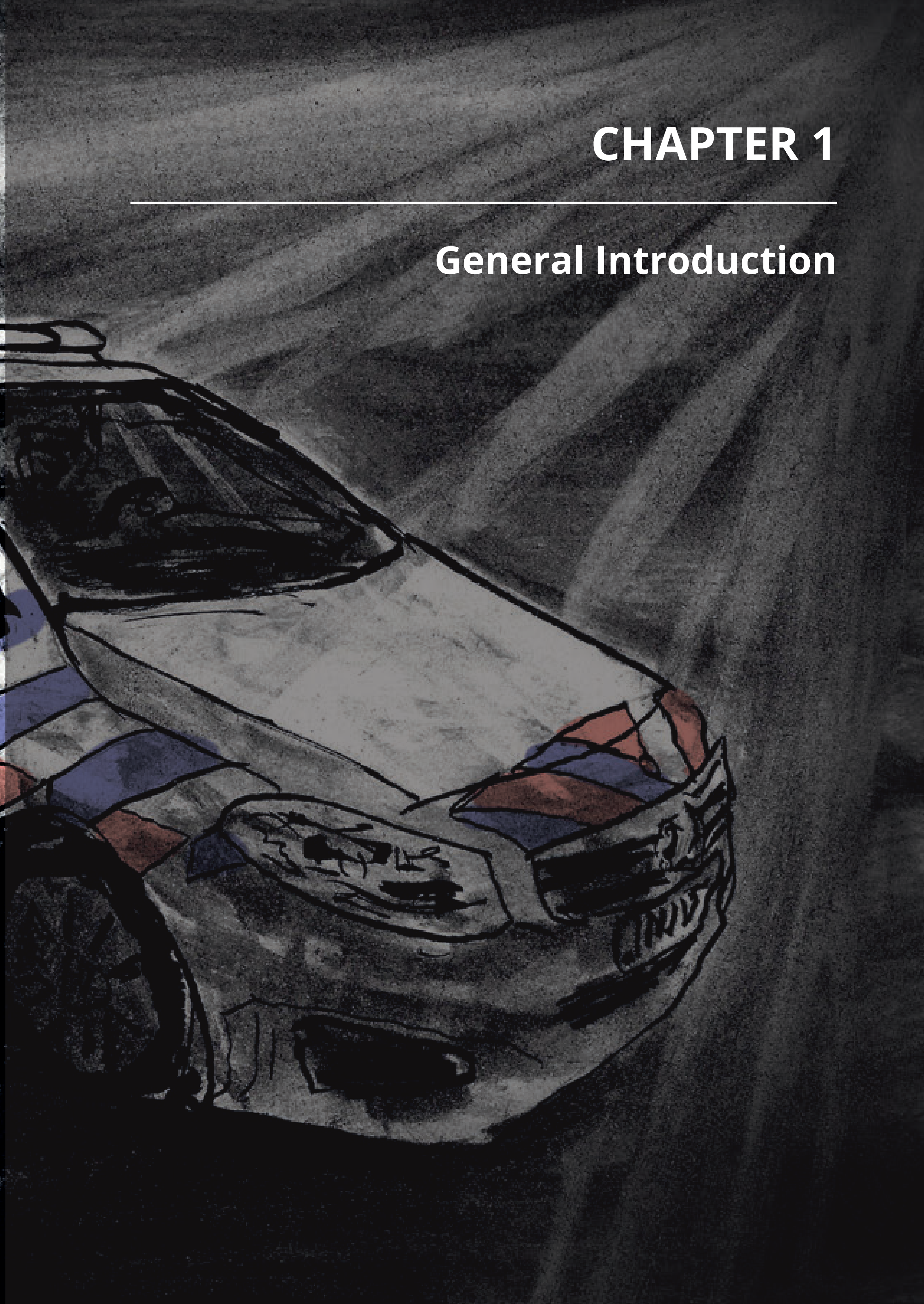
Manuscriptcommissie:

Prof. dr. E. S. Becker
Prof. dr. R.D.D. Oudejans (Vrije Universiteit Amsterdam)
Prof. dr. D. Sander (University of Geneva, Zwitserland)

TABLE OF CONTENTS

CHAPTER 1

General Introduction



GENERAL INTRODUCTION

The present thesis describes the design and validation of a game-based biofeedback training intervention in Virtual Reality (VR), aimed at helping police officers to improve psychophysiological control during active decision making when stressors are present. Acute stressors lead to a plethora of physiological changes in the brain and body, including increased heart rate and decreased heart rate variability (HRV), and can have detrimental effects on acute action decisions, such as shooting decisions in police officers (Hashemi et al., 2019; Nieuwenhuys & Oudejans, 2011). Although biofeedback and breathing trainings have been successful in increasing HRV and although it is well known that HRV is positively linked to better cognitive performance and emotion regulation (De Witte et al., 2019; Lehrer et al., 2020; B. Yu et al., 2018), those trainings typically take place at rest (while sitting behind a computer). They bear the disadvantage that the skills do not transfer well to naturalistic situations that demand acute action under challenging and stressful situations (Parnandi & Gutierrez-Osuna, 2017, 2019). In order to increase the chance of skill transfer we developed a game requiring police officers to control their psychophysiological state in a stressful and active action context. Below I will first explain the background for the need for this new intervention for police officers who often have to perform in stressful contexts. Next, I will detail the psychophysiological effects of stress and how those effects can be influenced through training, specifically by HRV training through biofeedback. Then I will move on to explain the potential of VR to develop an optimal training context. I will describe several stages of the intervention development and the lessons learned. The introduction section concludes with a description of the thesis layout.

Police performance in stressful contexts

When we feel threatened, stressed or scared, our body reacts with evolutionary engrained physiological response patters: The brain shifts into a state fostering immediate risk assessment (Blanchard et al., 2011; Zhang et al., 2022) and rapid defense mechanisms (Hermans et al., 2011), which promote attentional narrowing toward the threat (Rued et al., 2019) and habitual reflexes (Wirz et al., 2018). While highly beneficial for life-threatening situations that require immediate fight or flight, in stressful circumstances we are left with higher-order cognitive processes (such as executive control) impaired, our attention biased and our behavior somewhat inflexible (Lojowska et al., 2015; Maier et al., 2015; Wirz et al., 2018). Those are not the best circumstances for decision-making, since our body's only concern is to escape the threat (Hagenaars et al., 2014; Harrison et al., 2015), thus biasing us toward actions of self-preservation (Cornwell et al., 2017; Davies, 2017; Marenin, 2016). For most of us living in today's western Europe, far from war theatres, our stress reactions are triggered by a vast variety of events, but the majority of those events are not *actually* life-threatening. Yet, those sometimes-useful stress reactions may have an impact, for instance when they make one more impulsive or more prone to choose immediate rewards instead of pursuing long term goals (Maier et al., 2015). Indeed, deficient stress regulation has enormous consequences for people's lives, including their professional performance, their physical health, and their mental health (Gradus, 2017). Despite the fact that suboptimal stress management may not be directly life threatening for most people in modern societies, there are groups for which it can be life threatening, to themselves and others. This especially holds also for first responders, such as police officers, firefighters and paramedics.

This thesis takes the specific case of police officers who, as other first responders, are facing a paradox: Their line of work puts them into situations eliciting strong stress responses, yet they are asked to act with extreme control, calm and benevolence (Graef, 1990). Failing to behave with impeccable proportionality (i.e., using the correct amount of force and coercion required by the situation, but not more) can lead to excessive use of force, mistrust from civilians (Miethe et al., 2019), tarnished image of the profession (Worrall, 1999), and ultimately the death of innocents or of the officers themselves (Marenin, 2016). Excessive use of force is often reported when the officer is attempting to de-escalate a tensed situation (Ellrich et al., 2011). Specifically, self-preservation stress reactions have been suggested to be linked to excessive use of force (Baldwin et al., 2022; Nieuwenhuys et al., 2015). Thus, stress induction can negatively impact acute police performance, but it can also have longer term consequences for police officers themselves. This is reflected in the high prevalence of physical injuries (West et al., 2017), poor sleep quality (Fekedulegn et al., 2016) and stress-related mental health problems (Carleton et al., 2018; Leppma et al., 2017). Those negative conditions are worsened by the fact that police officers often have to deal with both pressure from the street and from the organization they belong to (Davenport, 1999; Graef, 1990; Van Der Velden et al., 2010). In light of all these challenges and the potential detrimental effects of stress responses on police performance and wellbeing, there is increasing interest in the following question: How do we train police officers to perform efficiently under stress?

Due to the specific conditions in which stress regulation has to happen in police work, training efforts have been undertaken by police academies to reduce the performance drop during stress induction. Indeed, it has been shown that when police officers were performing shooting exercises, the performance dropped when a physical threat was introduced (e.g., electrical shocks and plastic bullets; Hashemi et al., 2019; Nieuwenhuys & Oudejans, 2011). It has additionally been shown that shooting training under threat mitigated these negative effects of threat on shooting performance, with training benefits potentially lasting several months (Nieuwenhuys & Oudejans, 2011; Oudejans, 2008; Oudejans & Pijpers, 2009). However, these positive training results mostly applied to the precision of shots fired and did not hold for decisions to shoot or not. For instance, when making shooting decisions in response to armed opponents to be shot at, or unarmed citizens to be spared, the presence of acute threat led to a significant drop in police officers' decision-making performance, reflected in more unarmed citizens being shot (Nieuwenhuys et al., 2015). Importantly, this drop in decision-making performance induced by the threat could not be overcome by training under threat.

In light of the findings that stressful situations and exposure to threat elicit bodily responses that can bias our decision-making towards impulsive reactions (Baldwin et al., 2019; Hashemi et al., 2019; Morgado et al., 2015; R. Yu, 2016), we aimed to increase control over threat-induced psychophysiological reactions under threat. Before detailing the exact research questions, the following section provides an overview of important psychophysiological reactions under threat.

Psychophysiological effects of stressors and how to influence them

Threat response and switch to action

As described by Roelofs & Dayan (2022), when facing a potential threat, immediate defensive reactions are supported by rapid reactions of the autonomic nervous system. The sympathetic system—with noradrenaline as primary neurotransmitter—promotes tachycardia (heart-rate acceleration), hyperventilation, increased muscle tone, blood pressure, sweating and pupil dilation, all reactions geared to generate rapid *fight-or-flight* reactions and to activate the cortisol responses. However, when threat is still at distance or it is not entirely clear yet what action to take, parallel activation of the cholinergically-driven parasympathetic system puts a break on the cardiac and motor system and causes us to *freeze*. Freezing is characterized by upregulation of both sympathetic and parasympathetic arousal, but dominance in the parasympathetic system results in a net bradycardia (heart-rate deceleration, hypoventilation and motor inhibition that is so typical for freezing; Hagenaars et al., 2014; Roelofs, 2017; Roelofs & Dayan, 2022). Important for the present thesis, parasympathetic dominance not only results in heart rate (HR) deceleration and hypoventilation, it also results in an increased heart rate variability (HRV). This occurs potentially via vagally mediated mechanisms, including respiratory sinus arrythmia, which results in the acceleration of the HR during inhalation, and deceleration during exhalation (Hirsch & Bishop, 1981; Shaffer et al., 2014; Yasuma & Hayano, 2004).

Interestingly, the increased parasympathetic activity seen during the state of freezing is fundamental to preparing the body for optimal action (Gladwin et al., 2016). The reduced heart rate is linked to gathering information for improved decision-making (Klaassen et al., 2021), among others by enhancing perception (de Voogd et al., 2022; Lojowska et al., 2015; Rösler & Gamer, 2019). Once we have made the decision, we see parasympathetic withdrawal and the already alerted sympathetic branch of the autonomic nervous system enables fast acting, notably with temporary increases in strength (Jerath et al., 2006). These defensive cascades *can* happen largely automatically, and adequate tuning may be essential for optimal coping with threat (Fragkaki et al., 2017; Niermann et al., 2019; Roelofs et al., 2023).

The increased sympathetic activity in the active fight-or-flight-state can be of paramount importance for survival when fast action has to be generated. However, when it is associated with reduced parasympathetic activity, a decrease in HRV can take place, which has been linked to impaired decision-making capacities (Forte et al., 2022; Ramírez et al., 2015). Indeed, high HRV is suggested to be a proxy of ventromedial prefrontal cortex mediated activities that guide flexible control over behavior (Thayer et al., 2012).

Restoring balance in the autonomic nervous system

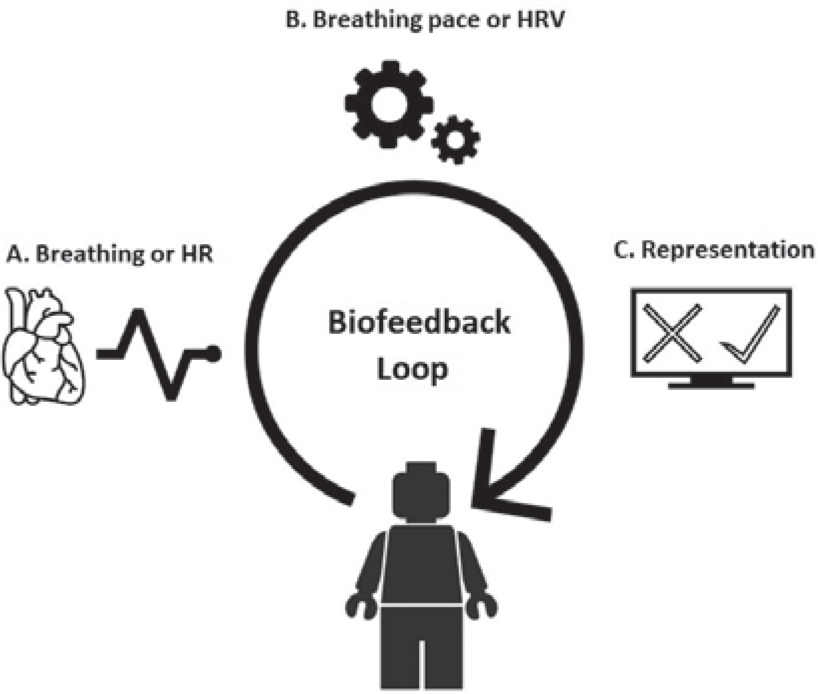
In order to bring back more parasympathetic influence in stressful encounters, breathing control is perhaps the most promising direction as it is less invasive than medication and direct vagus nerve stimulation (Ben-Menachem et al., 2015). Indeed, deep abdominal breathing stimulates vagal activity which results in increased HRV through respiratory sinus arrythmia (Yasuma & Hayano, 2004). This increase of breathing-related HRV can be measured by increases in HR during inhales, and slowing-down of the HR during exhales. After stress exposure, HRV has been consistently reported to decrease as a response to stressors (Johnsen et al., 2012; Pulopulos et al., 2018, 2020). Given that HRV is intimately tied to the psychophysiological sequelae of stressors and the balance of the sympathetic and parasympathetic nervous systems, increasing HRV has been suggested as an effective training target to mitigate stress reactions. Indeed, a large variety of studies suggest that improving HRV allows to increase cognitive control (Laborde et al., 2021), performance and decision-making under threat (Forte et al., 2021; Hansen et al., 2009), and more generally stress management (De Witte et al., 2019).

Biofeedback for HRV regulation

Several techniques can be employed to help modulate HRV, including breathing exercises, meditation and biofeedback. We selected biofeedback (BF), as it is well compatible with VR applications, allows to extract objective training data for *in-action* training and performs as well as breathing-based meditation techniques even in standard BF setups (van der Zwan et al., 2015). Additionally, it has been shown to be well applicable across a large variety of populations and disciplines (Goessl et al., 2017; Jiménez Morgan & Molina Mora, 2017; Lehrer et al., 2020; Lehrer & Gevirtz, 2014; Pizzoli et al., 2021). With this method, the physiological information—specifically HR—of the participants is analyzed in near real-time to provide them with information about their HRV (see figure 1.1). Thanks to breathing techniques (mainly deep abdominal breathing), the participants can then train to modulate HRV while being provided live information about their success rates. This type of closed loop BF training usually takes place by sitting still in front of a computer, while the HR data is recorded using a plethysmograph or other types of HR monitoring systems. Such HRV trainings have already been developed and implemented in the police curricula of the Netherlands, with limited benefits and mixed reception. The main critique on the current passive HRV training is that the breathing techniques taught in a calm classroom setup do not transfer to a stressful action context (van der Meulen et al., 2018; P. van der Velden et al., 2014). Finally, the current implementation has been reported to lack engagement elicitation. This negative appraisal is in line with findings suggesting the superiority of contextualized BF, especially in terms of skill transfer (Parnandi & Gutierrez-Osuna, 2017, 2019).

To summarize, the aim of the current thesis was to investigate how HRV could be trained in ways that would promote transfer of the learned HRV control skill to real-life stressful situations. To that end, we developed a VR game that included real-time biofeedback in a closed loop system (see Figure 1.1), that provided an active action context that elicited psychophysiological stress and was engaging. We started by **validating potential VR training environments** in terms of arousal elicitation and capacity to measure meaningful behaviors while training. We further tested the **causal role of biofeedback** presentation on psychophysiological control, and the **transfer** of such control skills in a police-relevant decision-making task. Additionally, we monitored the **appraisal** of the training by the participants to provide a more encompassing account of the mental processes involved in the training. The experiments described in this thesis were conducted in a population of police trainers rather than police students, as they could more expertly provide feedback on the training procedure and assess its benefits. Hence, we tested a total of 118 police trainers in two different studies, in a VR environment that we designed specifically to train HRV through BF, in a stress-inducing action context with decision-making. The choice for VR was based on its immersive capacity and the enhanced engagement this elicits in the player. The rest of the current chapter details how the intervention was designed.

Figure 1.1
The biofeedback loop.



Note. During a biofeedback (BF) training the user is put in a closed loop. The loop can be divided in three components. First, a BF parameter (A) is extracted from a physiological modality. In the study presented in chapter 4 of this thesis, the parameter is breathing (lower thoracic expansion) while in chapter 5 the parameter is heart rate. Second, the signal is processed (B) to evaluate if the current state of breathing pace and depth, or Heart-Rate-Variability (HRV) matches a predefined target state. This calculation results in a BF score that gets translated into a representation (C) consisting, in our case, of a manipulation of the players' visual field in VR. *Adapted from* Brammer et al. (2021).

Intervention development

For the final design, we considered a "zombie-shooter" game dynamic as a suitable candidate. In short, the player was immersed in a VR environment and asked to shoot approaching zombie targets matching a radio description. Other zombies had to be left reaching the player unharmed. The game additionally reacted to the player's physiology by narrowing their field of view when the measured physiological state of the player was not indicative of relaxation. This game format was selected since similar gaming formats involving zombies coming at the player from all sides have proven to be a reliable way to induce stress and a feeling of threat (Lin, 2017; Reichenberger et al., 2017). Additionally, non-living targets alleviated ethical considerations around the risk of de-sensitization of police officers through repeated shoot/don't shoot decisions in a realistic police training (Hinte, 1971; Williams & Clarke, 2019). Pragmatically, the rationale for this choice can be summarized by the idea that creating a *real* feeling of threat is more achievable by an *unrealistic* yet "believable" context than by a virtual simulation of reality.

Engagement for an effective intervention

The lack of engagement reported in standard BF setups further highlights the often-overlooked problem of defining the processes leading to behavioral improvements (Kazdin, 2011). For a training to be successful, the trainee must engage with it and understand its value. Poorly designed interventions can be unsuccessful as they don't engage the trainee enough. This fact has been seen with the wearable sensors (used for BF) that are often dropped by the users shortly after purchase (Ledger & McCaffrey, 2014), and lack of engagement has also been found in game-based interventions (Scholten & Granic, 2019). This lack of engagement has been theorized to hinder behavioral change, specifically in BF learning processes (Weerdmeester et al., 2020). Further, providing an engaging and relatable context for BF training could help the trainee to better understand how to apply the learned physiological regulation skill in real life. It has been shown that when physiological regulation is practiced in a relevant context, the attention given to arousal and its reappraisal by the participant is higher than when the skill is practiced in a neutral or relaxing setting (Driskell & Johnston, 1998). Training police officers to control their psychophysiological arousal in stressful encounters therefore requires an engaging and relatable context for the skill learned during training to transfer to real life.

A game for training police officers

In terms of design goals, the VR game had to: **(1) create an engaging experience, (2) elicit psychophysiological stress, (3) provide an active decision-making context that feels relevant for police officers and (4) have an imbedded mechanic to provide feedback on the participant HRV.** As explained in the previous paragraph, the choice of a VR game-based intervention solved naturally the first two requirements of the game, namely the engagement problem (Allcoat & von Mühlenen, 2018) and the stress induction reliability problem (Lin, 2017; Riva et al., 2019). Additionally, using a game-based environment facilitates the third requirement as a game allows the introduction of scientific paradigms in the game mechanic to provide relevant decision-making opportunities in game (e.g., shooting zombies can become a Go/No-go task). The same holds for the last requirement, as a game provides easy opportunities for BF incorporation, for example by linking the field of vision of the player to their psychophysiological state.

To drive the development of the VR game, a guiding model was developed (see Table 1.1). This model details the principle through which each psychological mechanism could be investigated in the VR environment, and is recommended for game-based interventions (Scholten & Granic, 2019). In the model, each psychological mechanism is mapped to an evidence-based technique commonly used to investigate such psychological mechanism. Those evidence-based techniques are translated into a game mechanic, which in turn allows to measure a proximal outcome of the player performance that directly maps to the original psychological mechanism investigated. Thus, the model ensured that each measurement performed on a player while immersed in the VR game related to relevant psychological mechanisms that we aimed to train or influence.

Table 1.1

Guiding model for the incorporation of evidence-based mechanisms and techniques in the VR game environment

Psychological mechanism	Evidence-based technique	Game mechanic	In-game outcome	Envisioned long term goal
Psycho-physiological regulation	Bio-feedback	Width of the field of view linked to HRV score	Breathing/ HRV biofeedback score increase	Improved stress management in real contexts
Response inhibition	Go/No Go	Hostile (Go) vs. Benign (No Go) zombies according to dispatch information	Decreased false alarms	Improved and less biased decision making in presence of stressors
Bias resistance	Priming	Benign zombies partial matching (body type, but not eye color) with dispatch target information	Fewer primed false alarms	
Spatial attention	Target detection	Targets coming from multiple sides	Fewer zombies reaching the player unspotted	Improved situational awareness

Note. The items of the last column (Envisioned long-term goal) are part of the model as they served as a rationale for the incorporation of the relative psychological mechanism in the game. Testing those effects in real-life policing contexts was however outside the scope of the present thesis.

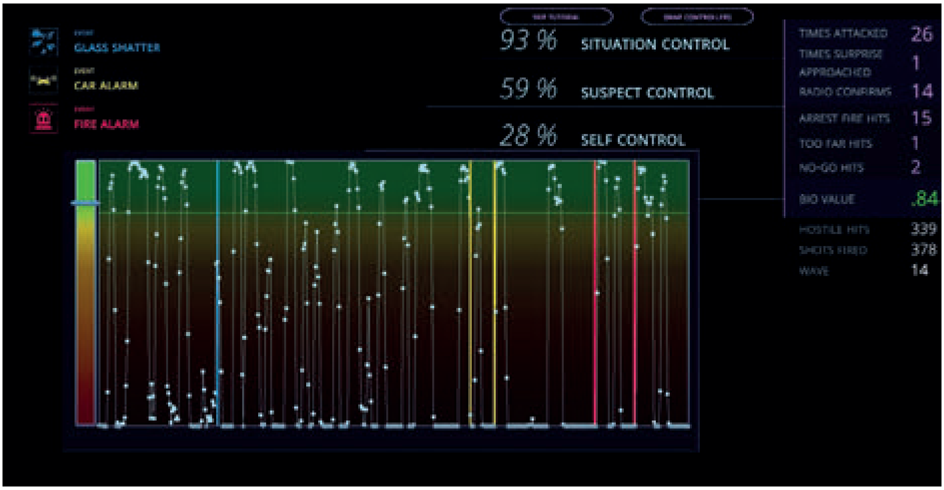
Additionally, and also in line with the design principles listed by Scholten & Granic (2019), we involved stakeholders in the creation of the training to ensure that it addressed the diverse needs and perspectives of the police. Some crucial features of the game, like the use of non-realistic targets to avoid unwanted transfer of behavior to real-life and the implementation of the BF as a "tunneled" vision impairment come from this collaboration. Another important feature that this early collaboration with the police stakeholders allowed is the extension of automated feedback from just the BF score to behavioral markers of decision making and spatial attention. Those extra feedback features, presented to the players at the end of each VR training round provided the player with more information to "chew on" until the following game session (See Figure 1.2). As those composite scores provided multifaceted feedback to the participant, they presented the additional benefit of diminishing the risk of having too simple of a metric to discriminate participants. Indeed, participants in a police center tend to compare their scores as professional self-efficacy is an important aspect of police officers' feeling of self-worth (Love & Singer, 1988).

The elements of Table 1.1, are discussed in the following paragraphs. Since there were several psychological mechanisms for which our game is designed, several separate challenges were faced with the conceptualization and implementation of the game, as well as difficulties in translating the evidenced-based principles of scientific tasks into a novel, appealing and dynamic environment.

Psychophysiological regulation. Psychophysiological regulation (see Table 1.1) is the central psychological mechanism on which the training focused. As mentioned earlier in this chapter, improving performance under stress in real-life policing contexts is thought to require a mitigation of the stress response. According to the recommendations of Hayano & Yuda, (2019), we focused on HRV as a marker of stress regulation and how this can be influenced by breathing. As detailed in chapters 3 and 4 of this thesis, the psychophysiological regulation was achieved through the evidence-based technique of BF. It was initially implemented by rewarding slow abdominal *breathing*, which has been associated with increased HRV (Laborde, 2019). This was replaced in chapter 5 by directly rewarding *breathing induced fluctuations in HR*— also called respiratory sinus arrhythmia (RSA). This physiological marker is characterized by an acceleration in HR when the subject inhales, and a deceleration when the subject exhales, and has been suggested to be a more direct and reliable indicator of calmness (Hayano& Yuda, 2019). Studies have shown that this natural coupling between breathing and heart rate not only provides a feasible biofeedback parameter, but that it is effective (Lehrer et al., 2020) and also correlates well with established HRV markers (Bornemann et al., 2016, 2019).

Figure 1.2

Example of the scores presented at the end of each VR game session to the player.

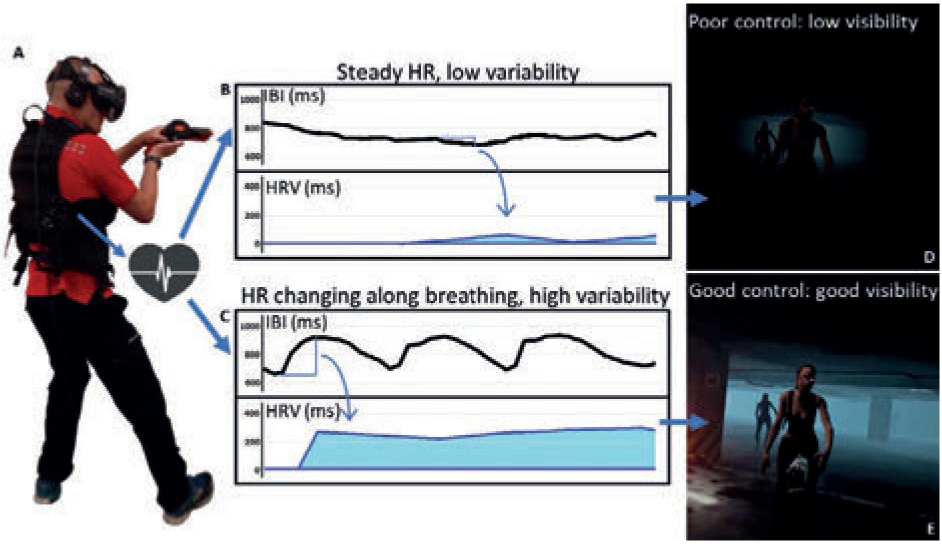


Note. When presenting this score panel to the participants, the experimenter would detail it, by mentioning that the three control scores (Situation-control, Suspect-control, Self-control) are built by using the more fine-grained scores presented in the rightmost portion of the screen. The central graph represents the evolution of the HRV/respiration of the participant throughout a VR session.

Irrespective of the physiological marker used for BF, its game mechanic was the same (see Figure 1.3 for an example of its implementation in chapter 5). Namely, when the player approached the desired physiological state their vision in the game improved, while moving away from that state reduced their vision by darkening their peripheral vision up to the point where only a small portion of the central field of view was still visible. This visual impairment was implemented to mimic the commonly reported phenomenon of “tunneled vision” appearing under stress, which is very familiar to police officers (Baldwin et al., 2022; Dirkin, 1983; Klinger & Brunson, 2009). The game recorded the BF performance of each player during a VR session, thus allowing to calculate BF scores to estimate the evolution of their psychophysiological regulation capacity. Additionally, the game automatically calculated the BF score of a whole session. This proximal outcome of psychophysiological regulation was displayed to the player at the end of each VR session. We expect that providing a capacity to manage psychophysiological stress in an action context will enhance the chance of transfer to real policing contexts, which is our envisioned long-term goal of the training. Skill transfer to different contexts has already been shown for neurofeedback training of the amygdala (Keynan et al., 2016) as well as HR-based BF in emotionally challenging (Peira et al., 2014) and stressful contexts (Parnandi & Gutierrez-Osuna, 2017, 2019).

Figure 1.3

A BF training game for Heart-Rate Variability (HRV) control under high arousal.



Note. Implementation of the biofeedback (BF) as closed-loop peripheral vision modulation to reflect the negative consequences of attentional narrowing (tunnel vision) occurring as a consequence of the stress response. The quality of the vision was linked to the BF score, calculated from the Local power HRV; (A) A police trainer in the VR training; (B) Exemplary traces of inter-beat-intervals (IBI) and relative HRV when the participant's HR is stable and does not fluctuate along with breathing. HRV is computed by measuring peak-to-through differences in IBIs, where larger fluctuations correspond to a higher HRV score; (C) IBIs and relative HRV traces when the participant's HR is fluctuating along with deep breathing; (D) Low visibility, associated with small breathing induced HR fluctuations (poor control); (E) High visibility, associated with larger breathing-induced HRV fluctuations (good control).

Response Inhibition. Another important requirement for our training was that the physiological regulation had to happen in an active decision-making context. This allowed us to monitor and train specific psychological mechanisms that are known to suffer when facing stressors. Response inhibition is the ability to withhold a response that is triggered relatively automatically by context, habituation or training. In policing realities, this could translate in the capacity to refrain from excessive use of force when facing insults or other situations where force seems a trivial solution to the problem the officer faces (Staller et al., 2018). In the training, response inhibition is implemented through a Go/No-go decision task, an evidence-based task where players must refrain from making a habitual response (shooting a target). Previous studies have shown that withholding such a response can be quite difficult, especially in high-stress situations where it may be a self-preservation attempt (Hashemi et al., 2019; Nieuwenhuys et al., 2015). The ability to withhold a response is also linked to prefrontal cortex activity (Föcker et al., 2018), which is impaired by stress responses (Hermans et al., 2014). In terms of game mechanic, response inhibition was implemented by asking the player to only shoot at zombies matching a description provided to the player by radio dispatch information, thus sparing “benign” zombies. The game kept track of each decision made by the player, thus allowing to calculate the number of false alarms (benign zombies shot by the player) for each session. This proximal outcome of response inhibition was displayed to the player at the end of each VR session. As an envisioned long-term goal, we expect biofeedback training to help police officers by improving their decision-making capacities, as already illustrated by the positive effect of HRV BF delivered in close proximity to action moments (Andersen & Gustafsberg, 2016).

Priming. Just as response inhibition, resisting biases is a psychological mechanism of crucial importance in a high-stress police encounter. Previous studies have shown that the radio dispatch information given to police officers prior to an encounter can bias their decisions towards excessive use of force (D. J. Johnson et al., 2018; Taylor, 2020). Priming is an evidence-based technique that was introduced in the VR game to measure and train (resistance to) dispatch induced bias. During a VR training session, the player faced several waves of zombie-like targets (see Figure 1.4). Each zombie wave was preceded by a radio dispatch message detailing the eye-color that hostile zombies would have (red, blue or yellow, a reliable identifier of the go targets) and the body type that these zombies were *expected* to have (small/large, male/female; an unreliable identifier of the go targets). This double-identifier feature for target identification allowed us to bias expectations. Throughout the VR session, while the dispatch information for target identification by eye-color was always correct, the information about the easier-to-identify body type became less reliable. This game mechanic allowed us to investigate to what extent dispatch information could bias the player towards shooting benign zombies (No-Go targets with the primed body type but the incorrect eye color) and whether this was associated to the physiological state of the player. The amount of priming-related mistakes of the player was recorded by the game, but this proximal outcome of bias resistance was not displayed to the player at the end of the VR session.

Figure 1.4
Sample of zombie types encountered in the VR game.

Note. All the possible combinations of eye colors (red, blue, yellow) and body types (small/large, male/female) could be encountered in-game.

Spatial Attention. The final psychological mechanism incorporated in the game was spatial attention. Psychophysiological stress alters attention in a variety of ways, from making us focus on more coarse visual features (Lojowska et al., 2015) and tunneling our vision (Dirkin, 1983) to distorting information processing and creating attentional biases (Aue & Okon-Singer, 2015). Since VR allows events to happen all around the player, it is a perfect environment to test how spatial attentional is allocated in the game. To that end, the time that a target spent in the field of vision of a player before being shot or reaching them (without being shot) was recorded. This metric provides insights on how many targets could approach the player without being spotted, and is particularly relevant for the officer's safety, as policing action usually requires cautious monitoring of the environment. The number of zombies reaching the player undetected was used as a proximal outcome for spatial attention, and displayed to the player at the end of each VR session. We expect this training element of the game to enhance situational awareness of police officers, as indirectly suggested by the increased situational awareness witnessed in drivers confronted with a training that pushes the need for enhanced situational awareness (Walker et al., 2009).

Early attempts at intervention development

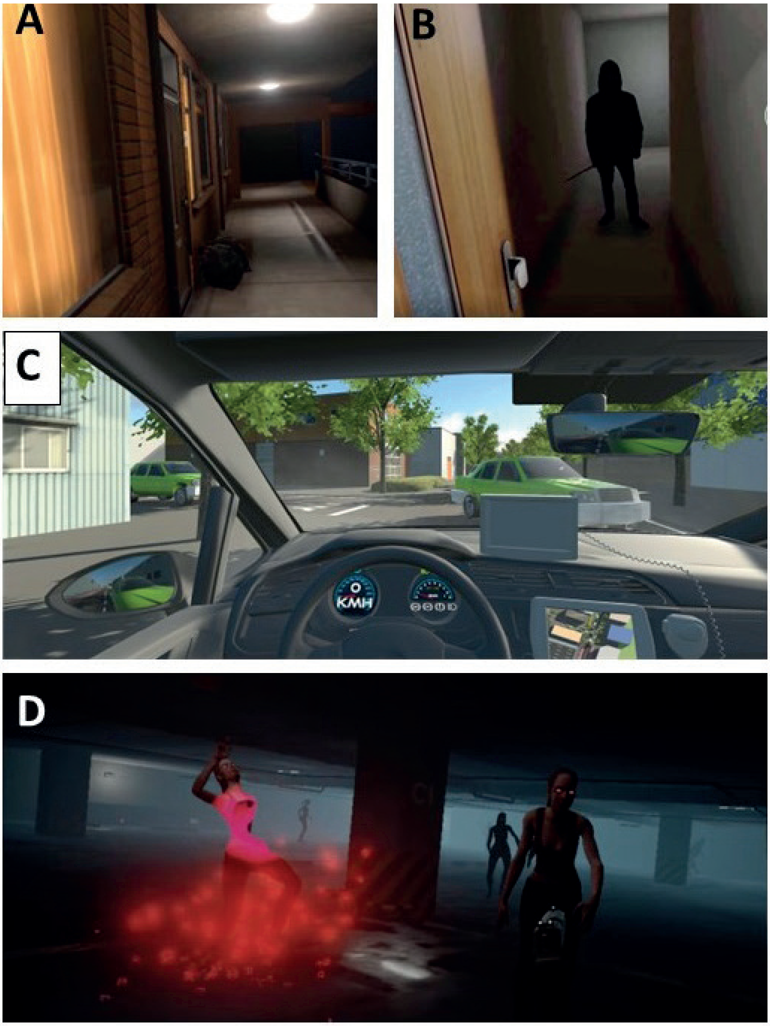
As shown above, there were a number of problems to be tackled to head toward improving the policing performance in presence of stressors and achieve the goal of skill transfer in HRV training. Several design attempts were made before we reached the final solution that was described above. These early attempts are hereby shortly documented as they further explain the process leading to our rather unconventional choice of training police officers with a "zombie-shooter" game.

From a methodological standpoint, the choice of Virtual-Reality to generate a training context for physiological control was driven by the possibilities of immersion and control that the technology grants (Parsons, 2015). Indeed, the complaints of the Dutch police concerning the lack of relatability of standard BF environments could not be overcome without proper immersion. Helpfully, VR has been praised as a very efficient way to create environments that are both representative of real life and allowing control to the finest details (L. C. Miller et al., 2019). VR is moreover considered a reliable way to elicit genuine emotional responses (Parsons, 2015) and to boost learning through engagement (Allcoat & von Mühlenen, 2018; Weerdmeester et al., 2017, 2020). Ultimately, and of particular importance for the current thesis, VR environments have been shown to be reliable fear and stress elicitation paradigms (Lin, 2017; Reichenberger et al., 2017).

For relatability and representativeness, in our first design attempt we tried to keep the VR environment as close as possible to realistic policing situations (see Figure 1.5A and 1.5B). The idea was that content similarity between training and real life would enhance the chances of skill transfer (Oudejans, 2008). Our early design attempts therefore included an ongoing house robbery, domestic violence scenes and similar realistic situations. The complex nature of those scenarios required the implementation of numerous features such as dialogues and physical interactions. Since those expensive features far exceeded the limited budget at our disposal, the scenarios were then reduced to a "simple" building search, where the participant would have to search several apartments in a building to find an armed robber. The idea was to give to the whole game a "shoot/don't shoot" dynamic, since a complex communication/handcuffing procedure was not manageable with the technology at our disposal. There, our problems (thoroughly listed in chapter 2) started. Indeed, this design was putting the participant at repeated risk of shooting innocent bystanders (which was not allowed by the police stakeholders for ethical reasons) and would interfere with taught police practices since many policing procedures had to be overlooked in the scenario. Creating scenarios with a large number of possible actions at any given moment would become very hard to design and would not allow for reliable measures of behavior.

Figure 1.5

Visuals of the earlier and current iterations of the VR game



Note. These visuals are screenshots taken from examples of playthrough moments of early and current versions of the VR game. We first investigated a building search scenario where the officer would have to (A) approach an apartment announced via radio dispatch as potential theatre of an ongoing burglary and (B) decide if the suspect had to be neutralized or not. In a later scenario version (C) the officer was asked to chase a getaway car. The final version of the game (D) was built around a zombie-shooter dynamic taking place in a poorly lit garage. Officers had to decide whether to shoot incoming zombies or not based on radio dispatch information.

The building search idea was therefore abandoned in favor of a “car chase” scenario, where the player would have to pursue an escaping car (see Figure 1.5C). The scenario had the advantage of reducing the risk for innocents being killed (the roads were empty), but made it very difficult to provide behavioral metrics that could distinguish between good and bad decisions made by the participant. The additional problem encountered with the “car chase” scenario was that it required the driver to frequently look on the side and change the speed of the car, thus reinforcing a mismatch between the information provided from the vision and from the vestibular system, ultimately leading to nausea (Padmanaban et al., 2018). Ultimately, participants also reported that the whole scenario felt rather dull and not very engaging. Due to the severity of the nausea symptoms experienced in pilot testing, as well as the difficulty to create an experience that was both engaging and stress-inducing while staying true to police realities, the scenario was ultimately discontinued.

For the final, successful design we went back to the drawing board and concentrated on the key role of the training environment (see Figure 1.5D), which consisted of creating a game eliciting a genuine feeling of threat, stress and sense of being overwhelmed. Pure “realism” was therefore abandoned to favor the “believability” of the stress inducing scenario, as the game designers of our team recommended (see chapter 2).

Thesis layout

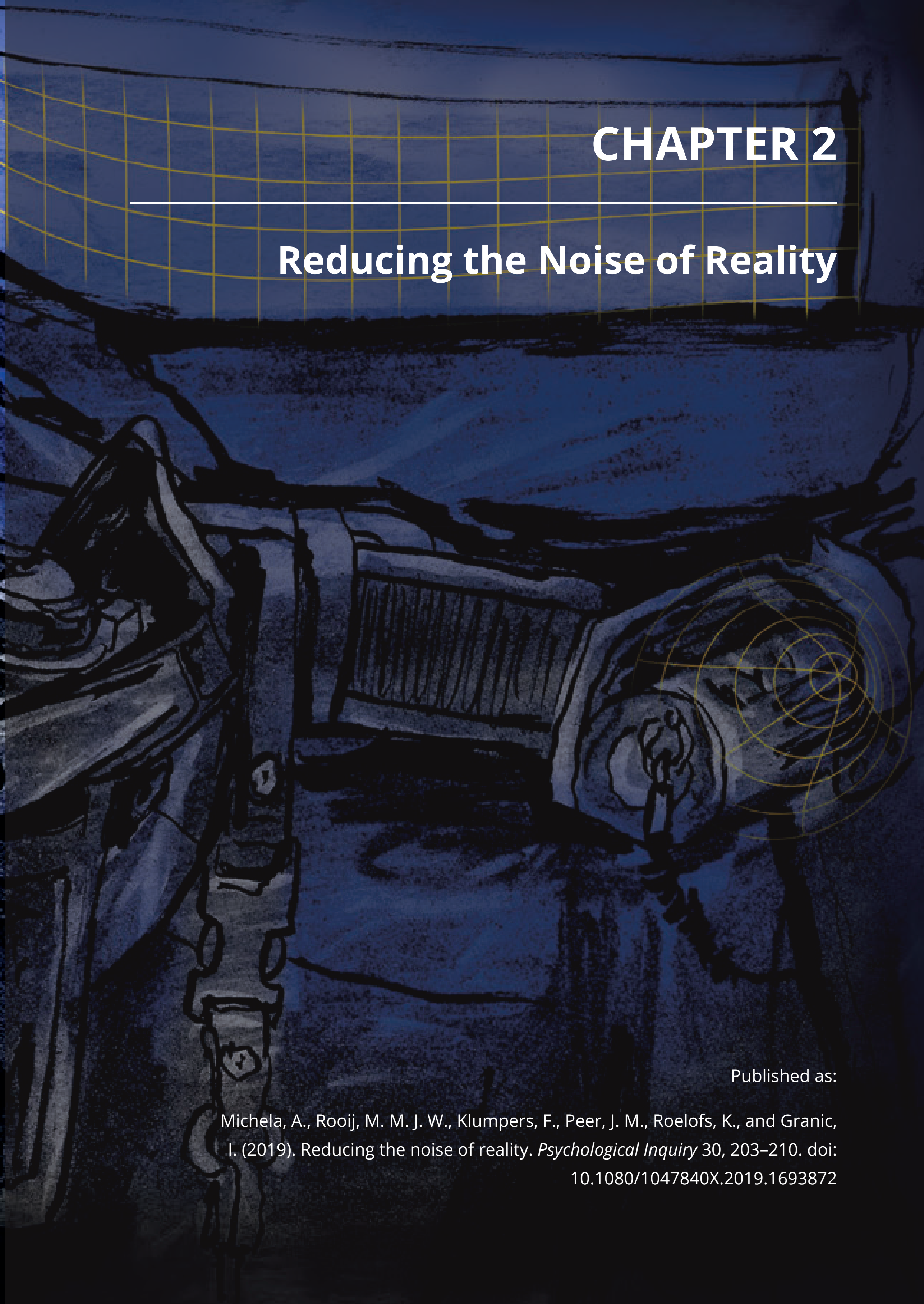
The present thesis investigated how a VR game-based BF training could be used to help police officers train control over their psychophysiological stress response, and ultimately improve their decision-making performance when facing stressful situations. **Chapter 2** is a theoretical article responding to L. C. Miller et al. (2019), who proposed a methodology to enhance the causal inference affordances of representative environments, thanks to VR. Our response focused on highlighting the boundaries and limitations of VR when used for scientific and training applications. Indeed, due to the relative novelty of the technology, many promises are made about its capacity to provide realistic, yet controllable, environments for scientific investigations (Baños et al., 1999; Blascovich et al., 2002; Cipresso et al., 2018; de la Rosa & Breidt, 2018; Pan & Hamilton, 2018; Parsons, 2015). Therefore, in chapter 2 we warn potential users for the complexity of developing realistic scenarios, and discuss the challenges of interpreting data that are measured in such complex environments, which can prove very hard to model. Hence, this chapter capitalizes on our initial experience in VR game design and highlights the main limitations considered and design principles followed when designing our VR game.

Just as designing a VR environment poses specific challenges, identifying and measuring suitable physiological markers for in-action biofeedback can prove very arduous. **Chapter 3** goes beyond theory and details the challenges and considerations that we faced in developing our own BF training, illustrated with pilot data. Specifically, the chapter deals with how *in-action* BF signal analysis has to deal with several problems, like movement artifacts, that make the usual trade-off between signal quality (requiring a delay of feedback presentation) and players' sense of controllability (requiring feedback as fast as possible) even more complex. Although the BF signal discussed in this chapter and the next (chapter 4) is different from the BF signal used in chapter 5, most of the considerations detailed in this chapter are still applicable to chapter 5.

Chapter 4 is a first qualitative small *N* study that investigated the effect of training police trainers for multiple days with our VR game. The aim of this study was to carefully document the evolution of the training over a 1-month period, comprising 10 training sessions. The results were used to determine the feasibility of biofeedback training in an active context, evaluate the likelihood of reaching a training plateau and monitor how the training was appraised and experienced by the participants. Many features of the experimental schedule and discoveries made in this explorative study were used to design the following study, a full-scale RCT.

Chapter 5 is the culmination of the thesis and tested the effectiveness of our VR game to train police officers in decision making in stressful situations. This chapter reports on a large-scale multi-day randomized control trial performed across many police centers in the Netherlands. In this study, the training schedules and design were based on the insights gathered in chapter 4. The BF target was adapted to HRV, to more directly reflect the psychophysiological state of the participant, and the transfer of the trained skill to a new police-relevant task outside VR was measured. This last study also included an exit questionnaire that allowed us to measure the degree of satisfaction the police trainers experienced in taking part in our training.

The final chapter, the **General Discussion**, summarizes and discusses all the findings of the current thesis, such as the effectiveness of the training and the presence of skill transfer. The chapter further summarizes the strengths and limitations of the present research and makes suggestions for future research. Among the strengths are the capacity of the chosen study designs to investigate causal relationships between different elements of the training intervention, as well as the reliability of the engagement elicited among the players. Weaknesses are mostly related to the lack of relation between physiological and behavioral results. Lastly, considerations are provided on the most important lessons regarding intervention designs, the use of VR, and the challenges of defining actionable outcome measurements for policing.



CHAPTER 2

Reducing the Noise of Reality

Published as:

Michela, A., Rooij, M. M. J. W., Klumpers, F., Peer, J. M., Roelofs, K., and Granic, I. (2019). Reducing the noise of reality. *Psychological Inquiry* 30, 203–210. doi: 10.1080/1047840X.2019.1693872

A Commentary on: Causal Inference in Generalizable Environments

Miller and colleagues (L. C. Miller et al., 2019) propose a novel and inspiring theoretical framework that merges systematic and representative experimental design. As such, the Systematic Representative Design (SRD) framework has the potential to move psychological science forward in a significant way. Essential to the authors' proposition is the default control group (DCG), an experimental condition that leverages the possibilities of new technologies such as Virtual Reality (VR) in order to create a close approximation to "real-life," but with the affordances of a tightly controlled experience. With this innovation, Miller and colleagues attempt to solve the incongruous needs, common in so many experimental designs, between generalizability (to everyday life) and experimental control (to claim causality).

Although the SRD framework could potentially result in a real shift in the way psychological research is conducted, it provides far from an 'off the shelf' solution. In particular the use of new technologies such as VR brings with it a number of complications - of both practical and theoretical nature – that are not fully addressed in the target article. With this commentary, we aim to contribute to the discussion on improving experimental design and thus empirical psychological research in general, by drawing from our experience in designing digital, game and neurofeedback-based interventions for mental health and behavioral change. Furthermore, we will suggest that there are important process similarities between SRD and game design, as well as common practical pitfalls. We will outline these links and discuss their implications in order to potentially further strengthen the SRD framework, especially for applied research purposes.

The potential of Systematic Representative Design

SRD is a promising framework that attempts to address, and provide solutions for, both causality and generalizability requirements in experimental design. This framework allows us to maintain a high level of experimental control in environments that are usually impossible to standardize without introducing a lot of noise in the measurements. Specifically, the strength and main novelty of the SRD approach is that it offers experimental control through a carefully constructed default control group (DCG) that serves as a highly controllable virtual substitute to reality.

The first benefit of a highly controllable, yet generalizable, environment in SRD is the reduction of noise and biases in the data that are collected in the testing environment. Specifically, the authors of the target article suggest that VR, unlike the typically austere psychology laboratory, allows for the design of an experimental environment that can mimic the natural, complex context in which behaviors of interest appear. In doing so, the VR experimental design limits some of the experimental biases (e.g., due to different motivations, semantic understanding, cultural or social expectations) that otherwise emerge from studying participants acting out-of-the ordinary in traditional lab contexts (e.g., Ceci, Kahan, & Braman, 2010). Moreover, in studies focusing on social interaction, the opportunity to use Virtual Agents (VA) in the VR setup helps researchers overcome the low internal validity and increased noise that is often introduced when human confederates are used (Kuhlen & Brennan, 2013).

On top of improving the reliability of the data being collected compared to naturalistic settings by increasing the consistency of the environment, SRD also aims to focus more precisely on the motivational core of participants' behaviors, keeping that core as consistent as possible across the full sample of participants. A large problem in "controlled" laboratory assessments comes from the range of unwanted individual differences that participants bring with them to a lab task (e.g., motivations, interpretations of the instructions, past experience with similar tasks, expectations, and so on). The SRD framework insists on taking these individual differences seriously and reducing the noise they introduce in experiments, by emphasizing the meaning of the action performed: Instead of relying on participants simply following task instructions, SRD experiments use contextualization (what is happening?) and interactive narrative (why is it happening and what should I do?) to prime participants to act consistently according to the role they are given. Thus, in participants' experience, they behave more in accordance with their natural internal motivation rather than the external motivation of complying with the experimenter. They *have a reason* to act as they are asked to. Indeed, "providing a role", has been shown to change behavior in simple setups like the ultimatum game, where being primed to imagine impersonating a banker significantly changed the way participants behaved (Lightner et al., 2017). In addition, the narrative context makes action *matter* to the participant. Stimuli become *affectively* salient and motivationally important as a result of these contextual enhancements (Parsons, 2015) for a review). In many research contexts, the increases in motivation and affective salience of stimuli should not only give rise to a better approximation of reality but also improve the reliability with which these effects are captured.

Drawbacks and challenges

Despite SRD being a sound framework from a theoretical perspective, there are important pitfalls that we anticipate researchers will face when they attempt to apply the framework to their own experimental work. We experienced many of these same pitfalls in our own research, in which we strived to design a suitable VR environment for studying decision-making under stress in police officers. We will review a number of these pitfalls both from an implementation and an ethical perspective, but also provide specific examples from our own research that illustrate the concrete challenges. Suggestions for solutions that we have found to address these challenges will be explained in the "proposed solutions" section.

Implementation concerns

One important aim of the SRD approach that is emphasized by the authors is the need for *generalizability to everyday life*. This concept resembles the more widespread concept of *ecological validity*, and will be used interchangeably in our discussions. Ecological validity can be defined as a combination of *verisimilitude* and *veridicality* (Franzen & Wilhelm, 1996). Verisimilitude is the level of believability or the extent to which an experimental task approximates the features of everyday life. Veridicality, in contrast, is the degree to which the performance of a participant in an experimental setup accurately predicts what that person would do in reality.

Enhancing verisimilitude. Enhancing the verisimilitude in an experimental setup can be deceptively appealing, but may introduce several problems. In an SRD context, enhancing verisimilitude requires isolating a target behavior, identifying the "most frequent setting" in which that "behavior of interest" appears and also the "relevant script components", and so forth. In other words, an SRD experiment working towards maximal verisimilitude may seem to aim at approximating a simple snapshot of reality - faithfully reproduced in VR and/or using by using VA. However, designing this type of VR environment leads to several problems of feasibility that need to be considered even before an SRD approach is taken, especially in VR.

First, among the most common drawbacks of VR are nausea and motion sickness: Up to 80% of participants are impacted by these physical symptoms in VR experiments (Stanney et al., 2003). Most concerning for VR design using an SRD approach, nausea and motion sickness are most often a problem when the aim of the VR design is to elicit "natural behaviors." More specifically, only a relatively small part of the population cannot physically tolerate VR that simply reacts to the natural movements of the player (e.g., an object grows bigger in order to appear closer when the user bends forward towards that object), but the proportion gets much larger when unpredictable or non-user-initiated artificial movement is experienced (e.g., a car hits a wall unexpectedly; e.g., Stanney & Hash, 1998). For example, car simulation games in VR (like Project Cars, https://www.projectcarsgame.com/vr/) minimize this issue by providing wide tracks, a predictable trajectory, a reduced feeling of braking and frontward gaze fixation points. Considering the success of such games, VR could seem like a perfect environment to a researcher investigating, for example, reckless driving. The virtual environment allows for highly controlled measures to take place, and even very "risky" situations to be safely re-created. Yet, the nausea-reducing measures present in VR driving games cannot be used in most setups aiming to enhance verisimilitude, as exemplified by driving in a city: it happens in narrow tracks, frequent braking and the need to constantly monitor the surroundings, which results in gaze diversion from frontward fixation. Thus, with the current technology, enhanced verisimilitude increases unwanted nausea effects, which paradoxically, reduce the generalizability of the task, because nausea can be so unpleasant.

A second drawback with the pursuit of verisimilitude is the uncanny valley effect (Mori, 1970), which is the feeling of eeriness or revulsion experienced while interacting with robotic or virtual avatars that mimic human behavior almost, but not exactly, perfectly. In the pursuit of ecological validity, researchers tempted to faithfully reproduce a snapshot of reality have a higher risk to encounter this effect than researchers limiting themselves to simpler, less realistic, stimuli. Research on this uncanny valley effect lacks consistency regarding the prevalence and explanation of the effect (Cheetham, 2017; Lay et al., 2016; S. Wang & Rochat, 2017). However, it may be that the effect is related to a discrepancy between expectations raised by an anthropomorphic "entity" and its observed behavior (Złotowski et al., 2015). Therefore, in the current state of most available technologies, the uncanny valley effect can pose a severe threat to the ecological validity of a task, as participants who feel disturbed by the perceived un-realism of the experimental setup are less likely to behave naturally.

A last, more mundane, drawback regards the limited availability of adaptable and evolved VA and complex interactive environments whether in VR or not. Their development requires large financial investments and might require niche programming expertise. These high costs and long development times are often overlooked in grant proposals and research designs, whilst being of paramount importance for the success of any project aiming at using VA to enhance ecological validity.

Aside from the implementation concerns, enhancing verisimilitude by approximating a snapshot of reality also brings with it analytic concerns: Given the complexity and richness of VR environments, how valid are our conventional analytic methods? Most of the tests used in traditional laboratory experiments rely on specific assumptions that are easily violated in VR setups that attempt to simulate reality as closely as possible. Human behavior is complex, occurring at different levels of analysis (perception, attention, interpretation, action), and changing over time (i.e., moment-to-moment). The methodological challenge of capturing this complexity and dynamic nature is illustrated by Brehmer's (1992) early research that attempted to re-create a generalizable, ecologically valid decision-making environment for firefighters. In his setup, a series of interdependent decisions had to be made in real-time in an environment that changed both spontaneously and as a consequence of earlier decisions. Since all the decisions made by a player were interdependent, the standard analytic strategy which aggregates all decisions across time and contexts was not feasible anymore because the assumption of independence was violated in this setup. This is why Brehmer's focus had to move to the general tactics and strategies used by the firefighters to achieve their pursued goal rather than momentary decisions. The same could happen in SRD experiments: When broadly reproducing real-life situations in an SRD setup, actions and decisions will be embedded hierarchically, they will influence one another, and they will be contingent on prior actions and decisions. As a result, individual actions or decisions are not the correct unit of measurement and aggregating those measurements (decisions) violate most General Linear Model assumptions. More sophisticated hierarchical analysis approaches could be envisioned, but without a severe constraint on the participant's action range those analyses would quickly grow out of hand. In other words, if an SRD experiment consists of a VR-based snapshot of reality, it is more like an interactive "video" of reality than a series of independent snapshots. Thus that design does not allow analyzing single action or decision moments in isolation. Taken together, the lack of independence of measurement points prevents fundamental mechanistic questions to be addressed adequately in such a high-fidelity reproduction of reality.

Enhancing veridicality. Enhancing the veridicality of an experimental setup raises another set of concerns regarding the way the data can be analyzed. Since veridicality is defined as the degree of prediction of the experimental setup on participants' everyday behavior, it seems meaningful to enhance veridicality by incorporating "real life" elements into a validated experimental setup. It actually is the only principled way to proceed at enhancing the prediction power of a laboratory setup on everyday life's behavior. Yet, sadly enough, every element of "reality" added introduces proportional levels of complexity. No level of technological sophistication will solve this problem;

in creating a close to real-world environment, one gets the corresponding real-world complexity for free. In a situation comparable to "real life", behaviors can have wildly different explanations, or be due to the interaction of a large number of sub-systems. Therefore, by making an experimental setup richer to enhance veridicality, the researcher might not be able to exclude potential alternative explanations of a measured effect, especially when trying to make an inference about the underlying mechanisms. Effects that are established in a controlled environment could also disappear altogether in this more complex environment.

A concrete example from our own work might be important to clarify our point. In an early iteration of an experiment on decision making under stress for police officers, we attempted creating an ecologically valid version of the laboratory shooter task standardized by Gladwin, Hashemi, van Ast, & Roelofs (2016). This task requires (police) participants to perform several trials of shoot-don't shoot decisions. The main effect we desired to reproduce from that task was a reaction time difference in shooting responses between high and low threat conditions, as found by Hashemi et al. (2019). Reaction times are measured as the latency between a target stimulus appearing on a screen and the recorded response from the participant. In the laboratory task the target stimuli appear instantaneously (an opponent with a gun or phone), thus allowing for a precise measurement of the time until a participant responds, with millisecond precision. In an ecologically valid VR setup however, even target stimuli with a seemingly sharp onset like someone opening a door or taking an item out of their pocket are not instantaneous enough for a precise reaction time measurement. In addition, the time duration of those very actions actually lasts much longer than the differences in reaction times that we wished to measure between conditions. As if that was not enough, since VR allows participants to look in any direction, target stimuli could be missed altogether. Hence, the effects of threat on response times observed in the laboratory task could disappear in the VR environment.

Also, when enhancing veridicality by adding "reality" to an established experimental setup, the problem of trial duration can also become a challenge. This concern can again be illustrated by our attempts to create an ecologically valid version of the shooter task mentioned above. In this task, participants were asked to perform 150 trials of shoot-don't shoot decision making. The task can be completed in about an hour and contains enough trials to reliably measure the within-condition effect (i.e., high versus low threat). In this set-up, it is virtually impossible to include contextual elements to make the decision similar to what police encounter in "real life". To have an idea of what a more ecologically valid example could be and what would be the implications for task duration, let us consider the task used by R. R. Johnson et al. (2014). In this task, deadly use of force decisions were inspired by real situations from police practice and trials lasted an average of two minutes. Transforming the shooter task by Gladwin and colleagues according to R. R. Johnson's example, while keeping the same number of repetitions to reliably detect the effect, would make it last almost five hours (without breaks). Thus, because of feasibility reasons, the effects reported in Gladwin and colleagues' actual experiment may not have been discovered, and could probably not be replicated, in a setup characterized by higher veridicality. This difference raises the question of the number of repetitions allowed by an SRD experiment relying on contextualization to elicit affective reactions, and therefore the effect size needed to reliably measure an effect with very few trials.

Ethical concerns

An extension to the implementation concerns we have outlined above are those related to what is ethically feasible in experimental setups. The effort to make an environment ecologically valid and able to elicit genuine behaviors of interest may raise a range of ethical concerns that may not apply to standardized, controlled studies. Specifically, as outlined by Pan & Hamilton (2018), there are several ethical risks that appear in immersive environments and interaction with VA: Enhanced personal disclosure can lead to privacy issues (Lucas et al., 2014; Rizzo et al., 2015), immersive environments could lead to changes in real-life behaviors through embodiment (Tajadura-Jiménez et al., 2017) and even false memories in children (Segovia & Bailenson, 2009).

As hinted by these studies, VR experiments could be considered to be emotionally charged environments because users are not only immersed in a story, with vivid graphics. It is also due to one's whole body – including gaze, body posture, physiological arousal, and so on – being directly impacted on by this enclosed, immersed simulation. As a result, emotional experiences and associated cognitive impacts may have long-lasting effects. Those effects might linger far longer than the actual VR experience itself and generalize outside the context of the experiment. Consequently, these studies suggest the potential risk of accidentally creating traumatic experiences in VR experiments that simply aim to assess behavior. Clearly, the point of using VR in a study is to increase "immersion", which in turn should elicit authentic emotional responses. If the responses are indeed authentic emotional experiences, they have the potential to impact the well-being of participants, as suggested by the use of VR for mood induction protocols (Baños et al., 2006), and the potential of those induction procedures to be used in clinical contexts (Herrero et al., 2014).

Moreover, the use of VR as a therapeutic intervention tool is a strong indication that it requires extra attention to potential side effects. Indeed VR is a promising tool used in the treatment of several disorders such as Post Traumatic Stress Disorder (Rizzo & Shilling, 2017), complicated grief (Botella et al., 2008), eating disorder (de Carvalho et al., 2017), and sexual disorders (Fromberger et al., 2018). Most of the leading scientists in those fields advise the use of extra care in considering the ethical issues specifically linked to VR. A comprehensive example of ethical guidelines can be found in the article of Madary & Metzinger (2016), where potential implications of the use of VR are reviewed extensively.

In our own work, we had similar issues around ethically designing an emotionally evocative task in VR, with high levels of ecological validity. In the previously described VR project, aimed at training decision making under stress for police officers, situations arose in which police participants could involuntarily shoot innocent bystanders. These situations seemed relatively benign and game-like to the VR simulation designers. However, our stakeholder, the Dutch Police Academy, quickly vetoed this training scenario because of the concern for triggering traumatic memories and the potential for desensitizing officers or reinforcing shooting behaviors that were against their training policies. Thus, it is important to keep these ethical considerations in mind from the start when designing VR tasks from an SRD perspective, to avoid creating unsuitable (and costly re-) designs.

Proposed solutions

To overcome the previously mentioned ethical and implementation challenges in the development of SRD experiments, we propose two main directions. The first is to extend the range of techniques proposed in the target article to experimental manipulations that include using powerful tools from neuroscience. We agree that techniques proposed by the authors like fMRI neurofeedback are a good start, but we see substantial additional advantages in terms of feasibility and opportunity for making causal claims from EEG neurofeedback applications, as well as brain stimulation techniques and other psychophysiological methods. Indeed, those techniques allow for a better use of VR capacities, by preserving the participant's head movement freedom. The second direction that we propose is the application of game-based approaches to SRD paradigms. We argue that it may be important to reconsider the importance of general *verisimilitude* and think about which specific elements of everyday life are necessary to claim generalizability. We suggest that less realistic game-based elements can often prove more effective in homing in on the core causal units necessary for generalizability claims.

Extending the use of SRD to other methods

One of the core promises of SRD is to allow for causal inferences by isolating mechanisms by testing the experimental condition against a default control group (DCG). Providing the experimental condition by modifying the DCG by changing parts of the VR environment or of the VA can be relatively straightforward for a certain number of applications (like changing the gender of an interacting virtual avatar). However, as previously mentioned, for more fundamental questions (e.g., such as investigating the neural underpinning of specific decision-making processes), difficulty can grow exponentially fast. This increased difficulty is mainly due to the need to exclude alternative explanations for observed effects and to make causality claims, which is what has historically driven experiments away from ecological validity (Burgess et al., 2006, for a narrative review). One interesting way to avoid or at least address these difficulties could come from the fields of neurofeedback, biofeedback and brain stimulation. Indeed, these techniques can provide a way to manipulate physiological parameters - and can therefore allow causal inferences - by being used to create an experimental condition without modifying the VR environment or VA behaviors.

Miller and colleagues alluded to neurofeedback as a field of applications that would benefits from the SRD framework. We agree, but argue that this benefit could go both ways, as neurofeedback has been used not only for interventions, but also as a tool to perform experimental manipulations in fundamental research (Sitaram et al., 2017, for a review). Moreover, contrary to the authors' suggestion that fMRI neurofeedback is ideal for incorporating into SRD designs, we suggest that there is a wide range of opportunities offered by the more VR-friendly EEG neurofeedback (as exemplified by the work of Vourvopoulos et al. 2019). This latter technique has already proven its potential in investigating fundamental neuroscientific questions, like the trainability of brain plasticity (Tomas Ros, Munneke, Ruge, Gruzelier, & Rothwell, 2010), EEG biomarker's connection to psychopathological symptoms (Tomas Ros, J. Baars, Lanius, & Vuilleumier, 2014), the normalization of scale-free dynamics in EEG in (T Ros et al., 2016) and research on stroke patients rehabilitation (Tomas Ros et al., 2017). Moreover, the spatial resolution of EEG neurofeedback has recently been extended by using machine learning to extract deeper, non-cortical, signals like those from the amygdala (as elegantly shown by Keynan et al., 2019)

Specific brain activity could therefore be modulated either endogenously with neurofeedback or externally with brain-stimulation techniques, in controlled SRD setups. These possibilities would remove the chore of providing an experimental condition from the VR environment design. In other words, in such an experiment the VR and VA components would be constant across the control and experimental group, and the experimental manipulation would happen in terms of brain training or stimulation only. Experimenters could therefore perform targeted mechanistic manipulation of brain activity on participants interacting with a generalizable environment. If the VR environment was built in such a way that it was, indeed, reliably eliciting the behavior of interest, it would allow direct claims to be made regarding the implication of specific brain activity patterns in everyday functioning. One example could be studying the effect of disrupting (or training) posterior alpha oscillations, which has been linked to visual attention (Rihs et al., 2007) in ecologically valid VR contexts. This could allow researchers to link the phase-amplitude coupling of the alpha-gamma waves (Pascucci et al., 2018) to behaviors generalizable to "real life", which could in turn prove very useful to better understand the neural underpinning of the daily attentional processes.

Finally, the range of techniques used to create the experimental condition in SRD could be further expanded to biofeedback protocols controlling non-cerebral psychophysiological markers. Biofeedback could be used to investigate well-established psychophysiological effects in environments that provide a higher degree of generalizability to everyday life than traditional experimental designs. A few examples could be studying (through training) the role of easily measurable markers like heart-rate variability for stress management (B. Yu et al., 2018) or anticipatory bradycardia for decision making (Roelofs, 2017). In these two specific examples, studying the contribution of well-established psychophysiological markers in generalizable contexts would pave the way for a wide range of affordable interventions aiming at changing, for example, the behavior of patients suffering from anxiety disorders.

Careful integration of reality: Make it a game!

We have argued that there are a host of potential pitfalls in designing SRD studies that aspire to maximize verisimilitude and veridicality, with the goal of increasing generalizability of research results and making stronger causal claims. An alternative and more promising approach to create an SRD experiment could come from the game design world. As we know, building the DCG requires isolating the core causal units needed for that generalizability. As long as these core units are maintained in the design, we may reduce realism in our VR simulations and still elicit the behavior of interest. We suggest that elements common in commercial games can offer such a solution by making the environment not *realistic*, but rather *believable* (Schubert, 2013). Where *realism* is achieved by faithfully reproducing reality, *believability* is achieved by engaging the player through several game mechanisms, like challenge or emotional narratives.

When Miller et al. suggest that an SRD experiment could look like "serious games" they run the risk of directing the player toward the pursuit of *realism* instead of *believability*. Serious games are developed for educational interventions (e.g., skill training) and even if they usually attempt to include a "fun" component, it does not usually compare with the entertainment value of commercial games (Baranowski et al., 2016). We propose that using conventional gamification techniques incorporated in serious games will not be enough to obtain generalizability, largely because serious games are usually simplistic simulations that pay little attention to the emotional, engagement, and motivational underpinnings of player's experience (Schoneveld et al., 2019). This is why we think that serious games do not offer a solution for many applications of the SRD approach. However, we do advocate for game-based approaches that include *believable* core concerns of participants' motivations and engagement that are fundamentally embedded in an emotionally relevant context.

To illustrate the difference between a *realistic* "serious" game that feels artificial and a *believable* one that is experienced as relevant and motivationally engaging, we make the analogy of an SRD "serious game" being like an empty plate on which we attempt to grow a bunch of human cells, while a believable game is a petri dish containing all the essential nutriments required for those cells to grow. The cells on the petri dish survive, whereas the cells on the empty plate quickly die. Similarly in an SRD experiment, in many cases the behavior of interest in a study could not "survive" in the DCG if the context does not provide the correct motivational, affective, and engaging conditions. A successful SRD experiment relies on a tradeoff between approximating real-life conditions (by using VR and VA in a *believable* way) while accommodating the scientific imperatives. We argue that providing motivational, affective and engaging elements that elicit ecologically valid behaviors in participants can be best achieved by making it into an engaging and entertaining (not "serious") game. After all, what makes a good game greatly overlaps with the needs of SRD:

An immersive experience with an interactive narrative that elicits genuine affective emotions in the participant.

Importantly, game-based interventions can provide solutions for some of the potential pitfalls of SRD studies we mentioned earlier: The narrative scaffolding directs participants' attention to what is at the core of the experimental goals and narrows the spectrum of actions observable in the experimental setup. Moreover, a high quality narrative can also provide a *believable* reason for many repetitions of any particular behavior, an otherwise key potential pitfall we mentioned earlier in SRD studies. Using real games as a reference can also mitigate the ethical concerns by reducing the risk of learning transfer of unwanted aspects of the simulation. In turn, a well-designed game setup can also greatly reduce the technical difficulties like the uncanny valley effect in the development of the SRD task, thus increasing the implementation potential and generalizability to everyday life.

The final benefit of applying SRD in the form of real games, based on our own experiences in the previously described police project, is that it forces the researchers to adopt a more iterative stance in their design process, which is crucial for a successful experimental setup using VR. Working closely with game designers often means that researchers need to (at least partially) adopt design thinking principles in their development process (Scholten & Granic, 2019). Creating a suitable VR environment - with an immersive narrative and emotional experience - often requires more than just handing a list of requirements to game designers. These designers should instead be included in the experimental design process very early on, and help the researchers develop the task in an iterative procedure that will invariably challenge the scientist's original assumptions. Indeed, as the complexity of an SRD experimental setup - sing VR, VA, and potentially other technologies combined - is orders of magnitude higher than traditional experiments, the design process may often resemble the ones used in commercial game companies. This process, called the Rapid Iterative Testing and Evaluation (RITE) method, has been advocated for more than a decade (Wixon, 2003) and can prove useful for testing assumptions and design choices in small consecutive steps, instead of designing and testing the full complex setup at once. It requires the early involvement of stakeholders, to test and refine each prototype until reaching combined scientific and design goals. This is a time-consuming endeavor, but as Miller and colleagues themselves implied, isolating the behavior of interest and constructing a DCG are arduous tasks and require a very flexible mindset to be achieved.

Conclusions

The article by Miller and colleagues (L. C. Miller et al., 2019) suggests that the needed shift in the way we study psychology can be provided by Systematic Representative Design (SRD). We agree with them and hope that our commentary will help researchers undertaking this endeavor to be aware of the potential challenges of designing an SRD-informed study. More than a cautionary note, we aimed to raise issues about the conceptual conversion cost involved. The straightforward use of Virtual Reality and Virtual Agents to increase ecological validity by simply adding "reality" to a controlled experiment might rarely pay off due to feasibility issues, ethical complications and increased complexity in analysis methods. Our first suggestion is to widen the range of techniques used to create the experimental condition to EEG neurofeedback, biofeedback and brain stimulation. Our second suggestion is to draw inspiration from the game design world when designing an SRD experiment, both for design process and final result, and aim at creating *believable* environments instead of *realistic* ones. If these kinds of cross-disciplinary approaches are taken, we are convinced that SRD will lead to genuinely novel psychological research protocols that address many of the problems of generalizability that the field has suffered from for decades.

CHAPTER 3

Breathing Biofeedback for Police Officers in a Stressful Virtual Environment: Challenges and Opportunities

Published as

Brammer, J. C., van Peer, J. M. †, Michela, A. †, van Rooij, M. M. J. W., Oostenveld, R., Klumpers, F., Wendy Dorrestijn, Isabela Granic and Karin Roelofs (2021). Breathing biofeedback for police officers in a stressful virtual environment: challenges and opportunities. *Frontiers in Psychology* 12:586553. doi: 10.3389/fpsyg.2021.586553

† These authors have contributed equally to this work

ABSTRACT

As part of the Dutch national science program "Professional Games for Professional Skills" we developed a stress-exposure biofeedback training in virtual reality (VR) for the Dutch police. We aim to reduce the acute negative impact of stress on performance, as well as long-term consequences for mental health by facilitating physiological stress regulation during a demanding decision task. Conventional biofeedback applications mainly train physiological regulation at rest. This might limit the transfer of the regulation skills to stressful situations. In contrast, we provide the user with the opportunity to practice breathing regulation while they carry out a complex task in VR. This setting poses challenges from a technical – (real-time processing of noisy biosignals) as well as from a user-experience perspective (multi-tasking). We illustrate how we approach these challenges in our training and hope to contribute a useful reference for researchers and developers in academia or industry who are interested in using biosignals to control elements in a dynamic virtual environment.

Keywords: biofeedback, virtual reality, stress exposure, user experience, physiological computing

INTRODUCTION

Acute physiological stress impairs performance by causing deficits in motor control, cognition, or perception (Andersen & Gustafsberg, 2016; Nieuwenhuys et al., 2009; Nieuwenhuys & Oudejans, 2010) and can negatively impact mental health in the long term (Maguen et al., 2009). By teaching acute stress regulation, biofeedback could help preserve performance in challenging situations, and lessen the detrimental impact of repeated stress responses (Andersen & Gustafsberg, 2016). Since police are frequently confronted with demanding situations that require quick, high-stakes decisions, police forces, including the Dutch police, recently started introducing biofeedback to their training curricula (van der Meulen et al., 2018). However, the physiological regulation skills are usually exclusively taught at rest which might limit their transfer to stressful situations (Bouchard, Bernier, et al., 2012).

To make physiological regulation skills more robust to degradation under stress, we developed a training that combines biofeedback with a demanding task in virtual reality (VR). VR is increasingly used for stress-exposure training since it offers the opportunity to create immersive and stressful, yet controlled environments (Pallavicini et al., 2016). However, to date, the majority of biofeedback trainings do not leverage the potential of VR (Jerčić & Sundstedt, 2019). While providing a more immersive environment than screen-based biofeedback applications, current VR biofeedback applications require the user to stay relatively motionless and to solely focus on the biofeedback (Rockstroh et al., 2019; Van Rooij et al., 2016). In contrast, we provide police with the opportunity to recognize and regulate their physiological stress response while they carry out a demanding task in a stressful environment. Specifically, we provide an environment that requires the user to regulate their breathing while making fast decisions based on ambiguous, constantly changing information. We will refer to this kind of biofeedback as stress-exposure biofeedback. The promise of stress-exposure biofeedback has already been demonstrated in a military population, albeit in a non-VR setting (Bouchard, Bernier, et al., 2012).

Compared to conventional biofeedback applications, stress-exposure biofeedback introduces challenges from a technical and user-experience perspective. Here, we summarize and illustrate challenges in three critical areas: (1) the choice of a biofeedback parameter, (2) the implementation of the biofeedback processing, and (3) the biofeedback representation in the virtual environment. We hope to demonstrate the feasibility of stress-exposure biofeedback by illustrating each of these challenges with our implementation. Further, by sharing our experiences, decisions, and considerations we hope to contribute a useful reference for researchers and developers in academia or industry who are interested in using physiological signals to control elements in a dynamic virtual environment.

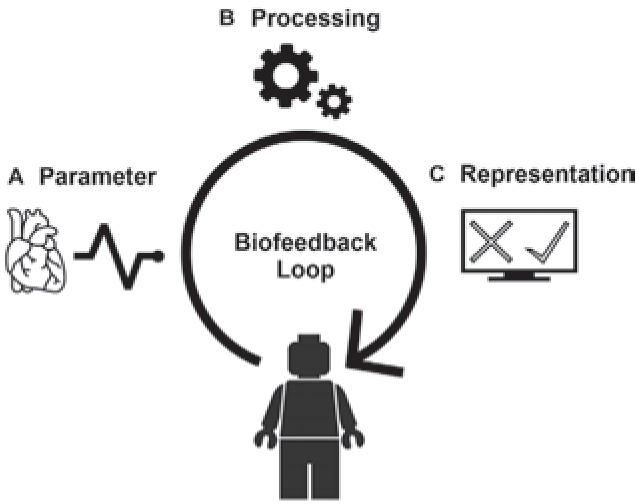
What is Biofeedback?

People are usually not conscious of their autonomic physiology, let alone able to regulate it (Price & Hooven, 2018). Biofeedback reveals internal physiological processes and provides guidance on how to change them, which can reduce anxiety and facilitate coping with stress (Tolin et al., 2020; B. Yu et al., 2018). In the following, we discuss how a conventional biofeedback application works and then show how it can be adapted for stress-exposure biofeedback.

Let’s consider an example of a trainee who is taught to downregulate their heart rate (i.e., the biofeedback parameter, Figure 3.1A). Electrocardiogram electrodes measure the electrical activity of the heart, which is send to a processing unit (Figure 3.1B). The unit estimates the trainee’s heart rate and applies a decision criterion that determines if the heart rate increased or decreased compared to the last measurement. The decision criterion is based on a biofeedback target, for example a decrease of 10%. Finally, the biofeedback representation reveals the outcome of the biofeedback processing (Figure 3.1C): for example, a green screen in case of a decrease, a red screen in case of an increase, or a blue screen if no change occurred. By rewarding the downregulation of the heart rate, the biofeedback system guides the trainee to the target through operant learning (Weerdmeester et al., 2020).

In the calm environments of conventional biofeedback applications, the trainee can fully focus on regulating their physiology to meet the biofeedback target. However, in a more stressful, demanding context, this is no longer possible since the environment might distract the user from the physiological regulation. Additionally, stress-exposure biofeedback poses technical challenges since acquiring and processing physiological signals is more challenging in dynamic conditions compared to resting conditions. That is, stress-exposure biofeedback creates additional demands both for the user and the developer. We will discuss these demands based on our application, in the context of the three challenges mentioned before: (1) the choice of the biofeedback parameter, (2) the real-time processing of the biofeedback parameter (evaluating match with biofeedback target), and (3) the representation of the biofeedback in the training environment.

Figure 3.1
The biofeedback loop.



Note. Biofeedback is a form of human-computer interaction which puts the user in a closed real-time loop. We divide this loop in three components that are discussed throughout the paper. First, a biofeedback parameter (A) is extracted from a physiological modality. For example, heart rate derived from an electrocardiogram. Second, in a series of processing steps (B) the current state of the parameter is evaluated relative to a target state. The processing results in a biofeedback score that expresses how well the parameter's current state matches the target state. The biofeedback score can be qualitative (match vs. no match) or quantitative (degree of matching). Finally, the biofeedback score is translated into a representation (C) consisting of (a combination of) visual, auditory, or tactile feedback which can be embedded in a variety of media, such as VR.

CHALLENGE 1: Choice of the biofeedback parameter

Prioritizing Controllability

To account for the trainee's divided attention during stress- exposure biofeedback, their control of the biofeedback parameter should be as easy and direct as possible. A variety of physiological modalities are related to stress and can serve as a basis for a biofeedback parameter, such as electroencephalography, heart rate variability or breathing (Tolin et al., 2020; B. Yu et al., 2018). These modalities differ in terms of their controllability and one of the easiest-to-control physiological modalities is breathing (Nacke et al., 2011; Parnandi & Gutierrez-Osuna, 2019). This is why we chose breathing rate as our biofeedback parameter, with a biofeedback target of4 to 12 breaths per minute (Russo et al., 2017), which is considerably lower than human breathing rates under cognitive or physical load (Hidalgo-Muñoz et al., 2019; Nicolò et al., 2017). Slow breathing affects the autonomic nervous system by increasing vagus

3

nerve activity and evoking a shift toward parasympathetic dominance (Russo et al., 2017). This might help regulate physiological arousal in an emotionally or cognitively challenging situation. In summary, breathing seems to offer both controllability and the ability to regulate physiological arousal. We evaluated the controllability of the biofeedback parameter and the achievability of the biofeedback target in a sample of nine police trainers. Each of them completed 10 training sessions over the course of three weeks. Each session lasted about 15 min and was played with or without biofeedback. Sessions with and without biofeedback were alternated in order to get an impression of the transfer of the physiological regulation skill.

The pilot data suggest that the biofeedback parameter is controllable and that the biofeedback target is achievable in a stress-exposure context. We observed that mean breathing rates decrease over sessions (Figure 3.2A, upper panel) and are lower in biofeedback sessions compared to sessions without biofeedback (Figure 3.3A). Many of the mean breathing rates fall within the biofeedback target range of 4 to 12 breaths per minute (e.g., Figure 3.3A). Similarly, participants continuously improve their biofeedback scores over the training sessions (Figure 3.2A, lower panel) and their mean biofeedback scores are higher in biofeedback sessions compared to sessions without biofeedback (Figure 3.3B, see challenge 2 for details on the biofeedback score). Moreover, the decreasing trend in breathing rate and increasing trend in the biofeedback score shown in Figure 3.2A do not seem to merely reflect the participants' habituation to the stressful environment. This is evident by the biofeedback-induced session- by-session fluctuations on top of the decreasing- (Figure 3.2A, upper panel) or increasing trend (Figure 3.2A, lower panel). These fluctuations seem to be an indication that, following biofeedback sessions, participants transfer the physiological regulation skill to subsequent sessions without biofeedback. Finally, we found the session averages of breathing rates and biofeedback scores to be strongly related (Figure 3.2B). This indicates that the biofeedback score is a valid representation of the extent to which participants manage to achieve the biofeedback target.

The Costs of Controllability

However, there are downsides to choosing a biofeedback parameter that is easy to control. As illustrated earlier (Figure 3.1), biofeedback is a form of human-computer-interaction (HCI). The human in the HCI is used to *immediate* and *invariable* control over the computer (Attig et al., 2017; Limerick et al., 2014). If someone presses the "k" key on their keyboard they expect the letter to appear on the screen instantaneously (immediacy). Also, they expect the letter to always be "k," not "o" occasionally (invariability). Immediacy or invariability cannot be guaranteed in a biofeedback system.

Variability Is More Noticeable

The non-deterministic nature of human physiology introduces variability to the biofeedback system. Consider the example of heart rate downregulation again: At some point the trainee may notice that they can decrease their heart rate by exhaling deeply. However, two identical outbreaths (in terms of duration and depth) don't necessarily produce the same decrease in heart rate. In general, a biofeedback parameter cannot be controlled in a deterministic manner. That is, even if a trainee consistently applies a specific regulation strategy, they will achieve variable outcomes in terms of the behavior of the biofeedback parameter. Variability is a greater challenge for biofeedback parameters that are under more direct control, because the trainee will have a clearer sense of their current physiological state. Consequently, they will more easily notice variability-induced discrepancies between their perceived physiological state and the biofeedback representation.

Delay Is More Noticeable

The biofeedback processing introduces a noticeable delay between the recording of the biosignal and the presentation of the feedback. To get a reliable and accurate estimate of a physiological state, it usually has to be integrated over longer time windows. For example, estimating instantaneous breathing rate requires at least a full breathing cycle and several breathing cycles have to be averaged to obtain a reliable and accurate estimate in the presence of measurement artifacts. Consequently, the biofeedback presented to the trainee will not pertain to their instantaneous physiological state. This can violate the trainee's expectation of immediacy, especially during abrupt shifts in the physiological state. For example, when someone shifts from slow, deep breathing to a markedly faster breathing rhythm or vice versa, the response of the biofeedback representation can appear sluggish.

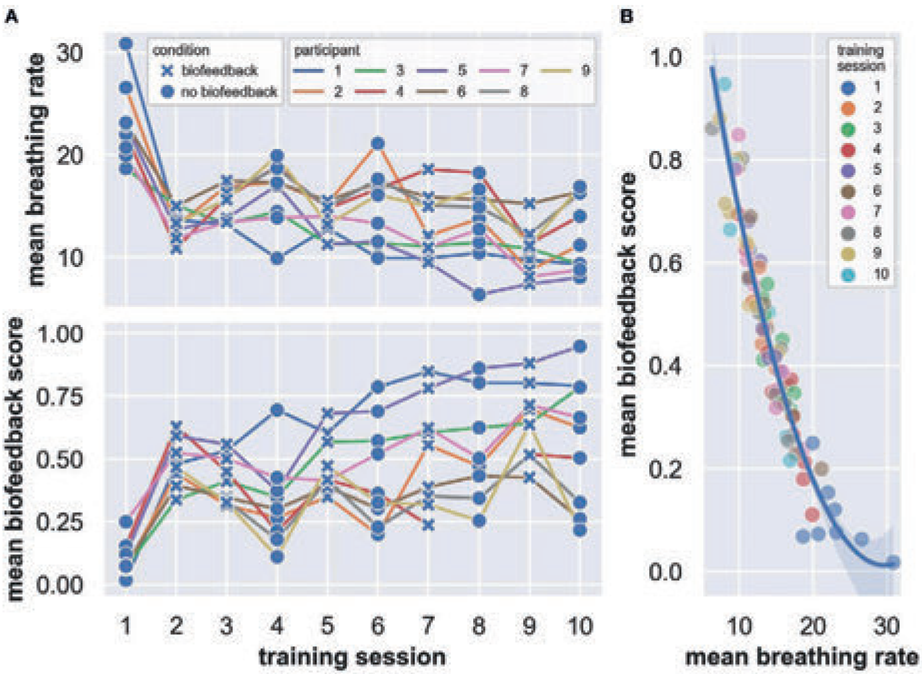
In summary, stress-exposure biofeedback benefits from a biofeedback parameter that is relatively easy to control. However, controllability comes at the cost of more salient variability and delay. Nevertheless, we think that controllability outweighs these costs especially since both variability and delay can partly be alleviated during biofeedback processing (Challenge 2) and the careful design of the biofeedback representation (Challenge 3).

CHALLENGE 2: Implementation of the biofeedback processing

The goal of biofeedback processing is to map the biofeedback parameter to the biofeedback representation. This involves two steps: First, the current state of the biofeedback parameter (e.g., breathing rate) has to be estimated. Second, the extent to which the current state of the biofeedback parameter approaches the biofeedback target has to be evaluated (e.g., breathing rate between 4 and 12 breaths per minute). Based on how closely the trainee matches the target, we then compute a biofeedback score that is ultimately reflected in the biofeedback representation (see Challenge 3). Supplementary Material 1 contains details on the hardware and software used for the biofeedback processing.

Figure 3.2

Mean breathing rates



Note. (A, upper panel) and biofeedback scores (A, lower panel) over training sessions (alternating with and without biofeedback). (B) Quadratic fit characterizing the relationship of the session means of breathing rate and biofeedback score. The shaded region indicates a bootstrapped 95% confidence interval for the quadratic fit

Estimating the Current State of the Biofeedback Parameter

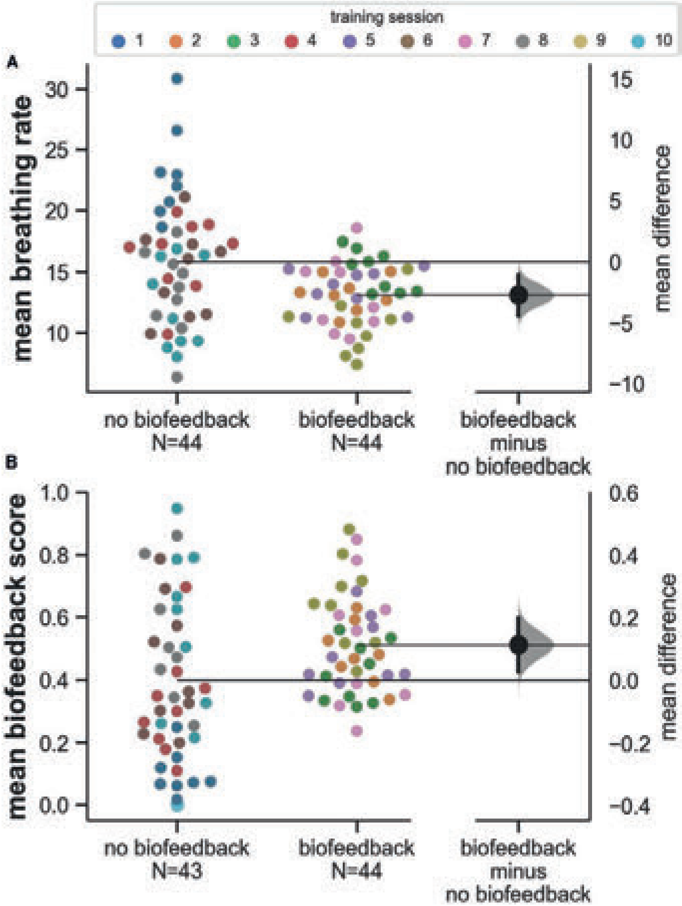
The first processing step is to estimate the current breathing rate from the raw data, which comes from a breathing belt around the trainee's lower abdomen. The raw data contains a phasic pattern with inhalation peaks and exhalation troughs, and conceptually, the instantaneous breathing rate is based on the temporal difference between moments of the same phase (e.g., inhalation peaks or exhalation troughs). Supplementary Material 2 contains a detailed description of how we estimate breathing rate.

Unfortunately, the raw sensor data does not exclusively reflect the dynamics of the biofeedback parameter. Instead, it contains artifacts that can originate from the measurement environment or unrelated physiological activity. For example, our breathing belt tracks breathing by measuring changes in torso circumference. However, since the trainee is standing and moving their upper body, the data contains movement artifacts that are in the same frequency range as (fast) breathing and cannot easily be filtered out (Supplementary Material 2 and Supplementary Figure 3.4). Regardless of the physiological modality, artifacts tend to be more prevalent in stress-exposure biofeedback compared with biofeedback at rest and from a user-experience perspective they contribute to both the problem of variability and delay (see Challenge 1).

Artifacts increase variability, which can frustrate the trainee because it can make the biofeedback target seem unattainable. This problem can be alleviated by making the biofeedback target less specific. When the target range is narrow (e.g., breathing at 6 breaths per minute), the estimated breathing rate will more often be "off target" due to artifacts. In contrast, when the target range is broader, the influence of artifacts is less perceptible since the wider margin compensates for artifact-induced variability in the estimated breathing rate. However, if the target range is too broad the training goal can lose specificity from a user's perspective.

Additionally, artifacts can increase delay. The presence of artifacts makes estimating the breathing rate from a short segment of data unreliable due to a low signal-to-noise-ratio. To increase the signal-to-noise-ratio, longer segments need to be processed (Hassan & Anwar, 2010). However, this means that at each point in time, the current estimate of the breathing rate and corresponding biofeedback representation do not exclusively reflect the most recent physiological state. Therefore, there is a trade-off between delay and the reliability of the biofeedback: More reliable estimates of breathing rate from longer segments come at the cost of more delay. A good compromise between reliability and delay allows for a reliable estimation of the biofeedback parameter from a technical perspective, while still feeling relatively responsive to changes in the biofeedback parameter from a user's perspective.

Figure 3.3
Comparison of mean breathing rates



Note. (A) and biofeedback scores (B) in sessions with and without biofeedback. In sessions without biofeedback, the biofeedback score was computed and recorded but did not affect the game. The distributions on the right side of panels (A,B) display bootstrapped 95% confidence intervals for the mean differences between the conditions (Ho et al., 2019). Note that the condition differences are mainly driven by the first training session.

Comparing the Current State of the Biofeedback Parameter to the Biofeedback Target

In a second step, the biofeedback processing quantifies how much the current state of the biofeedback parameter matches the biofeedback target. This matching is then expressed quantitatively (e.g., percentage) or qualitatively (e.g., binary) in the form of a biofeedback score. The computation of the biofeedback score can differ widely between biofeedback applications. However, regardless of the specific application, the computation of the biofeedback score presents the developer with seemingly small choices regarding algorithmic parameters that can profoundly influence the user experience. For example, related to our application, we already mentioned choosing the upper and lower bound of the biofeedback target range. We illustrate additional parameter choices related to the computation of the biofeedback score in Supplementary Material 2. Making these choices based on iterative user testing is crucial to ensure a satisfactory user experience (Scholten & Granic, 2019). User testing is greatly facilitated by the ability to visualize the raw data and intermediate processing steps as well as the ability to adjust parameters in real-time. Therefore, we implemented a dashboard that allowed us to fine-tune the biofeedback processing in real-time to immediately experience the effects of different parameter settings (Supplementary Material 2 and Supplementary Figure 3.3).

CHALLENGE 3: Biofeedback representation in the virtual environment

Finally, the biofeedback score needs to be presented to the trainee in a meaningful and intuitive way. In our virtual environment, the trainee finds themselves at the center of a poorly lit parking garage where they are surrounded by zombies that can either be benign or hostile, which is indicated by their eye-color or body shape (Supplementary Material 3 and Supplementary Figure 3.5). These indicators change several times throughout the training, which is announced via radio dispatch calls that mimic a suspect description. The trainee has to shoot the hostile zombies while leaving the benign zombies unharmed. In collaboration with our advisors at the Dutch police, we made an effort to steer clear of the "shoot 'em up" genre of video games (i.e., reflexive shooting at uniformly hostile adversaries) by designed the shooting task such that the player is primed to make, careful, deliberate shooting decisions. Additionally, the task engages behaviors that are universally relevant to police: The trainee has to rely on good situational awareness, be constantly vigilant to changes in information, and be able to override response biases by flexibly incorporating these changes in their decisions (Di Nota & Huhta, 2019). At the same time, by eliciting police-relevant behavior in an overtly fictional environment, we sidestep the necessity to simulate realistic police incidents and avoid overtraining the officers to idiosyncratic elements of a realistic simulation (Michela et al., 2019).

In this environment the biofeedback representation needs to be as salient and intuitive as possible, since regulating physiology becomes part of a multi-tasking exercise. The player has to allocate cognitive resources to both the decision task as well as the physiological regulation, which can worsen performance on both tasks (Wickens, 2010). Additionally, high task demands in a multitasking context can increase heart- and breathing rates (Fairclough et al., 2005). Together, these findings suggest that deliberate physiological regulation may be especially challenging in a multitasking context. Further, the multi-tasking bears the danger that the biofeedback is misattributed to behavior instead of the physiological regulation. For example, the trainee might attribute a poor biofeedback score to shooting a benign zombie rather than their fast breathing. This misattribution can be prevented by presenting the biofeedback such that it intuitively represents physiology in the task context. We use the analogy of tunnel vision which is relatable for Dutch police since they are introduced to this concept during their academy training (van der Meulen et al., 2018). Figure 3.4 illustrates how the trainee's peripheral vision widens as the breathing gets slower and deeper.

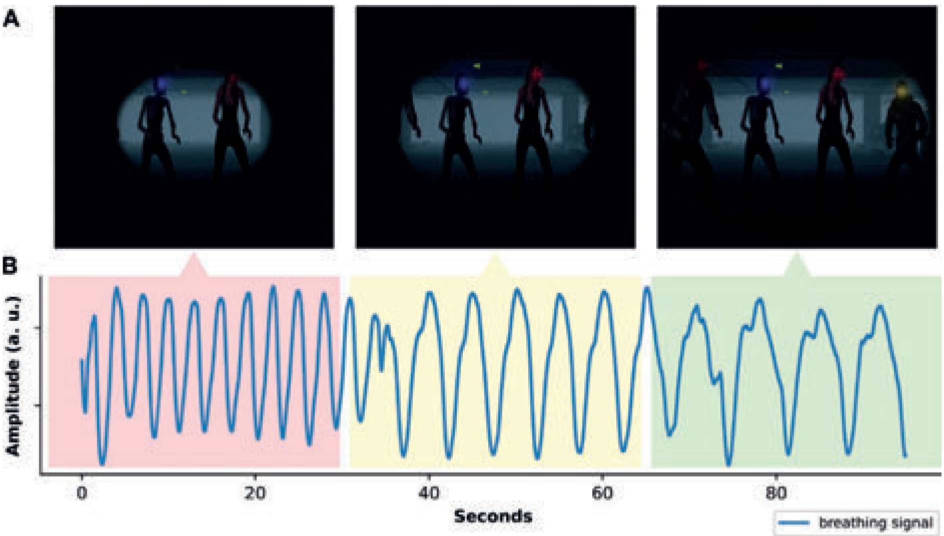
This is particularly salient since the decision task requires the trainee to monitor all 360° of their surroundings, which makes losing peripheral vision costly. The tunnel vision is amplified by modulating the brightness of environmental lights. In general, the biofeedback representation should intuitively fit into the context and training goals of the application to facilitate immersion and engagement. This can lead to fundamentally different representations of the same biofeedback parameter. For example, DEEP, another breathing biofeedback training in VR, teaches the user to leverage their exhalation to propel themselves forward and their inhalation to float upward in a virtual underwater environment (Van Rooij et al., 2016).

Moreover, to ease multi-tasking, the biofeedback representation should only interfere with game play if this is intended (such as the tunnel vision), not for ergonomic reasons such as graphs that are placed inconveniently in the trainee's field of vision. In the same vein, we chose to not include a more explicit biofeedback representation such as the commonly used statistical graphs (Sun et al., 2017) to avoid burdening the trainee with monitoring yet another element in the environment.

Another crucial element of the biofeedback representation is its stepwise introduction. Before the trainee enters our multi- tasking environment, they are guided through a breathing tutorial that gradually introduces them to the regulation skill. The trainee starts with a breathing exercise that requires them to breathe along with a visual pacer (Supplementary Material 3 and Supplementary Figure 3.6A). Once they feel comfortable with the breathing skill, we demonstrate the effects of the breathing regulation. This demonstration includes an explicit component (bar graph) providing clear feedback on the current physiological state as well as the more implicit environmental effects described earlier (tunnel vision and environmental lights) (Supplementary Material 3 and Supplementary Figure 3.6B). Note that the latter are implicit only in terms of the concrete representation, not in terms of saliency. Once the trainee has a good understanding of how their breathing affects the environment, we remove the bar graph and they practice the regulation skill in a simplified version of the decision task (Supplementary Material 3 and Supplementary Figure 3.6C) before entering the full-fledged training. The gradual introduction of the regulation skill in progressively more challenging contexts avoids overwhelming the trainee with the demands of multi-tasking and is believed to facilitate skill transfer (Driskell et al., 2001; Driskell & Johnston, 1998).

Lastly, in designing the biofeedback representation, it is helpful to know the behavior of the biofeedback parameter as early as possible. This includes being familiar with its temporal dynamics as well as extreme patterns. Regarding the latter, it is useful to account for the possibility that trainees "get stuck" in a physiological state and consequently struggle to meet the biofeedback target. In this scenario, it is important for the biofeedback representation to be designed such that the trainee can still function in the environment. For example, in our environment the trainee always retains a minimum of visibility even with the worst biofeedback score (see Figure 3.4).

Figure 3.4
Biofeedback representation and its relation to breathing



Note. The peripheral vision in the VR headset (A) responds to the user's breathing rate. (B) Shows the raw breathing signal of a user who transitions from a fast breathing rate to a slower breathing rate that matches the biofeedback target more closely

DISCUSSION

Designing and implementing a stress-exposure biofeedback training requires a re-thinking of conventional biofeedback training. This introduces challenges around (1) the choice of a biofeedback parameter, (2) the biofeedback processing, and (3) the representation of the biofeedback. We examined these challenges from both a technical as well as a user- experience perspective and illustrated the feasibility of stress- exposure biofeedback with examples from a breathing-based stress regulation training for police.

We highlighted the importance of controllability of the biofeedback parameter as well as the attainability of the biofeedback target. Additionally, we showed how seemingly small algorithmic decisions during the real-time computation of the biofeedback can have far-reaching consequences for the user experience, and emphasized the importance to arrive at these decisions during iterative user testing. Finally, we point out the relevance of a salient, intuitive biofeedback representation that is introduced gradually, and tailored to the task context and goals.

In demonstrating the feasibility of stress-exposure biofeedback, we hope to advance this biofeedback paradigm and to help pave the way for studies that explore its potential to diminish the short- and long- term consequences of repeated stress-exposure. Above all, we hope this paper and its Supplementary Material provide a useful reference for developers and researchers in academia or industry who are interested in using physiological signals to control elements in a dynamic virtual environment.

Ethics Statement

The studies involving human participants were reviewed and approved by the Ethics Committee Faculty of Social

Sciences, Radboud University, Nijmegen, Netherlands. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Authors Contributions

JB wrote the manuscript, conceptualized and implemented the biofeedback, and analyzed the pilot data. RO conceptualized and implemented the IT infrastructure (software and hardware) for the biofeedback. RO and FK supervised the implementation of the biofeedback. JB, JP, MR, FK, AM, IG, KR, and WD conceptualized the VR environment and decision task. WD coordinated the pilot data collection. AM planned and conducted the pilot data collection. All authors reviewed and contributed to the manuscript.

Funding

This work was supported by the NWO Creative Industry grant # 055.16.139 2795 awarded to KR, IG, and FK.

Acknowledgements

We would like to thank Ken Koontz (https://www.kenkoontz. com/), Erik van den Berge (https://twitter.com/followerikvdb), and Thijs Doornbos (https://tiohoan. wordpress.com/) for their dedication in designing and implementing the VR, as well as the numerous police officers who generously volunteered their time to test the training throughout its development. We would also like to thank the originators and maintainers of Python (Van Rossum & Drake, 2009) Matplotlib (Hunter, 2007), PyQtGraph (PyQtGraph, 2020), PyQt (Riverbank-Computing, 2020), SciPy (Virtanen et al., 2020), NumPy (Van Der Walt et al., 2011), Spectrum (Cokelaer & Hasch, 2017), Seaborn (Waskom et al., 2020), dabest (Ho et al., 2019), and EEGsynth (SciCrunch, 2020), whose open-source software was used extensively throughout this project. The Lego minifigure icon used in Figure 3.1 and Supplementary Figure 3.1 was created by Okan Benn for the Noun Project and is distributed under a Creative Commons License.

Supplementary Materials

The Supplementary Material for this article can be found online at: https://www. frontiersin.org/articles/10.3389/fpsyg. 2021.586553/full#supplementary-material

SUPPLEMENTARY MATERIAL

1 Software and Hardware

We use a respiratory inductance plethysmography (RIP) belt[1] (Figure S3.1A) to acquire the analog breathing signal which is converted to a digital signal on a BITalino (r) evolution board[2] (Figure S1B). To obtain the biofeedback score, we then process the digital signal on a Raspberry Pi 4 Model B[3] (Figure S3.1C) with custom Python software using the open source EEGsynth library[4] . The biofeedback implementation can be found in the "biochill" module[5] of the EEGsynth library. See supplementary material B for details on the signal processing. The biofeedback score is then published to a Redis channel[6] (Figure S3.1D) that is monitored by a Unity VR engine[7] (Figure S3.1E) which renders the biofeedback representation in the virtual environment accordingly. Eventually, the virtual environment is presented using a HTC Vive headset[8] (Figure S3.1F). We would also like to point the reader to the Biocybernetic Loop Engine[9] as well as the Excite-O-Meter[10] as two additional software tools for integrating physiological measurements into a VR environment.

[1] https://www.biosignalsplux.com/products/sensors/respiration-inductive.html

[2] https://bitalino.com/en/plugged-kit-bt

[3] https://www.raspberrypi.org/products/raspberry-pi-4-model-b/

[4] https://scicrunch.org/resources/about/registry/SCR_018732

[5] https://github.com/eegsynth/eegsynth/master/module/biochill/biochill.py

[6] https://redis.io/

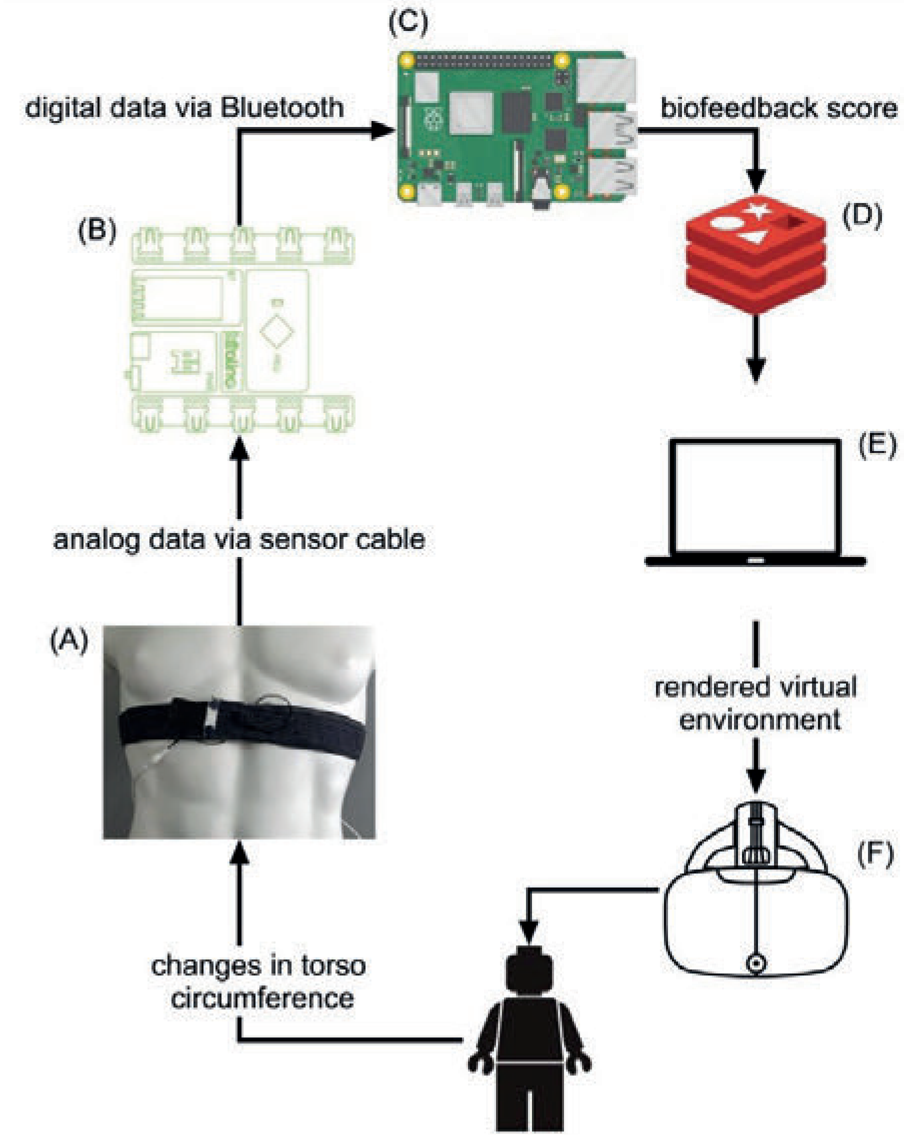
[7] https://unity.com/unity/features/vr

[8] https://www.vive.com/eu/product/#vive%20series

[9] https://doi.org/10.5220/0006429800450054

[10] https://sites.google.com/view/exciteometer/

Figure S3.1
Components of the real-time biofeedback loop.



3

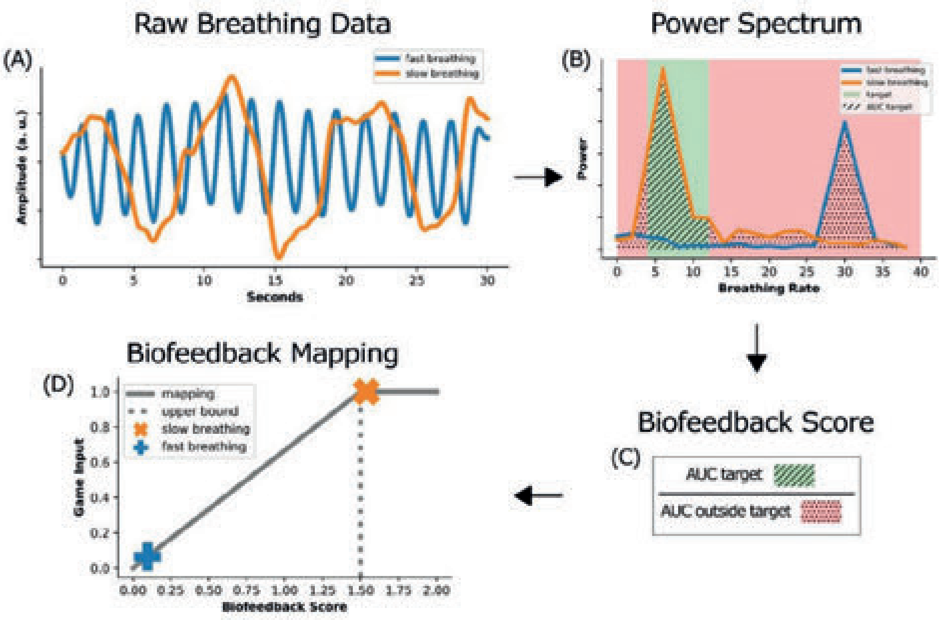
2 Signal Processing

In our real-time signal processing loop, we process 30 seconds of breathing data every two seconds (i.e., shifting window of 30 seconds with a hop size of 2 seconds and a resulting overlap of 28 seconds). We use spectral density estimation (SDE) to approximate the current breathing rate from each new breathing segment.

SDE indicates how much of a certain frequency can be found in a signal. Breathing frequency is the number of breaths per second, similarly to breathing rate, the number of breaths per minute. A plot of spectral density (aka power spectrum) has power ("how much") on the y-axis and frequency ("how fast") on the x-axis.

Figure S3.2

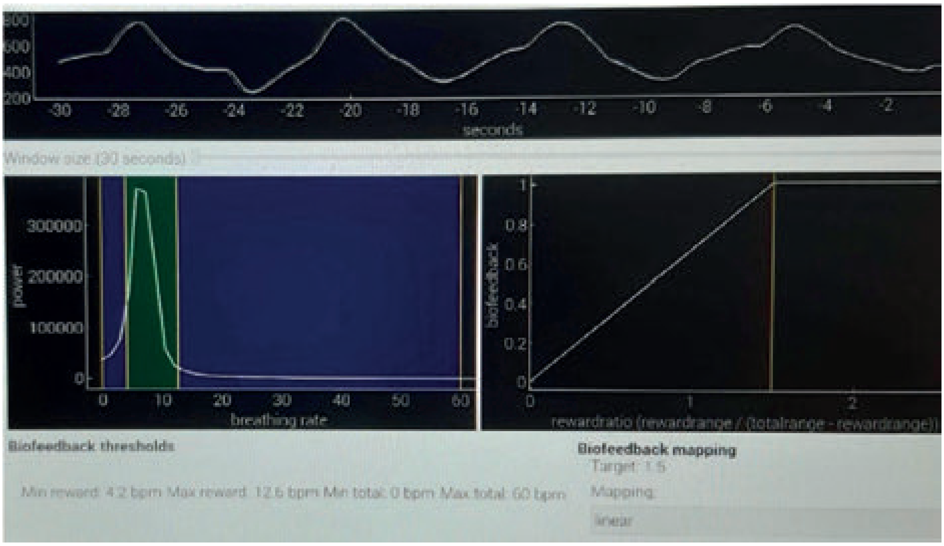
Sequence of biofeedback processing steps from raw data to game input.



Note. The blue and orange elements represent two breathing signals that are moved through the processing steps. The blue elements correspond to a signal with a fast breathing rate, the orange elements pertain to a signal with a slow breathing rate.

Figure S3.2B shows the power spectra of the two breathing signals in Figure S3.2A. The orange breathing signal contains a breathing rate of 6 breaths per minute whereas the blue breathing signal contains a breathing rate of 30 breaths per minute. We use the power spectrum to evaluate how much the current breathing rate matches the biofeedback target (range between 4 and 12 breaths per minute, green area in Figure S3.2B). We do so by integrating the area under the curve (AUC) of the power spectrum over the interval of the biofeedback target (striped area in Figure S3.2B). To obtain the biofeedback score, we divide this area by the remaining area under the power curve (dotted area in Figure S3.2B), which spans zero to 40 breaths per minute (red area in Figure S3.2B). In other words, the biofeedback score is the ratio of a) spectral density over the biofeedback target range relative to b) the spectral density outside of the target range (Figure S3.2C). Since the biofeedback score is used to control elements of the virtual training environment (game input, as discussed in more detail in Challenge 3 of the main article), it is convenient to have the score be in a standardized range. We map the biofeedback score to a game input value ranging from 0 (worst) to 1 (best) (Figure S3.2D). There are two decisions regarding this mapping, which have important consequences for the trainee's experience of the biofeedback. First, we needed to define an upper bound on the ratio that represents the biofeedback score, and second, a transformation function has to be chosen. We chose an upper bound of 1.5 (dashed line in Figure S3.2D), which means that the trainee obtains the most rewarding biofeedback (game input = 1) when the spectral density within the biofeedback target range is at least 1.5 times larger than the spectral density outside of the target range. When the upper bound is set higher, it is more difficult for the trainee to obtain the most rewarding feedback.

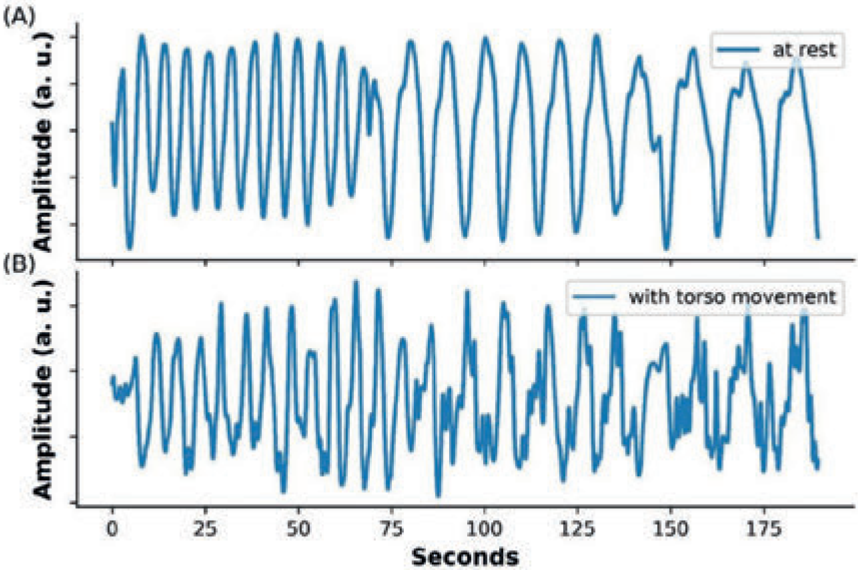
Figure S3.3
Biofeedback dashboard.



Note. The dashboard allows us to monitor the breathing signal and adjust a number of biofeedback processing parameters in real-time: 1. The length of the breathing signal that is processed every two seconds (upper panel). 2. The upper and lower breathing rate of the biofeedback target range and total range (lower left panel), 3. The kind of biofeedback mapping from biofeedback score to game input as well as the upper bound on the biofeedback score (lower right panel). The ability to adjust these parameters in real-time accelerates the process of homing in on a suitable biofeedback configuration during user testing.

Ideally the upper bound would be adjusted to the current skill level of the trainee (e.g., raise the upper bound as the trainee gets more skilled at controlling their breathing, to keep the training challenging). For the transformation function we chose a linear mapping of the biofeedback score to the game input (solid line in Figure S3.2D). This means that changes in the biofeedback score (up until the upper bound) will produce equivalent changes in the game input. In contrast, an exponential mapping produces more pronounced changes in the virtual environment at high values of the biofeedback score, and less pronounced changes at medium values. However, we want the biofeedback to be equally salient throughout its entire range.

Figure S3.4
Artifacts in a breathing signal.

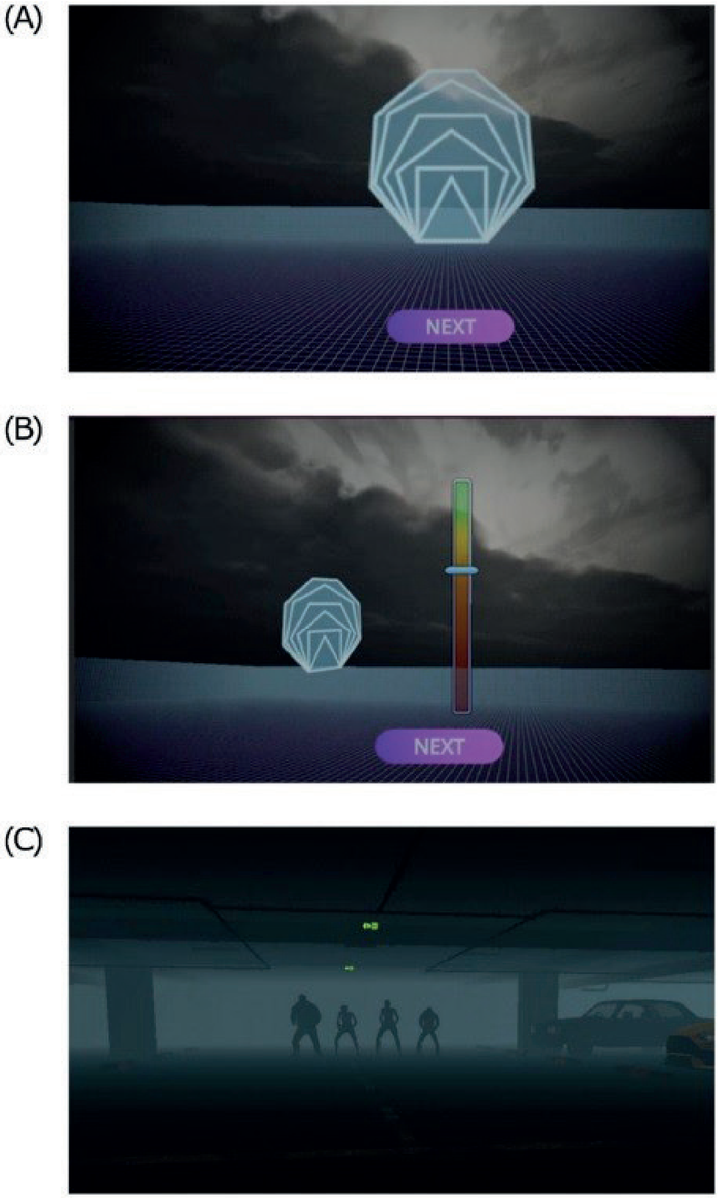


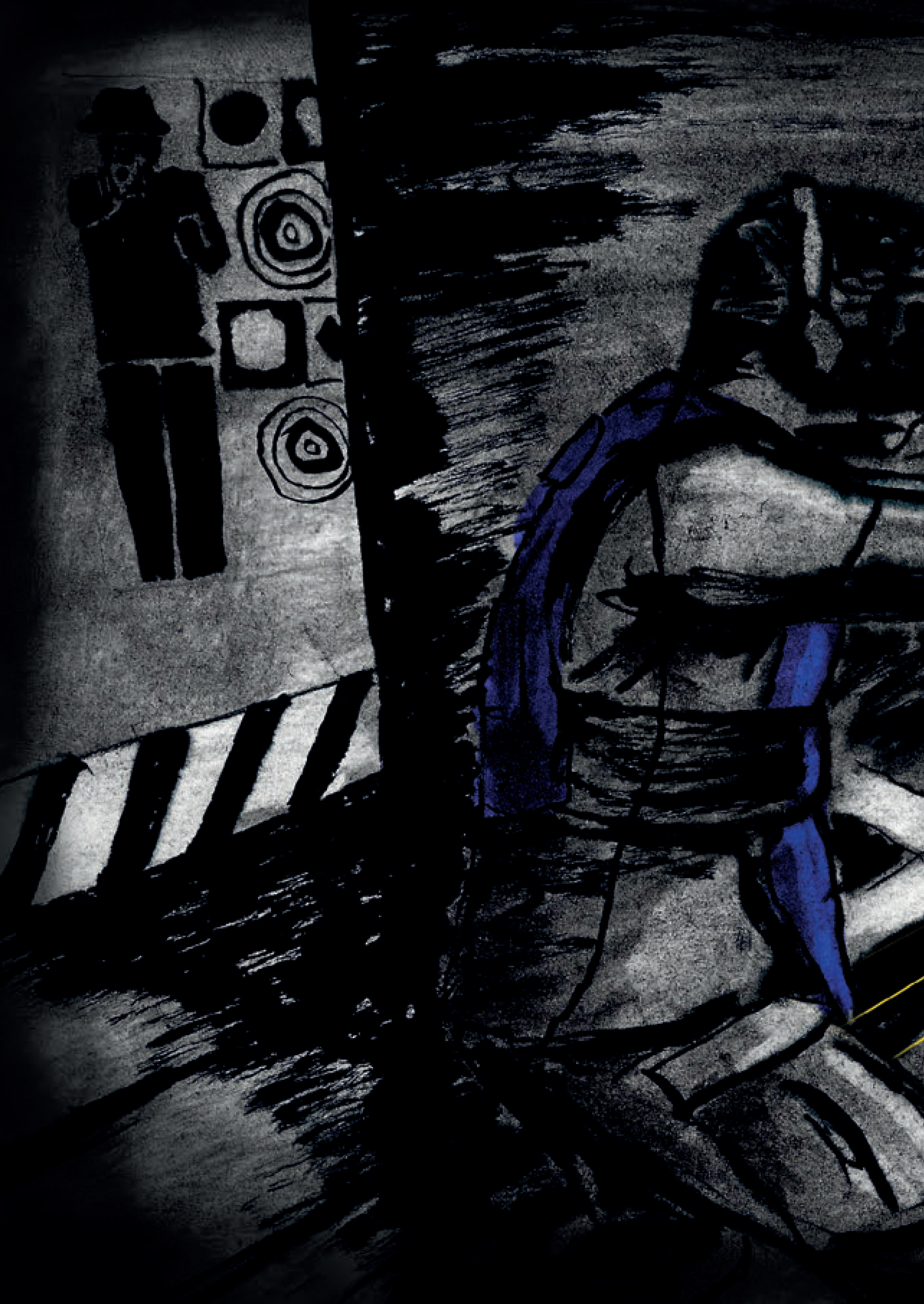
Note. A user is going through a paced breathing exercise, starting with a breathing rate of 10 breaths per minute (bpm), then shifting to 6 bpm, and finally breathing at 4 bpm. The exercise is conducted twice, while the user is standing in a relaxed position (A), and while the user is standing and moving their torso (B). Especially at lower breathing rates, the artifacts induced by the torso movement distort the signal. Note that the artifacts are in the frequency range of fast breathing.

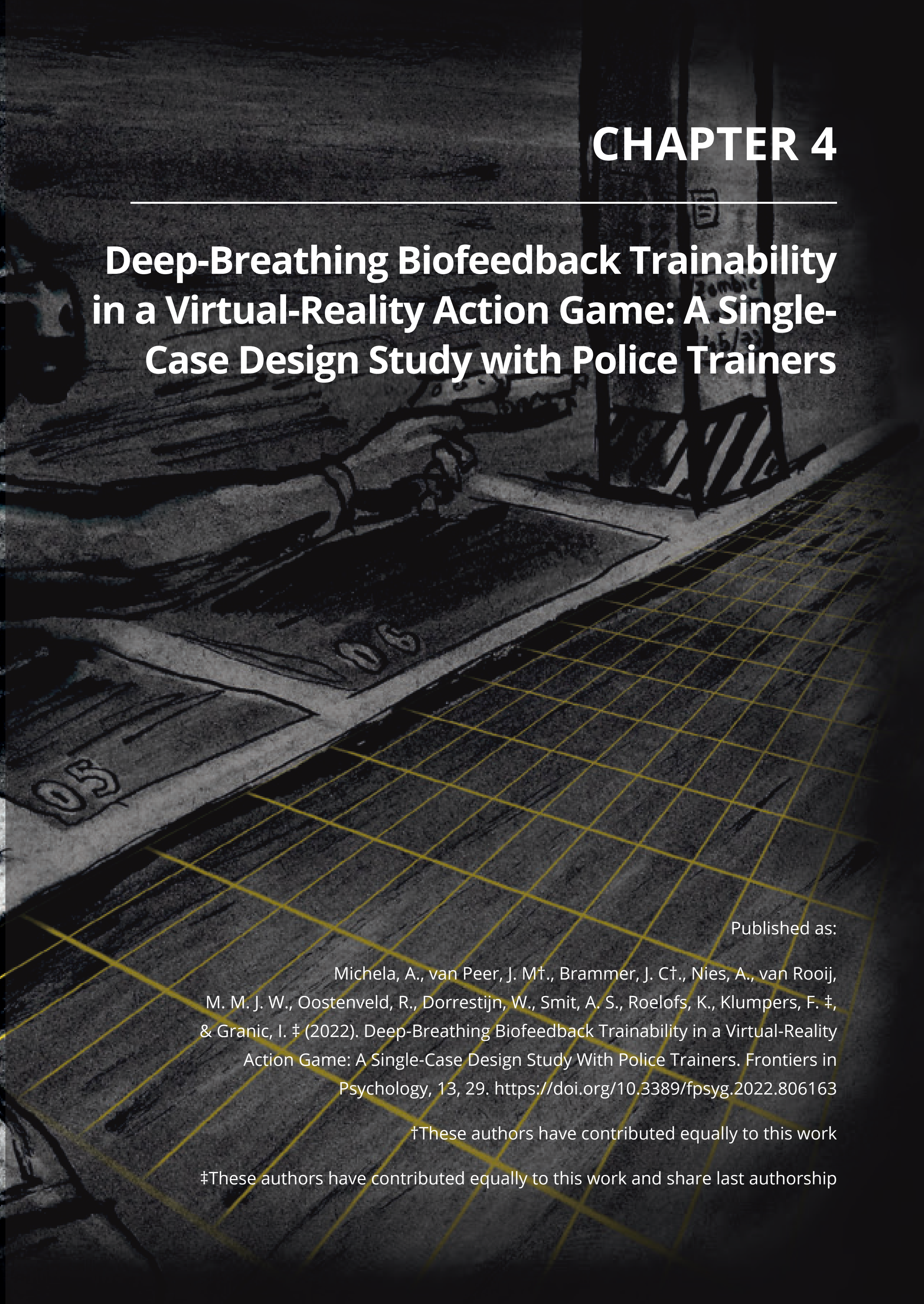
3 Virtual Training Environment

Figure S3.5
Avatar identifiers.

Figure S3.6
Breathing tutorial







CHAPTER 4

Deep-Breathing Biofeedback Trainability in a Virtual-Reality Action Game: A Single-Case Design Study with Police Trainers

Published as:

Michela, A., van Peer, J. M†., Brammer, J. C†., Nies, A., van Rooij, M. M. J. W., Oostenveld, R., Dorrestijn, W., Smit, A. S., Roelofs, K., Klumpers, F. ‡, & Granic, I. ‡ (2022). Deep-Breathing Biofeedback Trainability in a Virtual-Reality Action Game: A Single-Case Design Study With Police Trainers. Frontiers in Psychology, 13, 29. https://doi.org/10.3389/fpsyg.2022.806163

†These authors have contributed equally to this work

‡These authors have contributed equally to this work and share last authorship

ABSTRACT

It is widely recognized that police performance may be hindered by psychophysiological state changes during acute stress. To address the need for awareness and control of these physiological changes, police academies in many countries have implemented Heart-Rate Variability (HRV) biofeedback training. Despite these trainings now being widely delivered in classroom setups, they typically lack the arousing action context needed for successful transfer to the operational field, where officers must apply learned skills, particularly when stress levels rise. The study presented here aimed to address this gap by training physiological control skills in an arousing decision-making context. We developed a Virtual-Reality (VR) breathing-based biofeedback training in which police officers perform deep and slow diaphragmatic breathing in an engaging game-like action context. This VR game consisted of a selective shoot/don't shoot game designed to assess response inhibition, an impaired capacity in high arousal situations. Biofeedback was provided based on adherence to a slow breathing pace: the slower and deeper the breathing, the less constrained peripheral vision became, facilitating accurate responses to the in-game demands. A total of nine male police trainers completed 10 sessions over a 4-week period as part of a single-case experimental ABAB study-design (i.e., alternating sessions with and without biofeedback). Results showed that eight out of nine participants showed improved breathing control in action, with a positive effect on breathing-induced low frequency HRV, while also improving their in-game behavioral performance. Critically, the breathing-based skill learning transferred to subsequent sessions in which biofeedback was not presented. Importantly, all participants remained highly engaged throughout the training. Altogether, our study showed that our VR environment can be used to train breathing regulation in an arousing and active decision-making context.

Keywords: Police education, police training, Virtual-Reality, stress management, autonomic arousal, heart rate variability, biofeedback.

INTRODUCTION

The work of a police officer can be seen as an evolutionary paradox: in places and situations where most people would fall prey to survival instincts of self-preservation, police officers ought to act calm, with proportionality and benevolence. As a result, working as a police officer can elicit considerable stress, with high prevalence of work-related physical injuries (West et al., 2017) and stress-related symptoms (Leppma et al., 2017) which can lead to mental disorder (Carleton et al., 2018) and suicidal ideation (Di Nota et al., 2020). Additionally, police officers are under constant scrutiny from both internal and external sources, including the public display of mistakes to ensure accountability (Skolnick & McCoy, 1984). The nature of their dangerous work as well as being under ongoing surveillance puts a great deal of pressure on police officers.

A large body of literature has shown that, besides affecting physical and mental wellbeing (Stetz et al., 2007), stress can impair decision-making aspects which are crucial to the policing job. Experimental studies have consistently shown that under high levels of arousal, decision-making becomes more impulsive and less goal-directed (Porcelli & Delgado, 2017), which is related to impaired control by prefrontal brain regions under stress (Arnsten, 2015; Maier et al., 2015). The detrimental effects of arousal have also been demonstrated in police, where high arousal has been associated with impaired shooting accuracy, as well as decreased inhibitory control (Hashemi et al., 2019; Nieuwenhuys et al., 2015). It is essential that police officers manage their stress while on the job to avoid costly mistakes such as the inappropriate use of force.

To train coping under stress in police officers, recent studies investigated the training of behavioral skills under pressure, which improved basic perceptuomotor skills such as shooting accuracy under threat (Nieuwenhuys & Oudejans, 2011) or spatial orientation (Driskell et al., 2001). However, this type of training may not always improve relevant higher-level decision-making under stress, such as when asked to make shoot/don't shoot decisions (Nieuwenhuys et al., 2015). To counter this stress-induced performance drop and improve decision-making under threat, training efforts have moved beyond the focus on performance, by adding in police education elements to help officers directly manipulate the bodily stress response itself (Andersen & Gustafsberg, 2016; Bouchard, Bernier, et al., 2012; Mccraty & Atkinson, 2012). In some cases, this opportunity for physiological modulation was even delivered intermixed with action in realistic scenario-based environments (Andersen et al., 2018; Andersen & Gustafsberg, 2016; Di Nota et al., 2021).This type of biofeedback training mainly focuses on promoting emotional regulation by helping the officer control their physiological arousal by giving feedback on their objective physical stress level (e.g. indexed by heart rate parameters, see below) to improve physiological awareness and gain more voluntary control over their physiology (Weerdmeester et al., 2020).

After studies showed the usefulness of biofeedback to enhance stress regulation (Bouchard, Guitard, et al., 2012; McCraty et al., 2009) several police forces worldwide have implemented training programs in which a biofeedback training component is included. However, according to a recent survey in the Netherlands, this type of training was generally negatively appraised by police officers and did not result in substantial improvements in reported stress regulation (van der Meulen et al., 2018). These disappointing results might be attributable to the fact that in current practice, biofeedback is often delivered in a passive classroom setup, which may not feel relevant to police officers and thereby affect their engagement. Yet, engagement is a key prerequisite for behavioral change (Holzinger et al., 2006). To promote both engagement and generalization to real-life stressful situations (skill-transfer), it is important to train stress regulation within a progressively more active and thereby more representative action context (Seifert et al., 2018; Staller & Körner, 2019). Indeed, just as the current biofeedback applications often lack elements to make the learning process engaging and relatable, excessive stress might also prevent learning (Di Nota & Huhta, 2019). To better reappraise stress into challenge and to improve performance under stress (Jamieson et al., 2010), it seems necessary to design a training context in which stress is kept at a moderate, optimal level that is not too low, but also not overwhelming and excessive.

One way to address previous shortcomings in classroom-based training is by creating engaging environments in Virtual-Reality (VR) where physiological control must be exerted while performing decision-making actions in an arousing context. VR is a useful tool to create controlled – yet representative – environments, allowing to elicit high levels of arousal and engagement (L. C. Miller et al., 2019; Parsons, 2015). However, re-creating a highly realistic virtual police environment in VR can also have negative consequences on subjects' experiences due to slight mismatches with reality (Wilson & Soranzo, 2015) and the challenges in recreating genuine verbal and tactile interactions (Michela et al., 2019). To remedy this, the use of game mechanics (a set of rules and events defining the game experience) inspired by commercial videogames offers the possibility to re-create a genuine feeling of threat and immersion, moderate enough for learning to take place while at the same time boosting engagement (Allcoat et al., 2015; Cummings & Bailenson, 2016; Lin, 2017; Schoneveld et al., 2019; Slater, 2018). Game mechanics also present another advantage over realistic environments in VR, which is the ease of emotional elicitation and repeatability of the experience (Lobel et al., 2016; Michela et al., 2019;

Scholten & Granic, 2019). Ultimately, gaming contexts also alleviate the emotional impact that poor performance could have on feelings of professional self-efficacy, reduces the chance of overtraining motor responses that may be less adequate in real-life and limits reactivation of potentially traumatic experiences, since the actual action is rather far removed from real policing contexts (Michela et al., 2019). Evidence-based VR trainings are however currently scarcely available (Di Nota & Huhta, 2019) and to the best of our knowledge, no training has been described that makes use of the aforementioned VR and gaming assets. Therefore, we developed a VR environment that offers the player the possibility to train breathing-based biofeedback skills in real-time while immersed in an engaging active decision-making game (Brammer et al., 2021).

For the physiological training component, our VR application aligns with previous training practices in Dutch Police. Specifically, biofeedback training applied to police officers has mainly focused on abdominal deep breathing (Bennell et al., 2021; van der Meulen et al., 2018) , which has been shown to increase heart-rate variability (HRV) by respiratory sinus arrythmia (RSA; Hirsch & Bishop, 1981) reflecting parasympathetic nervous system dominance (Russo et al., 2017). The influence of deep breathing on the parasympathetic nervous system can be measured in both the low frequency and high frequency spectrum of HRV (Kromenacker et al., 2018; Shaffer & Ginsberg, 2017), as well as the coherence between breathing and HRV (Schwerdtfeger et al., 2020; Shaffer et al., 2014). This type of biofeedback has proven useful in a variety of training applications, from performance training for athletes (Jiménez Morgan & Molina Mora, 2017) to stress and anxiety management (Goessl et al., 2017). Importantly for police applications, an elevated HRV is also related to better performance under threat (Hansen et al., 2009), and was shown to be effectively modulated by biofeedback, hence reducing stress in contexts related to police realities (Andersen et al., 2015, 2018; Andersen & Gustafsberg, 2016; Bouchard, Bernier, et al., 2012) and also improving cognitive control (Laborde et al., 2021). In our VR environment, biofeedback was provided by modulating the width of the field of view according to breathing rate and depth, thus making real-time physiological regulation key to being able to perform well in the game.

For the behavioral assessment components during the training in our VR environment, we focused on one of the key decision-making processes that is affected by stress and related to police-relevant go/nogo decisions, namely shoot/don’t shoot decisions (Gladwin et al., 2016; Nieuwenhuys et al., 2015). Impairments in response inhibition are known to be especially high when individuals are primed to believe that they will need to take action and the stakes are high (D. J. Johnson et al., 2018; Taylor, 2020). Thus, we designed our game mechanic around these processes of response inhibition under stress and the impact of priming. Having those metrics imbedded in the same environment as the physiological training allows for direct measurement of the impact that physiological training has on performance.

The first overarching goal of the current study was design validation. First, we tested if the VR environment was, as hypothesized, successful in creating a challenging environment that elicits clear increases in levels of arousal (assessed via heart rate) and self-reported engagement. Second, we tested if our biofeedback a) increases slow and deep breathing and HRV, and b) raises physiological awareness for the participants. Third we explored whether our setup allowed extraction of meaningful behavioral metrics concerning response inhibition and priming. Behavioral measurements were all extracted from the VR environment, with the auxiliary aim of documenting interactions between behavioral metrices and physiology.

The second main goal of the current study was to perform a preliminary evaluation of the game's potential to train breathing-based biofeedback in an active decision-making context. We hypothesized that breathing biofeedback score would improve over the training, along with HRV. Moreover, biofeedback-driven physiological regulation skills should transfer to the same action context, when experienced without biofeedback. This preliminary proof-of-concept was performed by means of a withdrawal single-case experimental design (SCED), applied to a sample of nine trainers from the Dutch police, a difficult population to get access to and test due to their usual work load, yet highly valuable given that they contain all the critical insight in both the required skills for dealing with stress and the challenges surrounding teaching those skills to police recruits. These trainers took part in a ten-session training program. Withdrawal SCED has already been successfully used in investigating the potential of biofeedback (e.g., Bossenbroek et al., 2020). This design has the advantage of providing rich datasets to investigate the dynamical evolution of skill acquisition within and across sessions (P. L. Smith & Little, 2018), and inform future research about the often overlooked aspect of minimal training length required for efficient training (Di Nota et al., 2021).

METHODS

Participants

Participants were nine male police trainers with an average age of 43.2 years (*SD* = 6.45) and with an average of 18.4 years (*SD* = 8.6) of operational background as a police officer. Their average trait anxiety was 27 (range 23 to 34 on a scale of 20 to 80), which indicates participants in our sample to be non-anxious. Only three participants reported playing video games in their free time, to a maximum of 4 hours a day during weekends. Participants were recruited from a Police skill training center, in the Netherlands, hereafter referred to as IBT center ('Integraal Beroepsvaardigheids Training centrum'). Given that this was a proof-of-concept study, the number of participants was based on earlier studies using similar SCED methodologies (Bossenbroek et al., 2020; Ebbinghaus, 2005; J. D. Smith, 2012; P. L. Smith & Little, 2018). Participation was voluntary and handled by the coordinator at the IBT center. According to the rules of the Dutch Police, financial compensation of the police officers functioning as participants was not allowed. Therefore, for each participant a donation of 50 euros was allocated to a fund for the training of "PTSD dogs" (Actie ZeeHond, n.d.). The research procedures were approved by the ethical committee of the Faculty of Social Sciences of Radboud University Nijmegen (ECSW-2020-112). All participants provided informed consent in writing prior to participating in the study, in line with the guidelines of the Declaration of Helsinki (WMA, 2018).

Materials

Physiological recordings

Participants' breathing rate was measured using a respiratory inductance plethysmography (RIP) belt from Plux S.A. and a BITalino (r)evolution board (Batista et al., 2017). The heart rate of the participant was recorded by a Polar H10 chest strap, which extracts R – R intervals (i.e., the time between consecutive R-waves of the QRS electro cardiac signal). Both physiological recording units broadcasted their data to a Raspberry Pi 4 Model B, which was also used to calculate the breathing biofeedback scores with custom-made python software, based on the open source EEGsynth library (Brammer et al., 2021; Oostenveld, n.d.).

Virtual-Reality Material, Model, and Task

An HTC Vive setup was used to immerse the participants in a virtual environment that consisted of a poorly lit underground parking garage. The VR trackers were set 3 to 4 meters apart, giving the player a minimal play area of 4 square meters. One of the two VR controllers was wrapped by a 3D printed case, giving the controller the weight and shape of a gun. The second VR controller was attached to the vest of the participant and used as a dispatch-radio in the game.

Operationalization The VR environment was designed to incorporate game mechanics based on existing experimental laboratory tasks, as shown in Table 4.1. The paradigms listed were incorporated as they represent specific behavioral aspects known to be affected by stress and relevant in the decision-making processes. The first paradigm is emotional regulation, in our case physiologically influenced through breathing-based biofeedback presentation. Its implementation as a modulation of the width of the field of vision makes the biofeedback relevant for the player, since in active contexts it is easy to ignore the biofeedback if it does not interfere with action. Modulating the peripheral view directly impacts the difficulty of detecting approaching enemies in the VR environment.

Table 4.1

Operationalization of in-game tasks.

Experimental paradigm	In-game operationalization	In-game mechanic	Game output / proximal measure	Training outcome
Emotion regulation	Biofeedback	Width of the field of view linked to breathing pace	Breathing control and HRV scores	Physiological control
Response inhibition (Go/NoGo)	Shoot/don't shoot	Shooting at targets matching dispatch information	Accuracy	Response control
Priming	Dispatch bias task	Targets' (mis)match with dispatch information (body shape; eye color)	Accuracy based on body type of the target	Bias resistance

Note. Model for the incorporation of experimental paradigms in the game environment; HRV = Heart-rate variability.

The second and third paradigms were selected based on existing literature investigating police performance in decision making contexts. Specifically, response inhibition under stress, operationalized as go/nogo (shoot/don't shoot) decision-making , has been linked to increased error-rates with increased threat level (Hashemi et al., 2019, 2021; Nieuwenhuys et al., 2015) Similarly, priming has been shown to increase wrong shooting decisions in police officers primed with a radio dispatch indicating that an upcoming opponent was armed (Taylor, 2020). These two paradigms were implemented in the game in the form of a zombie shooting task (see below).

The Virtual Reality Game The game scenario unfolded as follows: Participants found themselves in a dimly lit parking garage, received dispatch information about hostile targets with a description of their features, were approached by (friendly and hostile) targets, and decided whether to shoot or not. Each game session lasted approximately 15 minutes. Targets were coming towards the participants from all directions, in fourteen waves. During each wave, the participant would receive dispatch information over the walkie-talkie to shoot the hostile but not the friendly targets, including a description of the hostile targets. All targets had two identifying features that made them recognizable as friendly or hostile. The *large identifier* was their body type (tall male, small male, tall female, or small female). During the game this identifier became less reliable to identify a target, as hostile targets had increasingly more varying body types (starting with 100% targets matching the description and decreasing to 50% in the last waves). The *small identifier* was their eye color (red, blue, or yellow); this was always 100% reliable. To summarize: The dispatch information for the eye color of hostile targets would always be correct, but the dispatch information for their body types was not completely accurate. The body type was visible from a far distance, while eye color was only visible when the target had approached to a close distance. Hence, identifying targets with the correct eye color and shooting at them on time was our implementation of the go/nogo task component, while the radio dispatch announcing suspected body types associated with the targeted eye-color was our implementation of the priming component.

Each time the player shot, the game recorded if the shot was a hit or a miss. In case of a hit (i.e., hostile target with correct eye color), the game logged the body type of the target and the distance between the target and the player. Shooting a friendly target (i.e., wrong eye color) was punished by a loud burst of noise. Each time a target reached the player without being shot, the game logged the body type of the target, whether it was hostile and the time that it had spent in the line of sight of the player. Hostile targets would then stay next to the player and attack them until shot. The player was notified of the attack by sound, and a red halo appeared framing their field of view. Players could not “lose” the game in the sense that they reached game-over, the hostile targets had to be all shot before the next wave would start. Please note, in contrast to more traditional laboratory assessments of go/nogo performance, different trial types were here presented at the same time (i.e., multiple targets approached the player at the same time). Hence, time pressure was created by the fact that multiple choices had to be made concomitantly, to simulate a short response window, required to maximize false alarm rates (Young et al., 2018). The maximization of false alarms was needed to provide enough improvement margin to the player and benefited from the non-realistic environment to mitigate risks in terms of professional self-efficacy in players.

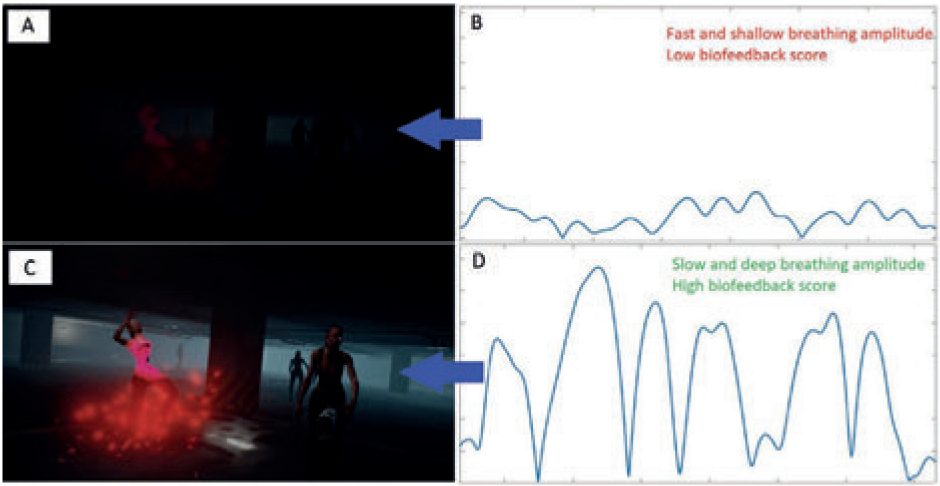
At the end of each session, the participant was presented with scores ranging from 0% to 100%, summarizing their performance on three metrics commonly used to describe police performance in the field: Control over the situation, control over the suspect and control over the self (Binder & Scharf, 1980; Huhta et al., 2021); see supplementary material 1 for the formulas used to calculate these scores). While the first two scores represented behavioral elements, the last one was a summary of the breathing pace performance rewarded by the biofeedback. The scores were calculated to make the scores of "control of the situation" and "control of the suspect" relatively easy: Participants could achieve high scores without excessive effort. On the other hand, the "control of the self" score directly represented the physiological control score, thus helping to nudge the players toward focusing on physiological control more in the following sessions.

Biofeedback parameter and implementation

In sessions in which the biofeedback was displayed to the player (ABBABABABA withdrawal design; A = without biofeedback: session 1, 4, 6, 8, and 10; B = with biofeedback: session 2, 3, 5, 7, and 9), a breathing pace of eight breaths per minute was rewarded by having optimal vision, with faster or slower breathing paces being progressively punished by reduced (tunneled) vision (see Figure 4.1).

Figure 4.1

Implementation of the biofeedback as peripheral vision modulation.



Note. The quality of the vision (on the left) was linked to the biofeedback score, calculated from the breathing speed and amplitude (on the right); (A) Vision with a biofeedback score equal to 0 (the player is not breathing at the rewarded pace); (B) example of fast and shallow breathing amplitude, corresponding to a breathing biofeedback score of 0; (C) Vision with a biofeedback score equal to 1 (maximum adherence to rewarded breathing pace) or in A-phase sessions in which online biofeedback was not presented to the player; (D) example of slow and deep breathing amplitude, corresponding to a breathing biofeedback score of 1.

The visual feedback on participants' breathing pace was implemented by reducing the vision of the player in the VR task proportionally to their non-adherence to the rewarded breathing pace. This "tunneled" vision served to help the player control their breathing pace in the heat of the moment, when task demand reduced awareness of their own physiological processes. Specifically, the biofeedback value rewarded slow and deep breathing by promoting high amplitudes of breathing paces in the 6 to 10 breath per minute range. It was calculated by performing spectral density estimation on the incoming breathing signal and then calculating the power of the signal within the target frequency range divided by the power outside the target frequency range (c.f. Brammer et al., 2021; Figure S3.2). Please note that the power was calculated as area under the curve, thus reflecting the amplitude (depth of breathing) within the target range. This value was updated at a rate of .5Hz, and ranges between 0 and 1. Hence, every 2 seconds, a breathing segment of the last 30 seconds of breathing data was analyzed, resulting in an overlap of 28 seconds between consecutive segments. It rewarded diaphragmatic breathing at a pace of eight breaths per minute. This target is faster than the six breaths per minute pace recommended by the literature to increase HRV (Ben-Tal et al., 2014; T. E. Brown et al., 1993; Hirsch & Bishop, 1981), as piloting revealed that such a low pace was too hard to achieve in action. The exact signal processing pipeline and data extraction can be found in the supplementary material of our previous article, detailing the development of the biofeedback parameter (Brammer et al., 2021). Importantly, the visual presentation of the biofeedback could be switched off, in which case the player would always have a full vision, independently from their breathing biofeedback score. This parameter was experimentally manipulated between training sessions to test the impact of biofeedback presentation (see procedure section).

Questionnaires

All questionnaires were administered in a paper and pencil form. The following questionnaire data were collected.

Trait-assessments of anxiety. Participants' trait anxiety was assessed in the first session using the Dutch State-Trait Anxiety Inventory with twenty items and four response options (1 = "almost never" to 4 = "almost always"; van der Ploeg, 1984). An example of a statement is "I feel satisfied with myself". A total score was calculated by calculating the sum of all item scores (range 20-80), with higher scores corresponding to less anxiety.

Prior gaming experience. Participants' prior gaming experience was assessed in the first session using two self-constructed questions. Participants were asked how many hours a day they played video games on a weekday and how many hours during a day on the weekend. Response options were: 1) I do not play video games; 2) less than one hour per day; 3) one to two hours a day; 4) two to three hours a day; 5) three to four hours a day; 6) more than four hours a day.

Pre-test questionnaire. The pre-test questionnaire (target approach analysis, see supplementary material 2) is a short self-constructed questionnaire inspired by questionnaires used to investigate plan-making in real-life policing situations (Adang & Timmer, 2005), administered before each VR session to the players. They were asked to indicate which one of the three control scores mentioned above they would focus on, what scores they expected to achieve as well as the scores they expected their colleagues to achieve. An open question concluded the questionnaire, to ask the player if they had a specific strategy in mind for the upcoming session. The aim of this questionnaire was twofold: Measuring the participants' training intentions and prime them to keep in mind the policing goals of the training endeavor.

Post-test questionnaire. After each session, the post-test questionnaire (after action report, see supplementary material 2), a short self-constructed questionnaire, was administered to the participants. The players were asked to indicate which score they thought they had achieved, on a scale from 0% to 100%, for each one of the 3 control scores (control of the situation, of the suspect and of self) and several open-ended questions to make them think about their performance. The aim of this questionnaire was twofold: Measuring the participants' performance estimates, and maintain awareness of the policing goals of the training endeavor. Of the three self-rated control scores, only the "control of the self" score was used for further analyses, where it was contrasted to the actual "control of the self" score achieved by the participant.

Threat and challenge appraisal. Participants' appraisal of threat and challenge during the game was assessed after each session on an eleven-item scale developed by Mendes et al. (2007), with seven response options per item (1= "totally disagree" to 7 = "totally agree). Six items were indicative of the threat aspect (i.e., "this task is demanding," "... is stressful," "... is distressing," "... is threatening"; "I am uncertain how I will perform"; "this task requires a lot of effort"). In addition, five items were indicative of perceived positive challenge (i.e., "I have the abilities to perform well," "I have the expectations to perform well," "performing well is important to me," "this task is a positive challenge," and "I am the type of person who does well on these tasks"). Two distinct scores (ranging from 1 to 7) were calculated for the threat and the challenge aspect, by averaging the related item scores.

Engagement. Participants' engagement was assessed after each session using seven items from the Intrinsic Motivation Inventory with seven response options (1= "totally disagree" to 7 = "totally agree) (Center for Self-determination Theory, 2019; McAuley et al., 1989). An example statement is: "I would describe this activity as very interesting". The negatively formulated statements ("I thought this was a boring activity" and "This activity did not hold my attention at all") were recoded. A total score was calculated by averaging the scores on all items (range 1 to 7). The higher the score, the higher participants' engagement during the training.

Procedure

After giving written informed consent and filling in the questionnaire about gaming experience as well as the trait anxiety questionnaire on the first session, each session consisted of participants putting on the respiration belt and the heart-rate belt, receiving a police vest, the controllers (radio and gun in the VR environment), headphone, and VR headset. Next, participants filled in the pre-test questionnaire, standing, while their physiological baseline was measured. The participants then performed the VR task for around 15 minutes until they completed all the waves of the game. In the first session, participants first played a tutorial in which they were instructed about processes such as confirming a radio message, followed by the actual VR task. The tutorial was also present in the second session, with additional information about biofeedback control and a breathing pace training. From session two onward, participants were instructed to breathe five seconds in and five seconds out while playing. After playing each session, participants filled in the rest of the questionnaires regarding their appraisals of the game. When the participant finished all questionnaires, they were presented with their average session scores of control over the situation, suspect and self.

To allow statistical inferences in small sample sizes, in this single case experimental design (SCED) study participants were invited ten times to the training sessions. A withdrawal ABAB design was used in which the experimentally withdrawn variable was the presence of online biofeedback on the participant's breathing pace, where slow and deep breathing was rewarded. The online biofeedback was presented to the participants by means of vision impairment while performing the VR task. Over the course of one month (September 2020), all participants completed the ten sessions (phases) following a withdrawal design (ABBABABABA). The majority of the sessions were separated from each other by one to five days. Due to participant scheduling limitations, some sessions had however to be done in the same day. In such a case, the first session always was an A phase, to prevent short term carry-over from the biofeedback presentation, displayed in B phases. In other words, we wanted to prevent the risk that players would apply the breathing technique in an A phase just because they had to apply it in the B phase that happened only minutes before. This security measure was implemented to ensure that looking at contrasts between a biofeedback B phase and a subsequent A phase would reflect as much as possible learning transfer and not mere short-term habituation. Continuous breathing-based biofeedback was added in the B-phases and withdrawn in the A-phases. The design is in line with the guidelines for small-N designs, by allowing a minimum of four repetitions of the addition-withdrawal procedure (Kratochwill et al., 2013).

Data preparation

Any identification information was removed from participant data and the data were securely stored on password-protected servers hosted by the Radboud University. The data was only accessible for approved members of the research team. The research data was not shared with the police organization, nor with the IBT center from which participants were recruited.

Physiological recordings. The physiological data (breathing biofeedback score and inter-beat 'R-R' intervals) were automatically saved at a rate of .5Hz. The data points were then synchronized with the game events and averaged before the start of the game to constitute the baseline. Since the baselines were inconsistent in length, the shortest recording (29 samples = 56 seconds) was used as a length reference, hence we only considered the last 29 samples portion for longer baselines. Per session, the inter-beat (R – R) intervals were interpolated to extract low (.04 Hz to .15 Hz) and high (.15 Hz to .4 Hz) frequency HRV, the low/high frequency ratio of HRV, and the coherence between the low frequency HRV and breathing. The coherence score was calculated by quantifying the similarity of the breathing and inter-beat interval time-series (Shaffer & Ginsberg, 2017). This quantification, bound between 0 and 1, is frequency specific. We therefore only considered the low-frequency HRV range as it is the range at which paced breathing would happen. The low frequency coherence metric is extracted as an index of relaxation, as Hayano & Yuda (2019) suggested that breathing induced fluctuations in the low frequency spectra of HRV reflects the presence of resting function. The scripts used to extract those HRV metrics can be found on GitHub (Brammer, 2020b) and were taken from the guidelines proposed by Shaffer & Ginsberg (2017).

Decision-making and response inhibition behavior. In-game events and actions were summed within a session to compute accuracy and signal detection measures. Hit scores were calculated by adding events where the participant shoots at an incoming hostile target before it reaches the player. Similarly, miss scores were the sums of hostile targets reaching the player before being shot, correct rejection scores were the sums of friendly targets reaching the player unharmed and finally false alarm scores were the sums of friendly targets being shot by the player. For each of those scores, we also recorded if the target involved in the event had the primed large identifier (body type announced by the radio dispatch as potentially hostile).

Data analysis

Missing and excluded data. Due to a technical issue with the heart rate belt recordings, the HR data of session 1 was missing for participants 3 and 5, although the participants did complete the training session. Moreover, due to material failure with the breathing belt used for biofeedback, the breathing data of session 8 was missing for participant 4, although the participant completed the full training session. Subject 9 was excluded from HRV-related analyses, since his high frequency component of HRV was more than 3 standard deviations higher than the rest of the group, for several sessions; this was due to a lack of accuracy in the R peak detection from the heart rate belt. This problem did not affect biofeedback measurements.

Environment design validation. Due to the small sample size, this section of the results is purely descriptive, as only the biofeedback data allowed to make meaningful statistical inferences thanks to the withdrawal (ABBABABABA) design.

Game arousal, challenge, and engagement. To descriptively assess, for each subject, the evolution of the level of arousal, challenge and engagement, individual trajectories were plotted alongside group average. For the HR, average baseline and in-game session scores were obtained. The in-game scores were obtained by subtracting the average baseline score from the in-game scores.

Biofeedback relevance. To assess whether our biofeedback manipulation was successful in influencing HRV, the average breathing biofeedback score within a session was correlated with the low and high frequency components of HRV and with the coherence between the breathing and low frequency HRV. The latter correlation is used to measure the extent to which low frequency HRV is influenced by slow and deep breathing, an indicator of resting function rather than sympathetic dominance (Hayano & Yuda, 2019). Since correlations were based on all 10 sessions of eight participants in total, where each session is a data point, we used a repeated measure correlation approach using the R package "rmcorr" (Bakdash & Marusich, 2017, 2021). Subject 9 had to be excluded from those analyses due to a measurement error.

To measure how awareness of the breathing control performance evolved over the course of the 10 sessions, we computed for each session the difference between participants' self-rating in breathing performance (an auto evaluation ranging from 0% to 100%) and the actual session score of "control of the self" (the percentage of the session that the player spent with a breathing biofeedback score of at least 0.8 over 1). Thus, the awareness score was calculated as: real "self-control" score – self rating "of self-control" score.

To further investigate the relevance of biofeedback, we looked at how often players mentioned breathing (e.g., "I need to focus more on the breathing"), biofeedback visual impairment (e.g., "Too much tunnel vision") and self-control score (e.g., "Focus on self-control") in the open-ended questions asked to the player before and after the VR task. Mentions of action-related elements (e.g., "Reload more often" or "Monitor 360") were scored as well.

Decision-making behavior. To measure decision-making (shoot/don't shoot) performance, we calculated the amount of hits, misses, correct rejection and false alarms per subject and session. Additionally, sensitivity (*d'*= [z(Hit rate)-z(False alarm rate)])) and response bias (criterion = - [z(Hit rate)-z(False alarm rate)]/2) were computed according to signal detection theory (McFall & Treat, 1999). Since some participants managed to avoid false alarms for an entire session, a loglinear correction was applied to the data to avoid infinite values (Hautus, 1995; Stanislaw & Todorov, 1999). To describe the effect of the priming, we calculated per session and subject the difference in false alarm rate between targets that had the body type primed by radio dispatch and those who did not [1].

Training efficiency: physiological control. Next, we tested the link between the presentation of online biofeedback in B-phases and the improvement in breathing control in action witnessed in all subjects. In this methodology, the goal is to search for the effect of repeatedly adding and removing a variable. Evidence for a causal

[1] Individual trajectories were plotted alongside the average of all participants, with standard error.

effect is then gathered through multiple analyses to ensure that the found effect is genuine. In our case, the experimentally added and removed variable is the presence of biofeedback, implemented as the tunneled vision described above. According to the requirements, of the SCED methodology (Kratochwill et al., 2013), multiple datapoints have to be extracted per session, which allows this section of the study to perform inferential statistics. The biofeedback values of a session were separated in 15 second bins and averaged for each bin. The biofeedback scores of individual participants were analyzed within and between sessions (A-sessions, B-sessions, B-A-sessions, and A-B sessions). The six features of the SCED visual analysis were level, trend, variability, immediacy of effect, consistency of data patterns across similar phases and degree of overlap of data (Kratochwill et al., 2013). A change of the data-patterns when there was a change in condition (addition or removal of biofeedback) indicates the biofeedback had an effect. To determine whether there was a significant intervention effect for a participant, a minimum of three such changes was needed for at least three of the six features (Horner et al., 2005; Kratochwill et al., 2013; Lane & Gast, 2014). Additional descriptive representations of HRV parameters were included in this section, since the main goal of breathing-based biofeedback was to calm participants by regaining parasympathetic dominance through HRV increase.

Single-Case Experimental Design Analysis. To illustrate the within and between session dynamics, the evolution of the biofeedback scores, evaluated at the 15 second bin level throughout the entire training, were plotted for each subject along with separated fitted trend lines for the A and the B phases. The trend lines were obtained with the MATLAB function “polfit.m”. To increase visibility of the data, the lines were smoothed with the moving average option of the MATLAB function “smooth.m”. Since the biofeedback score was measured on a time window of 30 seconds, the first two data bins of each session were discarded to avoid analyzing data from the baseline period.

Of the six SCED features used to test the role of biofeedback, data overlap is the one presented in the result section as it is most relevant to our design. For completeness, we report other aspects in the supplementary material 3. To test whether vision impairment was a salient enough way of providing biofeedback in B-sessions, and causally influenced the player’s behavior, data overlap of the biofeedback values of consecutive sessions was calculated by using the Kendall’s tau ranked correlation coefficient (Kendall, 1938). For each subject, the tau scores were then aggregated into two distinct scores (addition and removal) by combining effects of single contrasts (Manolov et al., 2015; Tarlow, 2017a, 2018). The addition score encompasses the tau scores for A-to-B phases contrasts, hence when comparing consecutive sessions where the first is played without online biofeedback and the second with biofeedback. Conversely, the removal score encompasses the tau scores for B-to-A phases (transition from with to without biofeedback). Kendall's tau was extracted by using an edited version of the R code by Tarlow (2017a). The editing consisted in removing Theil–Sen estimator used for baseline trend correction (Tarlow, 2017b). Baseline trend correction was not used for this dataset, as trend analysis revealed that a within session trend was often present but negative, hence correcting for it would lead to false positive results.

RESULTS

Goal 1: Environment design validation

Does the VR-game environment evoke arousal and engagement?

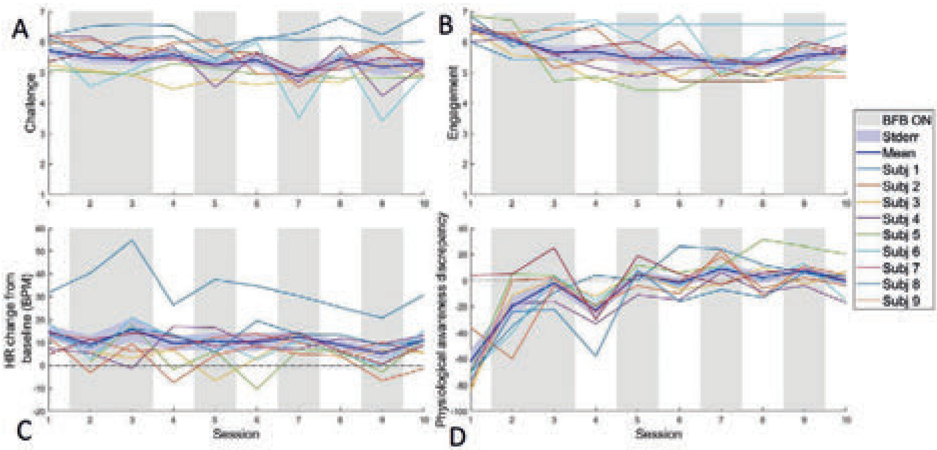
The VR-game elicited psychophysiological arousal and was experienced as challenging and engaging in the small sample at our disposition. This is indicated by the self-reported threat and challenge appraisals, the increase in HR and the self-reported engagement score (see Figure 4.2A to 4.2C). As expected, police officers reported consistently low levels of experienced threat, with an average of 2.39 on a 7 point scale (*SD* = .86; see supplementary material 4) over all sessions and participants. However, challenge scores were consistently high (Figure 4.2A) indicating that the participants experienced the game as positively stimulating (*M* = 5.36, *SD* = .65). In terms of absolute HR during the sessions, the average was 79 BPM. There was a marked peak in HR during the first session (*M* = 99 BPM) with substantial variation between subjects from *M* = 78 BPM (for subject 4) to *M* = 176 BPM (for subject 1). This strong variability was only witnessed for the first session. Relative to the baseline period immediately preceding the game start, an average increase of HR from baseline of approximately 10 BPM was observed (Figure 4.2C), similar to increases witnessed in established stress induction protocols (Vogel et al., 2015). As shown in Figure 4.2B also participants' reported engagement (scale 1-7) was generally high (*M* = 5.65, *SD* = .66), steadily decreasing until session seven, where the average engagement was still 5.3, corresponding to a moderately high level of engagement.

Does the biofeedback implementation influence HRV and facilitate physiological awareness?

In our limited sample, a strong and significant positive relation was found between the average biofeedback score and the coherence between low-frequency HRV and respiration in a session (r(68) = .72, $p < .001$). A smaller positive correlation was also found for low frequency HRV ($r_s(68) = .47$, $p > .001$) but not for high frequency HRV ($r_s(68) = .15$, $p = .212$). Hence, biofeedback scores correlated strongly with breathing induced fluctuations in HRV, a resting function index (Hayano & Yuda, 2019; Shaffer et al., 2014).

Figure 4.2

Stability of arousal and engagement during the course of the training.



Note. Changes across the whole training in: (A) self-reported challenge (1 = very low to 7 = very high); (B) mean engagement (1= very low to 7 = very high); (C) HR change, in BMP, from the 30s pre-test baseline (the dashed line represents a null change); (D) Difference between the self-reported estimation of biofeedback control score after the VR task and the real score (range 0 to 100%; plotted value = real score – reported score). Stderr = standard error of the mean; BFB ON = Sessions in which online biofeedback was presented to the participants.

To evaluate the participants' self-awareness of physiological control during the training process, the differences between the self-rating of physiological control and the objective biofeedback score are plotted in Figure 4.2D. A fast reduction in differences (overestimation) can be seen in the first half of the training process, with a notable yet short-lived increase in difference once online biofeedback was removed for the first time in session 4. Thus, subjects became considerably more accurate in their assessment of their physiological control over the course of the training.

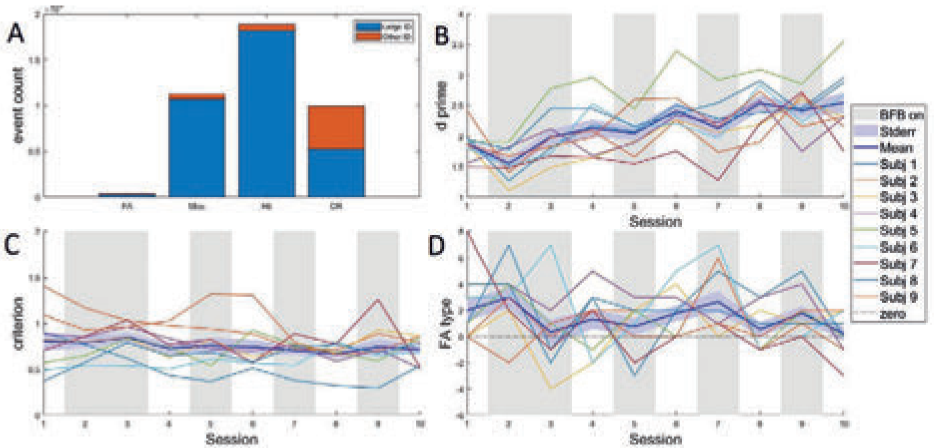
Additionally, the answers given to the pre- and post- questionnaires (see supplementary material 5), indicated that the participants mentioned breathing more after biofeedback sessions, both in the pre-test phase before the session (from five subjects in session 1 to all nine subjects in session 3) and in the post-test phase when debriefing (from four subjects in session 1 to eight subjects after sessions 2 and 3). Thus, biofeedback sessions successfully increased attention to breathing control.

Does the behavior in our application follow expected patterns relating to response inhibition under stress?

The distribution of correct and incorrect decisions in the game is presented in Figure 4.3 (see Table S6.1 in supplementary material 6 for details). Results showed that misses were by far the most likely type of mistake and police trainers avoided to make false alarm responses (Figure 4.3A). Over time, participants steadily increased in accuracy (sensitivity) as assessed by d prime (*d'*; Figure 4.3B), while their response bias (criterion) remained stable and conservative (Figure 4.3C). Interestingly, *d'* tended to be lower in sessions with biofeedback. Lastly, we investigated dispatch priming. As shown in Figure 4.3D, a higher proportion of friendly targets being shot (FA) was composed of targets whose large identifier (body type) was announced as presumably hostile in radio dispatch.

Figure 4.3

Behavioural metrics (in-game)



Note. (A) Go/nogo action distributions across sessions and subjects. Columns are sub-divided according to the large identifier of the target related to the trial. (B) *d'* from signal detection theory indicating sensitivity to target type (hostile/friendly); (C) criterion from signal detection theory indicating potential changes in strategy (D) Difference score of false alarm depending on the body type (primed body type – unprimed body type); Hit = hit; CR = correct rejection; FA = false alarm; Miss = misses; large ID = has the primed body type, announced as presumably hostile in radio dispatch; other ID = does not have a body type announced as presumably hostile in radio dispatch; BFB on = Sessions in which online biofeedback was presented to the participants; Stderr = standard error.

Goal 2: Training validation

Biofeedback score

The average biofeedback scores per participant are depicted in Figure 4.4A. The overall biofeedback score was M = .077 (SD = .16) in the first session (A_1), increasing to M = .497 (SD = .279) across all sessions with online biofeedback (B phases) and to M = .460 (SD = .266) across the subsequent sessions without online biofeedback (A_{2-5} phases). Six participants (2,4,6,7,8 and 9) showed higher biofeedback scores in B-phases (sessions with online-biofeedback), whereas three participants (1,3 and 5) had a higher average biofeedback score in A-phases (see supplementary material 3 for details).

Evolution of HRV

We here illustrate how beneficial the breathing-based biofeedback training is to influence HRV through RSA modulations. In Figures 4.4B to 4.4D the evolution of the low and high frequency components of HRV, and the coherence between low frequency HRV and the breathing pace are presented. While the low frequency HRV-breathing coherence (Figure 4.4C) parallels the increase over time shown by the breathing biofeedback score, displayed in Figure 4.4A, quite accurately, this pattern is less apparent for low (Figure 4.4D) or high (Figure 4.4B) frequency HRV. This indicates that our operationalization of the biofeedback parameter as a paced breathing reward successfully promoted higher coherence between the breathing and the HR (Schwerdtfeger et al., 2020; Shaffer et al., 2014), an index of resting function (Hayano & Yuda, 2019).

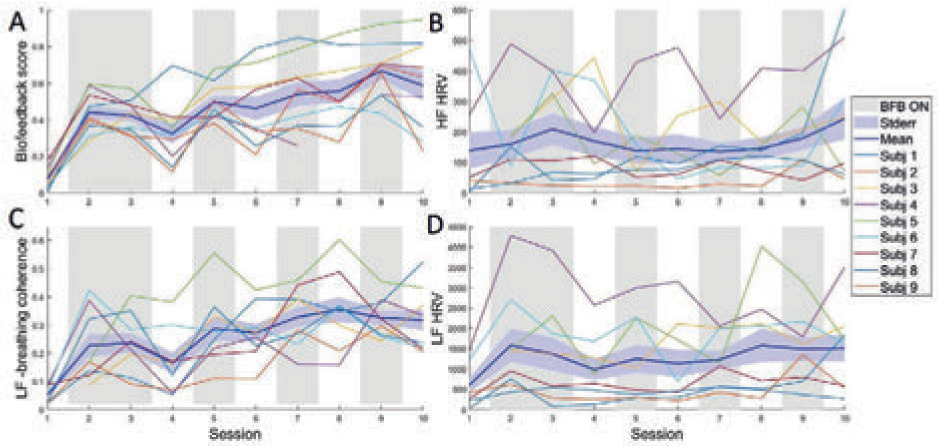
Singe-Case Experimental Design Analysis

The within and between session evolution of each participant's breathing biofeedback score (indicative of breathing control) is presented in Figure 4.5. Trend lines revealed that every subject showed a positive trend with increasing breathing biofeedback scores over time for both A and B phases, except subject 4 who showed no improvement in the B phases (red lines). Subjects 1, 3 and 5 displayed a particularly steep learning pace. Interestingly, those are also the subjects with a higher biofeedback average in A-phase sessions than in the B-phase ones.

Data overlap. To assess the influence of the breathing biofeedback presence (on or off), changes in breathing biofeedback score were related to the addition or removal of biofeedback by means of Kendall's tau (Table 4.2). Results indicated a consistent positive effect of the addition of online biofeedback across subjects (A to B phase).

Figure 4.4

Evolution of the breathing biofeedback score and the HRV indexes

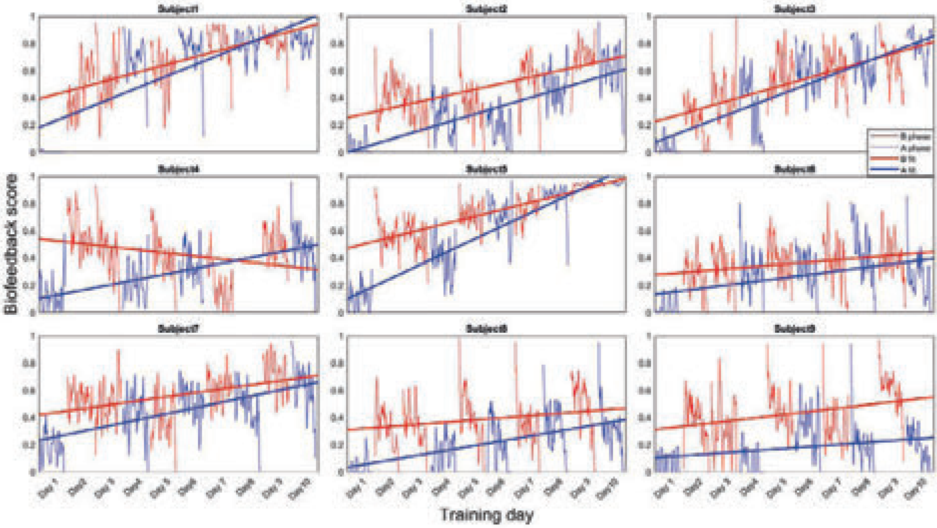


Note. Changes across the whole training in (A) the average biofeedback score; (B) the high frequency component of HRV (.15Hz to .4Hz); (C) the coherence between the low frequency HRV and breathing; (D) the low frequency component of HRV (.04Hz to .15Hz). BFB ON = B phases, where online biofeedback was presented; Stderr = Standard Error; LF-breathing coherence = coherence between the low frequency component of HRV and breathing; LF HRF = low frequency HRV.

Out of 36 individual transitions where biofeedback was added, 25 transitions were found to be positive and significant, with four subjects (2, 5, 8, and 9) each having three significant effects of non-overlap and a significant meta-effect when all A-to-B contrasts were merged, thus reaching formal criteria for a significant intervention effect. Importantly however, all nine subjects showed an effect in the same positive direction (Table 4.2). This suggests that, while there was a difference in the robustness of learning, participants generally improved in physiological control when online biofeedback was presented. Upon removal of biofeedback (B to A-phase), the pattern was more mixed. Out of 36 removal transitions, 14 transitions were found to be significantly negative. Only subjects 8 and 9 showed three significant repetitions of non-overlap, and only subject 9 had a significant meta-effect of withdrawal with in total six out of nine subjects showing a negative effect. Overall, the results support a causal effect of the biofeedback experimental manipulation, with seven participants (1, 2, 5, 6, 7, 8, and 9) showing a minimum of three significant non-overlap effects in the expected direction (Kratochwill et al., 2013). More elaborate SCED analyses can be found in the supplemental material 3, which according to formal SCED guidelines together indicated a moderate positive effect of our training procedure (Kratochwill et al., 2013).

Figure 4.5

Evolution of the breathing biofeedback score across the entire training period.



Note. Each datapoint represents the average biofeedback value for a 15seconds epoch. Scores range from 0 (no adherence to the rewarded breathing pace) to 1 (perfect adherence to the rewarded breathing pace). B phase = breathing score for B phase sessions (Session 2, 3, 5, 7 and 9); A phase = breathing score for A phase sessions (Session 1, 4, 6, 8 and 10); B fit = fitted line for the B phase sessions; A fit = fitted line for the A phase sessions.

Table 4.2

Kendall's tau non-overlap indices for consecutive sessions

Subjec	Kendall's tau									
	Addition (A-to-B phase contrast)					Removal (B-to-A phase contrast)				
	S1-S2	S4-S5	S6-S7	S8-S9	Meta	S3-S4	S5-S6	S7-S8	S9-S10	Meta
1	.8***	-.13	.21**	-.01	.26	.21**	.27***	-.16*	.03	.09
2	.67***	.16*	.47***	.29***	.42*	-.09	-.3***	-.12	-.1	-.15
3	.53***	.25***	.08	.06	.25	-.04	.02	0	.17*	.04
4	.6***	.35***	-.23**	*NaN*	.26	-.4***	-.12	*NaN*	-.05	-.19
5	.62***	.37***	.17*	.35***	.40*	-.26***	-.01	.1	.17*	.004
6	.52***	.14*	.11	0	.21	-.04	-.17*	.06	-.24**	-.09
7	.47***	-.03	.12	.29***	.22	-.05	.13	-.17*	-.09	-.04
8	.62***	.39***	.25***	.22**	.39*	-.34***	-.3***	-.02	-.24**	-.22
9	.56***	.61***	-.02	.49***	.43**	-.43***	-.19*	-.23**	-.55***	-.35*

* $p < .05$. ** $p < .01$. *** $p < .001$

Note. Grey cells indicate significant non-overlap between two consecutive sessions; i.e., S1-S2 = contrast between the first and the second session (the same applies for the other headings); Meta = Overall effect, obtained by merging the effect-sizes of the single contrasts.

DISCUSSION

We aimed to validate the design of a novel biofeedback training tool for in-action physiological regulation in police. Our first goal was design validation. We found that the VR game was successful in creating an engaging, challenging, and arousing environment. Our in-game biofeedback implementation succeeded in improving physiological awareness. Moreover, behavioral metrics of performance indicated suitability for probing response inhibition and priming effects under stress. For our second goal, investigating the training effectiveness, we additionally found support for the trainability of deep-breathing by presenting biofeedback in-action. These results, while preliminary given the small sample, suggest the feasibility and promise of influencing physiological control in an active decision-making context with tools like our new VR biofeedback application.

One of the rationales for using a VR context to train physiological control was the possibility to evoke strong emotions (Parsons, 2015). As expected, self-reports indicated a high sense of positive challenge (see Figure 4.2A and Supplementary Material 4), but a low sense of threat. This result indicates that the game is a good learning environment as it is not perceived as too threatening (Jamieson et al., 2013), although this interpretation should be cautious given the tendency of police officers to underreport socially undesirable emotions (Habersaat et al., 2021; Mccanlies et al., 2014). One additional possible explanation for the low threat score, as pointed out by Weerdmeester et al. (2020), is that the high level of challenge experienced by the participants throughout the training could partially be explained by the effect of biofeedback, which has been theorized to help participants better appraise threat into challenge since interoceptive signals of stress are dampened.

In terms of arousal, elicitation was successful in our VR environment as in-game HR increases from baseline were comparable to established stress induction protocols (Boesch et al., 2014; Vogel et al., 2015). However, we cannot establish the extent to which this arousal is of emotional nature or due to movement, as players were moving more during the game than at baseline. Yet, research by Gorini et al. (2011) showed effects of similar magnitude from exposure to a VR environment without participant movement.

We used VR to maximize the sense of engagement experienced by the participants, as it has been highlighted as an important nonspecific factor fostering behavioral change (Holzinger et al., 2006). Engagement also helps with accurate recollection of the participant's physiological arousal (McCall et al., 2015), thus improving biofeedback learning. The high and lasting engagement in our sample (see Figure 4.2B) may be ascribed to the advantages of VR (Riva et al., 2007), but could also be partially explained by peer pressure to perform, as our participants were, on their own initiative, actively comparing their performance scores with each other, a common practice in the police forces (Chen et al., 2013). This behavior provides support to the idea that reporting police-relevant scores at the end of each session helps strengthen engagement.

Just as the scores helped the participants to engage in the training, it also helped them to better estimate their performance and enhance physiological awareness, as illustrated by the improvements in physiological scores estimation (see Figure 4.2D). Similarly, the integration of biofeedback as peripheral view modulation was successful in eliciting awareness of breathing, when evaluated by self-report in open-ended questions (See Supplementary Material 5). This "tunneled vision" is known to be a relatable phenomenon for police officers (Dirkin, 1983; Klinger & Brunson, 2009). Indeed, as shown by the participant's written reports before and after action, mentions of breathing-related elements increased throughout the training.

Regarding the behavioral metrics, despite the small sample, we could show preliminary evidence that behavioral measurements were suitable for individual performance indexes to be extracted, per session (See Figure 4.3). It is worth noticing that decision-making stably improved throughout the training. Analyses of the mistakes showed preliminary evidence that priming the participants for certain targets' body types increased the chance for friendly targets to get shot when their body type partially matched the description, which is in line with priming tests performed in more realistic setups on police officers (Taylor, 2020). The game environment may therefore be suitable to investigate, in future studies, the *in-action* relation between physiology and response inhibition, as well as the effect of biofeedback on performance (Caballero Sánchez et al., 2016). Ideally, associations between in-game behavior and external measures of response inhibition under stress should be demonstrated in a larger sample, with a police relevant transfer condition, to further establish the validity of the in-game behavioral assessments. Importantly, in this game performance was affected by the biofeedback. Indeed, biofeedback was implemented as vision manipulation and thus could impair shooting and target recognition. A visual impairment linked to physiological measures created an artificial link between physiological control and behavioral performance. An illustration of this effect can be shown by fact that the sensitivity *d'* index tended to reduce in sessions with biofeedback. The artificial strengthening of the relation between physiology and behavior is a confound which, together with the changes in game difficulty during a session and the small sample available, made it difficult to investigate links between physiology and performance as have been previously reported (Hashemi et al., 2019). It is however anecdotally worth noticing that subject 5, the participant with the highest behavioral performance scores, was also the one with the highest biofeedback score in sessions without biofeedback. This transfer of physiological control skill to sessions where no online feedback on physiology was given suggests that successful transfer of breathing control skill may not come at the cost of reduced performance.

One of the main contributions of the current study was to investigate the proof-of-concept trainability of breathing biofeedback in an active context. Our results suggest that despite the fact that our subjects all received breathing-biofeedback in a classroom setup in the years preceding our experiment, no subject seemed able to control their breathing to satisfying levels prior to the introduction of biofeedback (see Figures 4.4A and 4.5). Additionally, our results show support for the trainability in active contexts of deep breathing, thus extending previous evidence showing deep breathing trainability in passive VR contexts (Rockstroh et al., 2019), but also in breaks during real-life police training scenarios (Andersen et al., 2018). Trying to apply emotion regulation techniques *in-action* is not a new idea, and has been implemented with techniques like neurofeedback (Schoneveld et al., 2016), and even breathing biofeedback (Bossenbroek et al., 2020), albeit in a rather meditative VR context. To the best of our knowledge, this is the first attempt at rewarding deep breathing in real-time while immersed in an arousing and active decision-making context. Additionally, only a few participants reached a plateau in biofeedback training after ten sessions. It is therefore unclear when the training benefits of such technique would stop. This issue should be investigated in future research, since there are institutional pressures to shorten training for police officers, sometimes to the point of rendering them useless (Di Nota et al., 2021).

Importantly, although the deep breathing in an active context was shown to reliably increase due to biofeedback, one limitation of this study is that a positive effect of the training on parasympathetic dominance indices could not be demonstrated and should be further investigated in future studies. While no claim can be made regarding parasympathetic control of the heart due to the fact that the heart is controlled by many systems simultaneously (Hayano & Yuda, 2019), we did identify a potentially strong link between adherence to the required breathing pace and a resting function index (the coherence between breathing and low-frequency HRV), as well as with an index of low-frequency HRV. We found no association for high-frequency HRV, which has been linked to increased performance in decision tasks (Gamble et al., 2018) and is usually seen as an index of parasympathetic dominance. The lack of results in high-frequency HRV might have a methodological explanation, as standing has been reported to dampen this HRV metric (Srinivasan et al., 2002). Additionally, training deep breathing transfers the breathing-induced HRV fluctuations from the higher to the lower part of the spectrum, hence increasing low-frequency HRV, possibly at the cost of high-frequency HRV (Hayano & Yuda, 2019). While low-frequency HRV may in fact be driven by parasympathetic nervous system activity during slow breathing (Kromenacker et al., 2018; Schwerdtfeger et al., 2020) it does not allow claims about the influence of the parasympathetic nervous system. Moreover, we found no consistent increase in either low or high frequency HRV across training, only in coherence. While coherence has been linked to improved cognitive and emotional function (Mather & Thayer, 2018; Schwerdtfeger et al., 2020), it is not possible, with the current dataset, to assess if the specific effect we report here is effective in producing such cognitive and emotional improvements.

The present study has other limitations, like the lack of a dedicated breathing pace baseline recording. Mostly, this study is limited due to the small sample size, which prevents not only inquiries on efficiency and training output (P. L. Smith & Little, 2018), but also on the interaction between behavior and physiology. A larger and more diverse sample, combined with a non-VR transfer task, would allow us to test the extent to which the training in VR transfers to physiological and behavioral patterns observed in real life, which is a critical test to evaluate the generalizability of the VR measures (L. C. Miller et al., 2019). A last limitation is the fact that our sample was, years prior to our training, exposed to a stress reduction course containing psychoeducational material and a short biofeedback session, as part of the mandatory training of all police officers in the Netherlands. However, the lack of evidence that the previous training and exposition to psychoeducational theories (that have a high prevalence in police curricula) had any effect in session 1 made us consider it as a non-confounding factor.

In terms of future work, our VR game could also be used to investigate the dynamical interaction between physiological state and decision-making in a more complex and engaging environment than traditional laboratory assessments. Indeed, little is known about the effectiveness of physiological control in the variety of phases through which police action evolves. While a fully realistic VR setup would contain too much variability and dimensions to measure performance reliably (Brehmer, 1992; Michela et al., 2019), testing well-known scientific measurements of decision-making in dynamical interactions could prove invaluable to relate performance and physiological control, but also to further investigate the external validity of those paradigms (Flake et al., 2017). Additionally, we recommend future studies to individualize the biofeedback target to the player, thus rewarding the breathing frequency and associated strategy that more efficiently raises HRV, rather than a fixed target like in the present study. Lastly, the output data includes positional tracking, and could be used to investigate potential movement confounds in the HRV measurement. More information on the ways movement can perturbate HRV biofeedback can be found in our implementation oriented publication (Brammer et al., 2021).

To conclude, while the generalizability of the results presented here needs to be assessed in future studies with larger samples, the present study showed the promise of a VR game to train physiological regulation in an arousing, police-relevant decision-making context. In addition, there seems to be support for targeting the behavioral action mechanisms that were included in the VR game. Our VR design opens new potential avenues for testing the impact of priming, arousal, and physiological control on (police-relevant) behavior in a more naturalistic context than traditional laboratory experiments. It will be important to design future studies to not only assess whether the VR game impacts decision-making under stress for police officers, but also if the physiological mediators of these effects are indeed the same that are targeted in the training.

Data Availability statement

The datasets presented in this article are not readily available due to privacy concerns, since the dataset contains performance indexes of police officers. Requests to access the datasets should be directed to the corresponding author, abele.michela@gmail.com

Ethics Statement

The studies involving human participants were reviewed and approved by Ethics Committee Social Science, Behavioral Sciences Institute Nijmegen, Thomas van Aquinostraat 4, ecsw@ru.nl. The patients/participants provided their written informed consent to participate in this study

Author Contributions

AM, AN, FK, and IG wrote the first draft of the manuscript. AM, JB, JP, MR, WD, KR, FK, and IG conceptualized the VR environment and experiment. WD provided the theoretical policing framework. WD and AS coordinated data collection. JB, JP, and RO conceptualized and implemented the biofeedback. AM, AN, and JB collected the data. AM, AN, JB, MR, and FK analyzed the data. All authors reviewed and contributed to the manuscript. KR, IG, and FK recruited funding for the project.

Funding

This work was supported by the NWO Creative Industry grant #055.16.139 2795 and CbusineZ. The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication

Acknowledgements

We would like to thank the police trainers from the Sevenum IBT center for their time and dedication. We would also like to thank Ken Koontz (https://www.kenkoontz.com), Erik van den Berge and Thijs Doornbos for their design and implementation work for the VR environment.

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg. 2022.806163/full#supplementary-material

SUPPLEMENTARY MATERIAL

Calculation of the scores presented to the participants at the end of the VR training session.

To give the participants a comprehensive summary of their performance, self-constructed “control scores” were presented at the end of every VR session. Those scores were named to match the way police officers themselves refer to their performance in real life. Namely, they mention that being in control of the self is needed to control the suspect, which is in turn needed to control the situation.

Control over the situation

Policing theory outlines the need for police officers to keep control of the situation, which entails keeping monitoring their environment and the communication with the commanding center frequent. In our game environment, these two dimensions were translated in 3 metrics, two accounting for environment monitoring (number of times a hostile zombie managed to reach and attack the player and number of times a zombie managed to reach the player unobserved), and 1 for communication (confirming reception of the dispatch information). Each dimension was weighted as 50% of the final “control over the situation” score. Control over the situation consisted of the number of times a participant was attacked by a zombie, the times a participant was surprised approached (by a zombie which the participant did not see) and the number of radio confirmations (confirming the message of which zombies, based on eye colour, to shoot by pressing a button on the radio). These variables were calculated using the following equations:

- Times attacked by zombie (TA) = (1- ((#first time hit by hostile) / (#total number of hostile)))*100
- Surprise approaches (SA) = (1 – (#hostile AND innocent unseen approaches) / (#total number of zombies))*100
- Radio confirmations (RC) = ((#radio confirmations made) / (#radio confirmations required in the game))*100
- Control over the situation = (TA + SA + 2*RC)/4

Control over the suspect

This score was summarizing the decision made by the player. The first measure is the arrest fire count. According to the Dutch police theory, police officers are allowed to shoot at the legs of aggressive incoming suspects from la large distance (15m to 25m), while shooting at the chest is recommended at shorter distances (up to 5m, after which other coercion tools are recommended for the officer's safety). The distance was slightly adapted in the game, but the principle of rewarding long distance shots was kept to encourage anticipating threats. This decision-making aspect was only weakly and linearly rewarded as many occasions to shoot at long distance are provided by the game.

Shooting mistakes were severely punished score-wise. Shooting mistakes entails shooting friendly zombies and shooting zombies before they were assigned an eye color (element allowing the player to decide if a zombie is friendly or not). Those mistakes were punished on a logarithmic scale, which allowed to punish the first mistake rather severely (~20% of final score was lost with the first friendly shooting mistake) and the following ones less severely, so that a player making many mistakes would not end up with a negative score.

Control over the suspect consisted of the amount of arrest fire hits (shooting a zombie from 13-17 meters), too far hits (shooting a zombie when the eye colour is not visible yet), and no-go hits (shooting a zombie with the wrong eye colour (innocent zombie)). These variables were calculated using the following equations:

- Arrest fire hits (AFH) = ((#hostile_hit in the 13-17m range) / (#hostile_hit))* 100
- Too far hits (TFH) = (1 - log10[(#Far_hit +2) / 2])*100
- No-go hits (NGH) = (1 - log10[(#innocent_hit +2) / 2])* 100
- Control over the suspect = (AFH + TFH + 3 * NGH)/5

Control over the self

Since the point of the training is to enhance awareness on the breathing control during the active decision-making context, this score was made purposefully more difficult to increase, hence nudging the player to pay more attention to this particular aspect of the summary scores for the following sessions. To that end, only moments of high breathing control (scores of 0.8/1 breath control) were counting toward the final "control of the self" score.

Control over the self (i.e. self-regulation of breathing) reflected the proportion of the breathing rate of the participant matching the required breathing pace of four to twelve breaths per minute during that specific period (e.g. wave, session). The biofeedback score was calculated using the following equations:

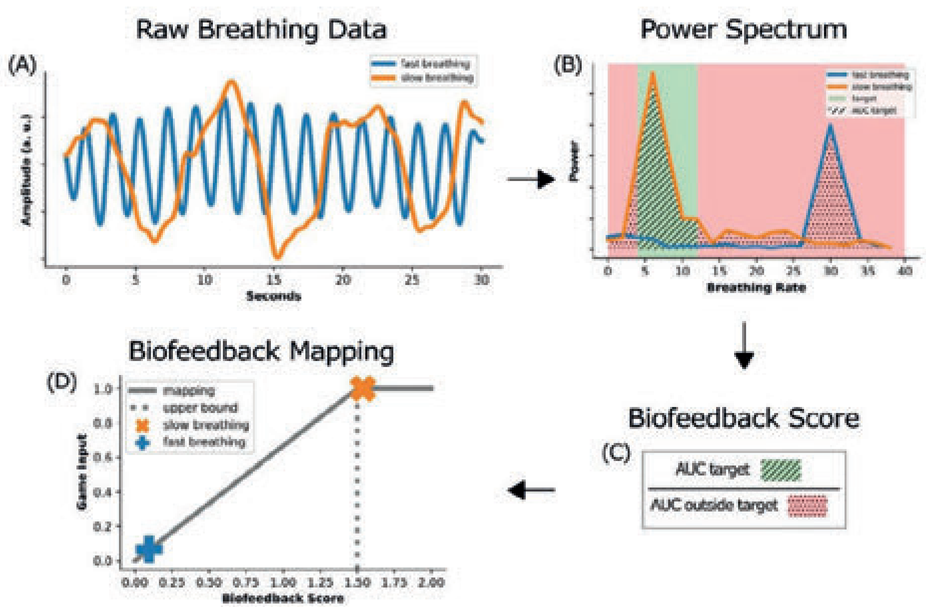
- Biofeedback score untransformed (BSU) = spectral density over the biofeedback target range of four to twelve breaths per minute / spectral density outside the target range of four to twelve breaths per minute (Brammer et al., 2021).

- Biofeedback score transformed (BST) = (BSU - (min(BSU))) / (max(BSU)) - (min(BSU))
- Control over the self = ((time BST higher than 0.8) / (total time)) * 100

The transformed biofeedback score ranges from 0 to 1. If BSU was larger than 1.5, A BST of 1 was returned since 1 is the maximum biofeedback score.

The process of calculating the biofeedback score is depicted in Figure S.4.1. Every 2 seconds, a breathing segment of the last 30 seconds of breathing data was analyzed, resulting in an overlap of 28 seconds between consecutive segments. Figures S.4.1A, S.4.1B and S.4.1D show two different breathing paces: a fast-breathing pace (blue) and a slow breathing pace (orange). Figure S.4.1B represents a power spectrum in which it is visualized that a breathing rate of 4 to 12 breaths per minute is the rewarded breathing pace. The power spectrum is used to evaluate how much of the current breathing rate matches the biofeedback target of 4 to 12 breaths per minute (the green area) in Figure A. 4.1B. To retrieve the biofeedback score, the green area under the curve (AUC target) is divided by the red area under the curve (AUC outside target). Next, the biofeedback score was standardized, as shown in Figure S.4.1 (Brammer et al., 2021).

Supplementary Figure S.4.1



Note. Reprinted from "Breathing Biofeedback for Police Officers in a Stressful Virtual Environment: Challenges and Opportunities," by J. C. Brammer, J. M. van Peer, A. Michela, M. M. J. W. van Rooij, R. Oostenveld, F. Klumpers, W. Dorrestijn, I. Granic, and K. Roelofs, 2021. Frontiers in Psychology, 12, p. 7 (https://doi.org/10.3389/fpsyg.2021.586553). CC BY, reprinted with permission.

Target approach analysis (TAA) and after action report (AAR)

Doel-Aanpak Analyse (TAA) (Target approach analysis)

1. Doel : waar ga je je in deze sessie op richten? (meerdere keuzes zijn mogelijk)
   Goal: what will you focus on in this session? (several choices possible)

❐ Controle over het zelf *(Control of the self)*
❐ Controle van de verdachte *(Control of the suspect)*
❐ Controle van de situatie *(Control of the situation)*

2. Risico's : *(Risks:)*
   Aantal verkeerde doelen __________________ *(number of wrong targets)*

3. Mag ik: wat is volgens jou de beste score die behaald kan worden in een sessie?
   (Can I: what do you think is the best score that can be achieved in a session?)
   Controle over het zelf _____%
   Controle van de verdachte _____%
   Controle van de situatie _____%
   Aantal verkeerde doelen ___________________

4. Kan ik : wat denk je dat je score zal zijn in deze sessie?
   (Can I: What do you think your score will be in this session?)
   Controle over het zelf _____%
   Controle van de verdachte _____%
   Controle van de situatie _____%
   Aantal verkeerde doelen ___________________

5. Plan van aanpak : *(Action plan:)*

Wat kan er fout gaan in deze sessie? en hoe ga je erop reageren?
(What can go wrong in this session? and how are you going to react to it?)

____

____

____

4

After Action Review

(Ging het goed, of ging het niet fout? Goed gegaan of goed gedaan?)

(Did it go right, or didn't it go wrong? Did it go right or did it right?)

1. Heb je veilig gewerkt? *(Did you work safely?)*
   Controle over het zelf _____%
   Controle van de verdachte _____%
   Controle van de situatie _____%
   Aantal verkeerde doelen ___________________

2. Wat maakt dat je wel/niet van je plan bent afgeweken?
   (What makes you deviate/not deviate from the plan?)

____

____

____

3. Wat doe je de volgende weer? *(What will you do again the next time?)*

____

____

____

4. Wat doe je de volgende keer anders? *(What will you do differently next time?)*

____

____

____

5. Wat doe je de volgende keer niet meer? (*What will you not do anymore next time?)*

____

____

____

6. Hoe voel je je? (Ben je klaar voor de volgende ronde?)
   (How do you feel? (Ready fort he next round?))

___

___

___

Uit: Beyond the Split Second, W. Dorrestijn 2014-2018.
(From: Beyond the Split Second, W. Dorrestijn 2014-2018. All rights reserved)

SCED analysis and tables

Table S4.3.1

Means and standard deviations of the biofeedback value per phase per participant.

Participant	A_1 phase, mean (SD)	B phases, mean (SD)	A_{2-5} phases, mean (SD)
1	.002(0.015)	.649(0.247)	.77(0.195)
2	.041(0.122)	.466(0.269)	.391(0.278)
3	.044(0.13)	.497(0.278)	.594(0.266)
4	.129(0.255)	.435(0.274)	.343(0.288)
5	.118(0.209)	.709(0.200)	.725(0.172)
6	.043(0.124)	.356(0.320)	.329(0.309)
7	.193(0.27)	.555(0.287)	.524(0.307)
8	.036(0.135)	.385(0.314)	.265(0.302)
9	.087(0.217)	.424(0.323)	.206(0.28)

Note. A_1 = first phase without biofeedback; B = phases with biofeedback; A_{2-5} = all phases without biofeedback (the first phase excluded); Biofeedback values can range between 0 and 1.

Level

To facilitate visual inspection of the evolution of participants' *level* in biofeedback score, figure 4.4A in the main article displays session averages for each participant. A consistent steep improvement in biofeedback control is clearly visible here between the first and the second training session, where the online biofeedback element is introduced for the first time. All subjects ended the training with a higher breathing score than in the very first sessions, and five subjects (1, 2, 3, 5 and 7) ended the training with a higher breathing score than in the first B-phase, thus suggesting that they learned how to control the biofeedback parameter. Two participants (1 and 5) approached a potential ceiling effect in the second half of the training.

Trend

The *trends* in biofeedback score within a training session were evaluated by means of a linear regression fitting (MATLAB function "polyfit.m"). A positive trend indicates an improvement in breathing control throughout the training session, whereas a negative trend indicates a decrease in breathing control over time in a session. The differences in trends between subsequent sessions were analyzed (decelerating or accelerating). An accelerating trend reflects a positive training effect since the training is designed to increase the biofeedback values (Lane & Gast, 2014). If the trend of the data changes with a change in condition (either addition or removal of biofeedback), this suggests the training has an effect. When a participant shows three of such changes, the training is deemed significantly effective (Kratochwill et al., 2013).

The *trends* in biofeedback score *within* a training session are presented in Table S4.6.2. As could be expected since the game gets harder throughout a session, a large proportion of the sessions (63 out of 90) displayed a negative trend, suggesting that players performed worse in breathing control over time. We also identified when changes in trend direction occurred in consecutive sessions, where biofeedback was added or removed. Five subjects (1, 2, 3, 5 and 7) each showed at least three effects in which the trend changed with a change in condition, thus providing evidence for training effectiveness. No subject except subject 3 displayed a trend in the first session, suggesting a potential floor effect in the first session.

Table S4.3.2

Slope of the linear regression trend, per session, per participant.

Participant	Session									
	1(A_1)	2(B_1)	3(B_2)	4(A_2)	5(B_3)	6(A_3)	7(B_4)	8(A_4)	9(B_5)	10(A_5)
1	0	.006*	.002	-.002	-.001	-.003*	-.003*	-.001	-.001	0
2	0	.004*	-.002	-.003	-.004*	-.002	0	0	.002	-.001
3	-.001	-.005*	-.002	-.009*	.002	-.002	-.002	-.005*	-.005*	-.003*
4	0	-.004*	-.006*	-.001	-.005*	-.001	-.005*		-.001	-.003
5	0	-.002	-.003	-.004	0	-.003*	.001	0	.001*	0
6	0	-.003	-.004	-.001	-.003	-.005*	-.004	-.007*	-.003	-.002
7	0	-.001	.002	-.001	0	-.004	.001	-.004	-.002	-.005*
8	0	-.007*	-.004	.002	-.006*	-.006*	-.004	-.004	-.006*	-.004
9	0	-.005*	.001	-.002	-.004	-.001	-.001	-.003	-.004*	-.004*

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$

Note. A_1 = first phase without biofeedback; B = phases with biofeedback; $A_{2\text{-}5}$ = phases without biofeedback (the first phase excluded); Grey cells indicate the session is part of a change; Data from session 8 of participant 4 is missing due to a technical issue.

Variability

Variability of the data refers to the fluctuation of the biofeedback scores per session (Horner et al., 2005; Kratochwill et al., 2013). Median absolute deviation (MAD) scores were calculated to compare variability between sessions and A/B phases (Levin et al., 2021). As shown in Table S4.6.3, the variability of the biofeedback values was higher in sessions with biofeedback (B phases) compared to sessions without biofeedback (A phases) for eight participants (1, 2, 3, 4, 5, 6, 8, and 9). Compared to the first baseline session (A_1), there was more variability in the subsequent sessions, both with (B phases) and without biofeedback ($A_{2\text{-}5}$ phases), in line with the earlier suggestion of a floor effect in the first session.

Table S4.3.3

Median Absolute Deviation (MAD) per session and phase, per participant

Participant	Session										
	1 (A_1)	2 (B_1)	3 (B_2)	4 (A_2)	5 (B_3)	6 (A_3)	7 (B_4)	8 (A_4)	9 (B_5)	10 (A_5)	$A_{2\text{-}5}$ phases	B phases
1	0	.283	.3	.122	.214	.096	.054	.066	.086	.07	.088	.187
2	0	.244	.243	.067	.245	.062	.222	.224	.146	.149	.125	.22
3	0	.105	.258	.287	.256	.224	.237	.242	.205	.09	.211	.212
4	0	.226	.311	.101	.231	.285	.12		.149	.217	.201	.208
5	0	.219	.192	.3	.143	.164	.082	.047	.018	.01	.13	.131
6	0	.118	.269	.223	.31	.29	.282	.321	.309	.092	.231	.258
7	0	.225	.272	.275	.279	.241	.213	.27	.121	.229	.254	.222
8	0	.249	.276	0	.304	.001	.272	.294	.242	.258	.138	.269
9	0	.274	.135	0	.263	.231	.219	0	.19	.029	.065	.216

Note. A_1 = first phase without biofeedback; B = phases with biofeedback; $A_{2\text{-}5}$ = phases without biofeedback (the first phase excluded). Data from session 8 of participant 4 is missing due to a technical issue.

Immediacy of effect

The immediacy of the effect of the intervention is usually examined by calculating the difference in level between the last three to five data points of one session and the first three to five data points of the next session to see whether the data immediately changes when biofeedback is either removed or added (Kratochwill et al., 2010). Some drawbacks to this approach are 1) ignoring autocorrelation between observations, 2) not using all the data of a session, and 3) no guidelines on interpreting the magnitude of the difference. Using a Bayesian estimator of abrupt change, seasonality, and trend (termed BEAST-model; Zhao et al., 2019) addresses all those issues. Therefore, the BEAST-model was used to assess the immediacy of effect. The BEAST-model assumes the change point (the change from one session to the next) as unknown and indicates based on the data after which data point in each session there was a substantial change in biofeedback values. The parameters of the BEAST-Model can be found in supplementary material 7. The training is considered effective for a participant when the estimated change point of the model corresponds three times with the actual change point.

The values in Table S4.6.4 indicate after which data point in each session there was a substantial change in biofeedback values according to the BEAST-model. Changes detected between session 2 and 3 were discarded, as both sessions are B phases. Eight participants (1, 2, 3, 4, 5, 6, 7 and 8) each showed at least three instances at which the estimated change point occurred at the true moment of change from one session to another (up to 10 samples before the end, or 3 samples after the start of a session).

Table S4.3.4

Estimations of the Bayesian change-points model transitions of A-to-B and B-to-A phases.

Subject	Sample of detected change in phase (distance from real sample)							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	82(10)	147(3)	210(-6)	353(-7)	422(62)	506(2)	619(43)	
2	76(4)	148(4)	290(2)	351(-9)	426(-6)	506(2)	570(-6)	650(2)
3	64(-8)	199(55)	338(50)	426(-6)	507(3)	660(12)		
4	74(2)	139(67)	208(64)	334(46)	434(2)	502(-2)	615(39)	
5	74(2)	136(64)	249(33)	321(33)	384(24)	497(-7)	650(2)	
6	106(34)	224(8)	290(2)	353(-7)	418(58)	506(2)	567(-9)	633(57)
7	73(1)	150(6)	213(-3)	281(-7)	362(2)	491(59)	611(35)	
8	72(0)	193(49)	290(2)	353(-7)	434(2)	536(32)	599(23)	660(12)
9	138(66)	202(58)	290(2)	377(17)	474(42)	536(32)	602(26)	

Note. A seasonal transition is considered to be correctly detected if placed up to 10 samples before the end, or 3 samples after start of a phase. The seasons to be detected are 72 samples long, each sample represents 15 seconds of gameplay; Grey cells indicate a correctly detected phase change

Summary of visual analysis

To summarize the visual analysis, all participants except subject 4 showed at least three effects for at least one index of the visual analysis. Four participants (1, 2, 5 and 7) reached formal significance as they showed effects in at least three main aspects of the visual analysis: trend, immediacy of effect and non-overlap of data. This result corresponds to a moderate positive evidence of an intervention effect, according to the guidelines for SCED analysis (Kratochwill et al., 2013). An overview, per participant, of how many replications of an effect for each main aspect of the visual analysis is included in Table S4.6.5.

Table S4.3.5

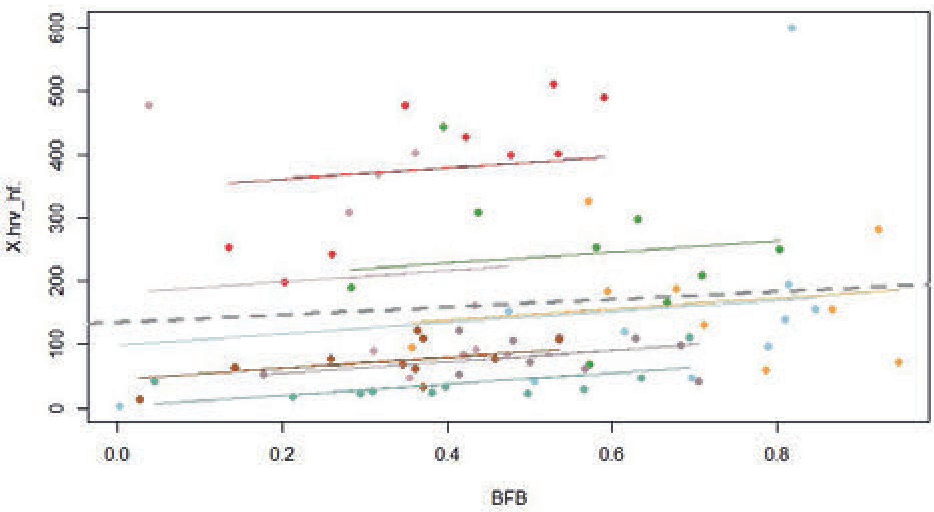
Overview of the replications of an effect for each main aspect of the visual analysis

Participant	Level	Variability	Trend	Non-overlap	Immediacy of effect	Positive effects triple repetitions	Positive effects overall
1	Y	Y	5	3	3	3	5
2	Y	Y	6	5	6	3	5
3	Y	N	4	2	3	2	3
4	N	Y	1	2	3	1	2
5	Y	N	8	5	3	3	4
6	N	Y	1	4	4	2	3
7	Y	N	7	3	4	3	4
8	N	Y	3	7	4	2	3
9	N	Y	3	5	1	1	2

Note. Grey cells indicate the presence of an effect for the aspects requiring a triple effect replication; Y = a positive effect was present; N a positive effect was not present

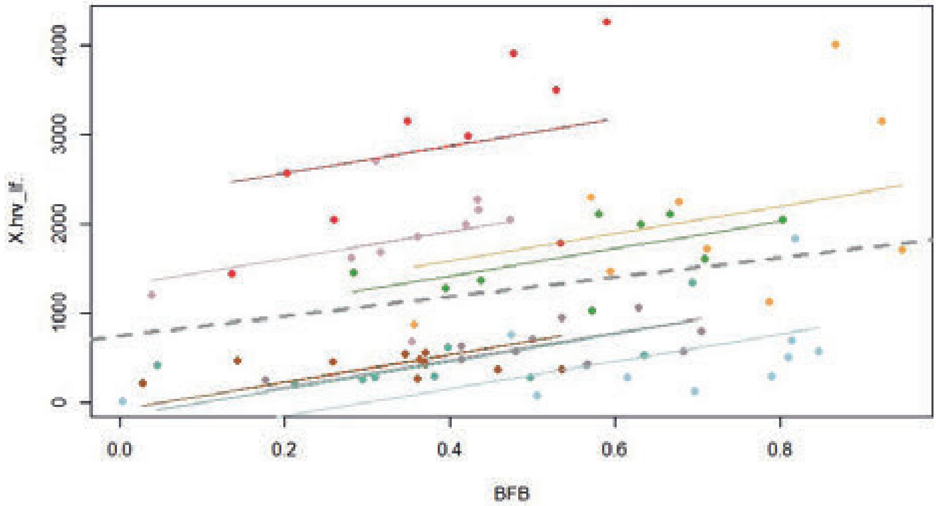
Threat and HRV analyses

Supplementary Figure S.4.4.1



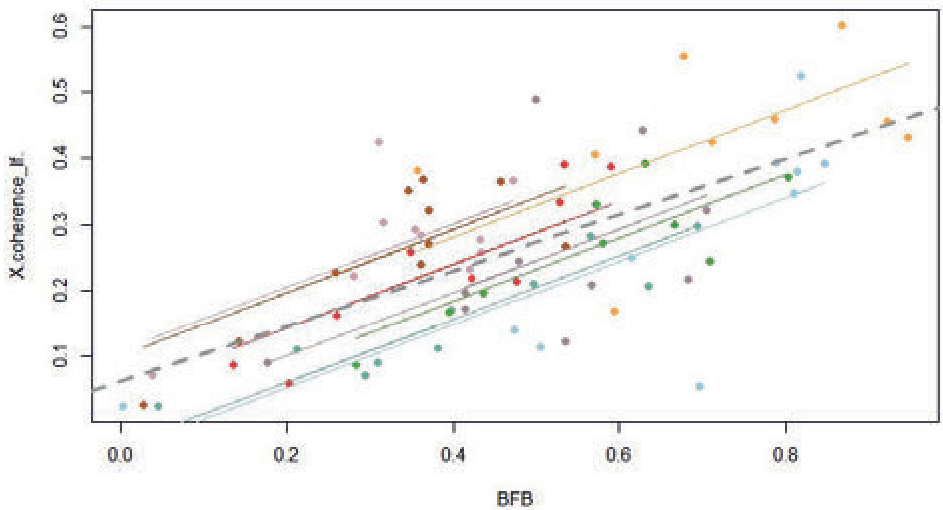
Note. Repeated measures correlation between High-frequency HRV and biofeedback score, evaluated at the session level; X.hrv_hf = high-frequency HRV; BFB = biofeedback score.

Supplementary Figure S.4.4.2



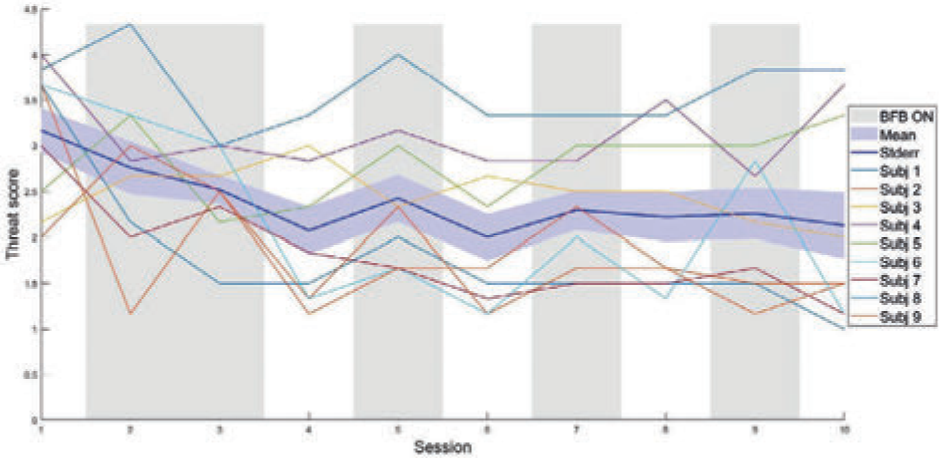
Note. Repeated measures correlation between Low-frequency HRV and biofeedback score, evaluated at the session level; X.hrv_lf = low-frequency HRV; BFB = biofeedback score.

Supplementary Figure S.4.4.3



Note. Repeated measures correlation between the low-frequency HRVcoherence with breathing and biofeedback score, evaluated at the session level; X.coherence_lf = low-frequency HRV coherence with breathing pace; BFB = biofeedback score.

Supplementary Figure S.4.4.4

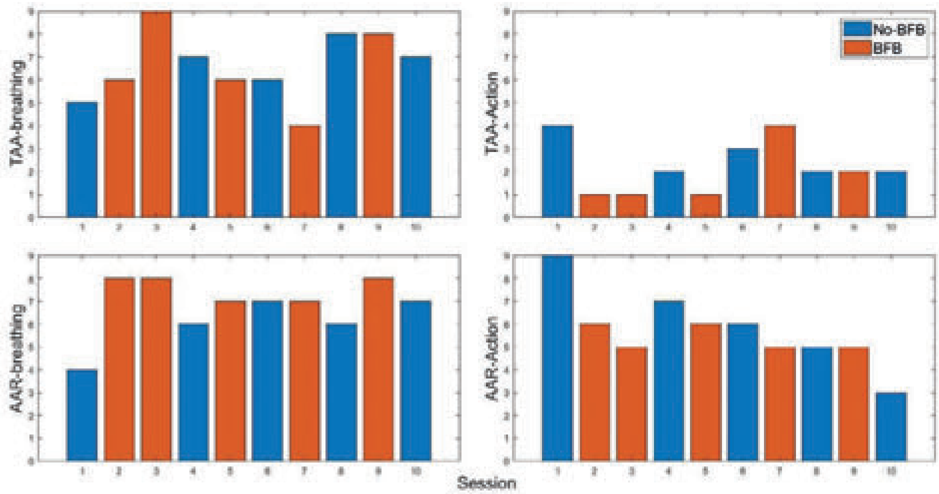


Note. Evolution of the threat-subscale of the Threat-Challenge appraisal questionnaire; Stderr = standard error of the mean; BFB ON = Sessions in which online biofeedback was presented to the participants.

4

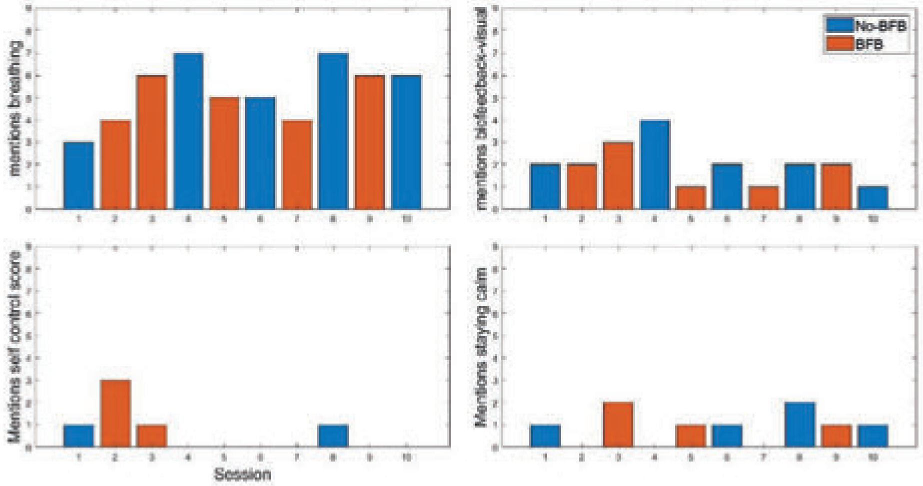
Target approach analysis (TAA) and after-action review (AAR) results

Supplementary Figure S.4.5.1



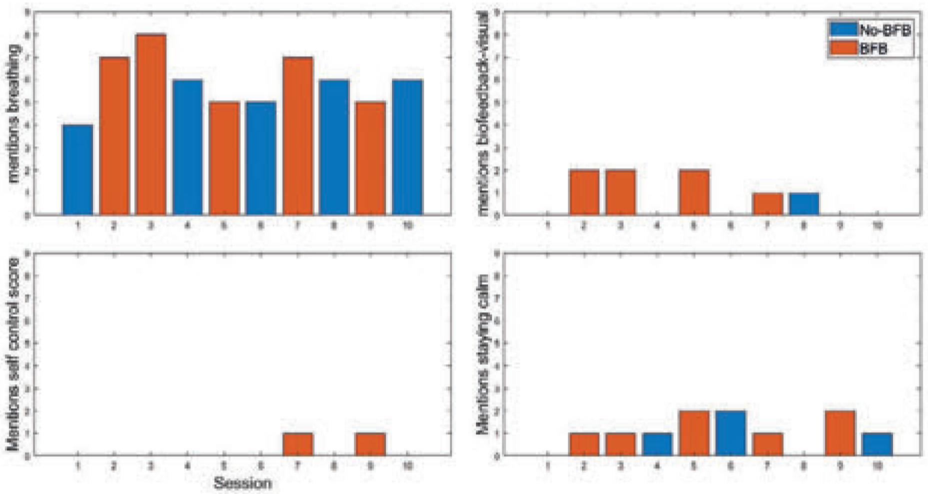
Note. The general occurrence of breathing and action-related keywords in the TAA and AAR questionnaires; No-BFB = Sessions without online biofeedback; BFB = Sessions with online biofeedback; -breathing = occurrences of participants mentioning breathing or biofeedback related keywords; -Action = occurrences of participants mentioning action related keywords

Supplementary Figure S.4.5.2



Note. The detailed occurrence of breathing-related keywords in the TAA questionnaire; No-BFB = Sessions without online biofeedback; BFB = Sessions with online biofeedback; mentions breathing = occurrences of participants mentioning breathing or respiration related keywords; mentions biofeedback-visual = occurrences of participants mentioning the graphical implementation of online biofeedback as visual impairment; mentions self-control score = occurrences of participants mentioning the self-control score directly; mentions staying calm = occurrences of participants mentioning the self-control score indirectly by referring at the state of calmness.

Supplementary Figure S.4.5.3



Note. The detailed occurrence of breathing-related keywords in the AAR questionnaire; No-BFB = Sessions without online biofeedback; BFB = Sessions with online biofeedback; mentions breathing = occurrences of participants mentioning breathing or respiration related keywords; mentions

Supplementary Figure S.4.5.3 (Contined)

biofeedback-visual = occurrences of participants mentioning the graphical implementation of online biofeedback as visual impairment; mentions self-control score = occurrences of participants mentioning the self-control score directly; mentions staying calm = occurrences of participants mentioning the self-control score indirectly by referring at the state of calmness

Behavioral priming effects

Table S4.6.1

Go/nogo actions distributions across sessions and subjects. Columns are sub-divided according to the large identifier of the zombie related to the trial.

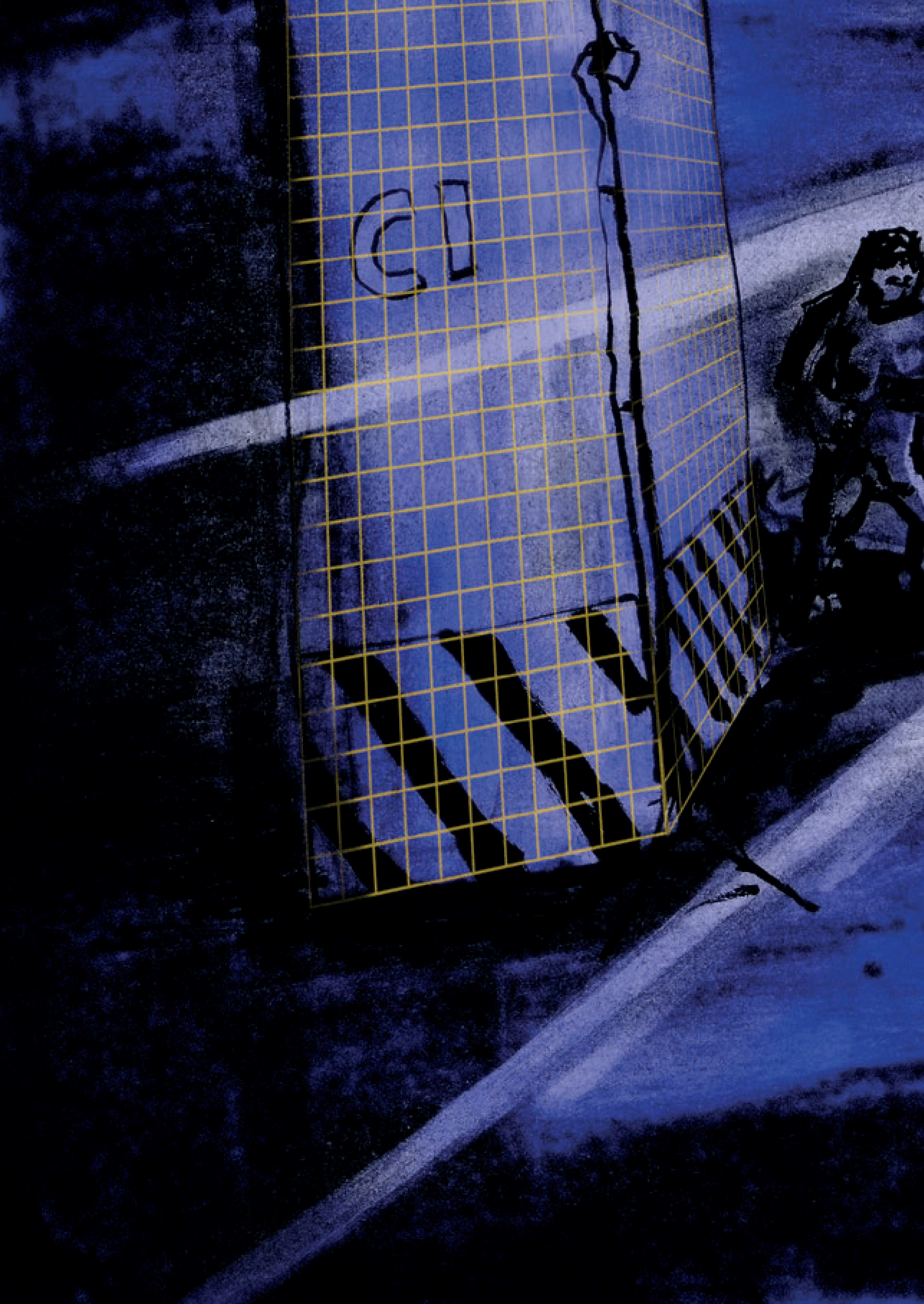
	Hit	CR	FA	Miss	Total
large ID	18169	5234	252	10708	34363
other ID	693	4718	122	576	6109
Total	18862	9952	374	11284	40472

Note. Hit = hit or true positives (hostile zombie shot before reaching player); CR = correct rejection or true negative (innocent zombie reaching player unharmed); FA = false alarm (innocent zombie shot before reaching player); Miss = miss or false negative (hostile zombie reaching player unharmed); large ID = has a large identifier (body type) announced as presumably hostile in radio dispatch; other ID = does not have a large identifier (body type) announced as presumably hostile in radio dispatch.

BEAST-model parameters

```
%% Set up the parameters needed for the BEAST algorithm
% Some of these parameters are the model specficiation parameters of BEAST
% (e.g., minSeasonOrder, maxSeasonOrder, minSetpDist_trend,
% minSepDist_Season); other parameters are just some input variables to
% control simulation behaviors or program outputs (e.g., samples,
% thinningFactor, seed, computeCredible).
%
opt.period    = 72;
opt.minSeasonOrder = 0;
opt.maxSeasonOrder = 1;
opt.minTrendOrder=0;
opt.maxTrendOrder=1;
opt.minSepDist_Trend  =  30;
opt.minSepDist_Season =  60;
opt.maxKnotNum_Trend =  10;
opt.maxKnotNum_Season = 10;
opt.maxMoveStepSize   = 10;
opt.samples = 720;
opt.thinningFactor = 1;
opt.burnin = 200;
opt.chainNumber=2;
opt.resamplingTrendOrderProb=0.2;
opt.resamplingSeasonOrderProb=0.17;
opt.omissionValue=-999;
opt.seed=100;
opt.computeCredible=0;
opt.computeSlopeSign=1;
opt.algorithm='beast';
opt.computeHarmonicOrder=1;
opt.computeTrendOrder=1;
opt.computeChangepoints=1;
%opt.timeDimensionIndex=3;
%% Run BEAST on "Y"
out=beast_default(Y, opt);
% extract "seasonal changes"
```

```
scp(f,:)=sort(out.scp);
% check if they fall in the 72 sample +3-10 range -> transition detected
```


CI

CHAPTER 5

Preparing the Heart for Duty: Virtual Reality Biofeedback in an Arousing Action Game Improves in-action Voluntary Heart Rate Variability Control in Experienced Police

Submitted

Authors:

Abele Michela[1], Jacobien M. van Peer[1], Robert Oostenveld[2,3], Wendy Dorrestijn[4], Annika S. Smit[5,6], Isabela Granic[1,7,†], Karin Roelofs[1,2,†] and Floris Klumpers[1,2,†]

[1] Behavioural Science Institute, Radboud University, Nijmegen, Netherlands
[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen
[3] NatMEG, Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden
[4] Faculty of Law, Radboud University, Nijmegen, Netherlands
[5] Police Academy of the Netherlands, Apeldoorn, Netherlands
6 University of Humanistic Studies, Utrecht, Netherlands.
[7] Faculty of Social Sciences, McMaster University, Hamilton, Canada

[†] These authors have contributed equally to this work and share last authorship

ABSTRACT

Adequate control over evolutionary engrained bodily stress reactions is essential to avoid disproportionate responses during highly arousing situations in police. Heart rate variability (HRV) biofeedback is a widely-used intervention aiming to improve stress regulation, but typically conducted under passive, low arousing conditions. By integrating closed-loop HRV biofeedback in an engaging Virtual Reality (VR) action game designed to contain the behavioral elements typically compromised under stress we trained in-action physiological self-control under high arousal to allow improved transfer to real-life situations. A pre-registered randomized controlled trial in 109 police trainers demonstrated 32% average increases in HRV, through the engaging and gamified closed loop biofeedback. The ability to voluntarily upregulate in-action HRV also transferred to conditions of VR gameplay without biofeedback (near transfer) and, critically, to a professional performance assessment outside VR (far transfer). These results suggest that real time-biofeedback in stressful and active action contexts can help train professionals such as police in real-life stress regulation.

Keywords: Biofeedback, Police training, VR, emotional control, decision-making, HRV.

INTRODUCTION

First responders such as police officers are routinely asked to make critical decisions under great pressure in the line of duty. Mistakes in such contexts can have grave consequences. Among the key factors contributing to inappropriate responses in such situations are psychophysiological arousal, subjective stress, and maladaptive emotional behavior (Miller, 2015; Staller et al., 2018). Indeed, stress-induced arousal impairs impulse control, results in attentional narrowing and habitual, rather than flexible, instrumental responding (Arnsten, 2015; Rued et al., 2019; Wirz et al., 2018). Moreover, longitudinal studies have shown that hyperarousal and alterations in neural systems linked to arousal control are key predictors of long-term stress symptoms (Admon et al., 2013; Roeckner et al., 2021; Zhang et al., 2022). Biofeedback (BF) has been employed as a targeted intervention to raise awareness of the currently experienced stress level, to increase the feeling of self-efficacy around the capacity to control one's own physiological reactions (Weerdmeester et al., 2020) and to train instrumental control over these processes (Andersen & Gustafsberg, 2016; Mccraty & Atkinson, 2012; Yu et al., 2018).

One of the most frequently used and successful BF interventions to control stress-induced arousal involves operant conditioning procedures to deepen and slow breathing which, in turn, enhances heart-rate variability (HRV; Lehrer et al., 2020). Slow deep breathing synchronizes the heart-rate (HR) with respiration, where the HR accelerates with inhalations and decelerates with exhalations (Yasuma & Hayano, 2004) which is indicative of relaxation (Hayano & Yuda, 2019) and linked to effective coping (Lehrer et al., 2020; Mather & Thayer, 2018). Mounting evidence has shown HRV upregulation is proven effective in reducing anxiety (Goessl et al., 2017), depression (Pizzoli et al., 2021), cognitive performance (Laborde, 2019), athletic performance (Jiménez Morgan & Molina Mora, 2017; Lagos et al., 2011) and generally helps police officers cope with stressful aspects of their job (Mccraty & Atkinson, 2012).

Despite its promise, widespread implementation of BF procedures is currently hindered by the fact that (1) current BF procedures are typically performed in a non-engaging way that requires high internal motivation and is not appealing for many users (Blum et al., 2019; Lüddecke & Felnhofer, 2022), and (2) trainings take place in a non-arousing, passive setting while application is typically expected in action and under stress, thereby hampering transfer to real world use (Burish & Schwartz, 1980; van der Meulen et al., 2018).

Contextualizing BF training in a game that creates a narrative to maximize engagement seems a promising direction to improve BF trainability and skill transfer (Gorini et al., 2011; Schoneveld et al., 2019). Virtual Reality (VR) has been shown to be effective as a tool to create an engaging and arousing active context to train HRV upregulation when it is most needed and most difficult to attain (Bossenbroek et al., 2020; Bouchard et al., 2012; Lobel et al., 2016; Rockstroh et al., 2019; Schoneveld et al., 2016). Designing engaging applications for in-action biofeedback is however a challenge as it involves adapting BF biomarkers to motion artifacts (Brammer et al., 2021) and serious games that can elicit genuine emotions and behaviors are notoriously hard to design (Scholten & Granic, 2019), even more so when the game should represent some level of policing reality (Michela et al., 2019).

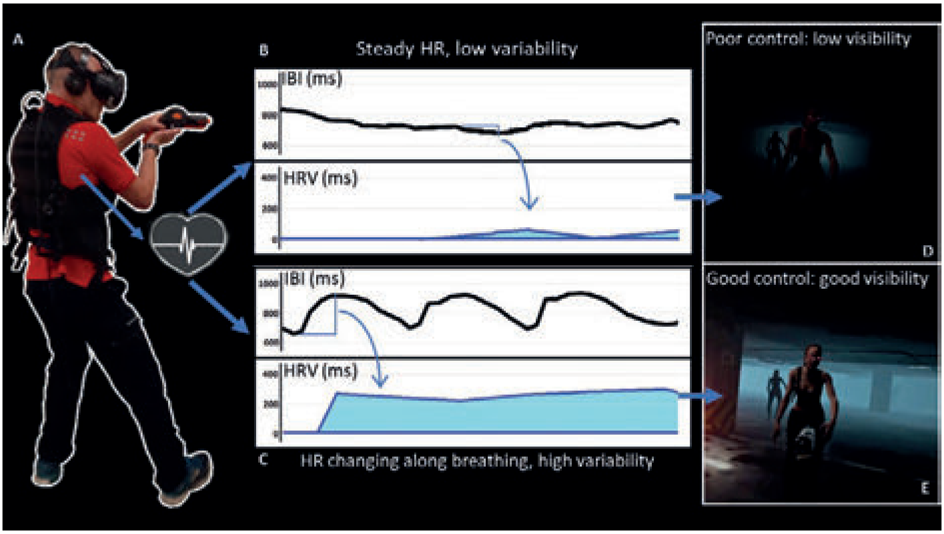
We recently developed a VR game for BF training in an active decision-making context in collaboration with the Dutch police (Brammer et al., 2021; Michela et al., 2022). The VR game called DUST (Decision Under Stress Training) draws inspiration from the popular genre of zombie shooter games which, even though they contain highly unrealistic narratives and stimuli, elicit high engagement and arousal that could be related to experiences in real policing situations, thus potentially leading to increased transfer (Kothgassner & Felnhofer, 2020). Initial proof-of-concept data from a series of case studies indicated the feasibility of training breathing control for police officers with DUST in a manner that was considered highly appealing, as well as its suitability to objectively assess impulsive and controlled behavior (Michela et al., 2022). Here we comprehensively test an adapted version with shortened VR sessions, a shorter training schedule, and with a feedback algorithm directly rewarding HRV instead of breathing pace to maximize training efficacy. This training targets local power HRV (Bornemann et al., 2019) which rewards respiratory fluctuations in HR, a widely used index of relaxation (Hayano & Yuda, 2019; Lehrer et al., 2020; Yasuma & Hayano, 2004), in setup that provides a high-level resistance to movement artifacts (Gilgen-Ammann et al., 2019).

In the present, pre-registered study (*https://osf.io/cdsbx*) we aimed to comprehensively test in a large police sample whether DUST can help in voluntarily increasing HRV in arousing action contexts, and most importantly whether this skill transfers to a professionally relevant real-life action context. To the aforementioned core game dynamics of engagement and arousal, we added mechanics representing psychological processes known to be impaired by stress-induced arousal: Go/No Go decision-making for response inhibition (Föcker et al., 2018; Hashemi et al., 2019), a priming task to assess bias resistance (Johnson et al., 2018; Taylor, 2020), and a

BF training mechanic for psycho-physiological regulation (i.e., feedback for HRV regulation; Hayano & Yuda, 2019; Laborde, 2019).

Figure 5.1

Illustration of biofeedback (BF) to train Heart-Rate Variability (HRV) under high arousal in DUST.



Note. The figure depicts the implementation of BF as closed-loop peripheral vision modulation to reflect the negative consequences of attentional narrowing (tunnel vision) occurring under stress, a relatable phenomenon for police officers: (A) A police trainer in the VR game-context representing an underground parking lot, with zombies approaching them in waves. Radio messages instruct the participants before each of the waves about the characteristics of the incoming “hostile” zombies to be shot at (Go targets), while other “benign” zombies are to be spared (NoGo/non-targets). Critically, the game additionally reacts to the real-time physiology of the participants, by restricting their field of view when HRV is low. (B) Example traces of inter-beat-intervals (IBI) and associated HRV when the participant’s HR is stable and does not fluctuate along breathing, (C) IBIs and associated HRV traces when the participant’s HR is in coherence with deep breathing (accelerating with inhalations and decelerating with exhalations); The in-game visibility was linked to the BF score, calculated from the local power HRV computed by measuring peak-to-through differences in inter-beat intervals, where smaller fluctuations (less coherence) correspond to a lower HRV score and worse visibility (D) while large HR fluctuations (high HRV) were associated with good visibility (E).

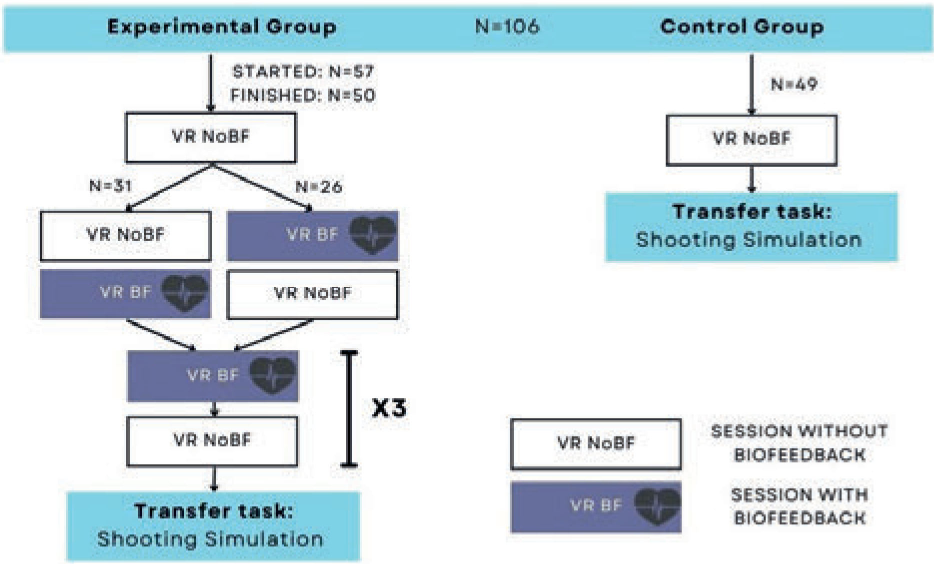
The study has three concrete objectives. First, to validate DUST as a believable and arousing virtual environment evidenced by significant increases in HR and the expected in-game behavior (Brammer et al., 2021; Michela et al., 2022). Second, as the subjective experience is a critical determinant of training motivation, we test the efficacy of the BF to improve HRV within the game, evaluate how the game is perceived and how it impacts self-efficacy and physiological awareness. As improvements in

HRV over time could be a simple consequence of repeated exposure to the game rather than BF, we assess the causal role of the BF by addition and withdrawal of the BF component in the game, and by manipulating between participants when the BF was introduced (introduced in session 2 for half of the participants; session 3 for the other half). Finally, and most critically, to measure if the acquired voluntary HRV control transfers to an independent context, we test whether the HRV control would transfer to a professional action context outside the VR game.

To this end, 109 police trainers were recruited from training centers across the Netherlands and pragmatically randomized (Di Nota et al., 2021) to an experimental group and a passive control group (See Figure 2). The experimental group participated in a 3-to-5-day training with a total of 9 short (5- to-7 minutes) individual VR game sessions, while the control group was only exposed to the VR game once without the BF component. Both groups took part in a police-relevant assessment task (shooting simulation) aimed at evaluating the benefits of the training in terms of skill transfer.

Figure 5.2

Participants flowchart and design outline.



Note. The experimental group received 9 VR sessions (in 3-to-5 days), alternating 4 sessions with online HRV BF and 5 without, while the control group only received 1 session, without BF. In the experimental group, half of the participants were randomly selected to receive BF for the first time in the second VR session, whereas the other half received BF only in the third session. This delayed introduction of BF is used to test the causal relation between BF introduction and HRV increases. After the training, both groups performed a police-relevant transfer task to verify whether physiology and behavior were improved by the training.

RESULTS

Objective 1: VR game validation

First, we evaluated whether the game produced the intended increase in arousal, evaluated as an *in-game* increase in HR from baseline. Across both the experimental and the control groups, in the first session of playing the game, *in-game* HR (M = 93.39, SD = 17.84) increased robustly when compared to baseline measures (M = 80.04, SD = 13.97; $t(83) = -14.554$, $p < 0.001$; see figure 3A). Also, across all successive sessions for the experimental group, mean in-game HR was highly significantly increased compared to baseline both in sessions with and without BF (N = 57, B_{bf} = 10.57, 99.9% CI [7.43, 13.70]; B_{nobf} = 9.91, 99.9% CI [6.97, 12.82]; further details see Supplementary Materials A). Despite the robust increase in psychophysiological arousal, police trainers rated the experience as mildly stressful (M = 3.54 on a 7-point scale), as a priori expected based on previously reported tendencies in this group to underreport experienced stress (Habersaat et al., 2021).

Given the arousal induced by our game, we expected difficulties in the inhibition of automatic response tendencies. To test this, a priming element was added to the game by means of radio dispatch information. The radio message preceding each zombie wave contained two pieces of information to describe target zombies (e.g., “Target zombies will have red eyes (1), and we expect them to be large males (2)”). While the first information (eye color) was always accurate but hard to identify, the second piece (morphology) was easily recognizable but became less accurate as the game progressed. This information pairing was expected to lead to an increase in false positive responses against zombies that had the primed body type, but a different eye color (and were therefore actually not to be considered hostile targets). While there was no difference in the first session between false alarm rates for primed and unprimed non-targets ($t(110) = -0.52$, p = 0.604; Mprimed = 1.65, SDprimed = 1.82, Munprimed = 1.75, SDunprimed = 1.617), later sessions (2 to 9) showed that false alarms were on average higher for the primed non-targets ($B_{priming}$ = -0.56, 99.9% CI [-0.91, -0.23]; Mprimed = 1.43, SDprimed = 0.96, Munprimed = 0.86, SDunprimed = 0.61). Thus, our game produced the expected increases in arousal and required participants to inhibit primed pre-potent responses.

Objective 2: Effectiveness and appraisal of the training

Every participant of the experimental group started with a session of gameplay without biofeedback to get accustomed to the gameplay. When considering the sessions after the first try-out (i.e., sessions 2 to 9, see Figure 3A), presentation of the BF consistently resulted in higher HRV (regardless of whether it was introduced in session 2 or 3; *N* = 57, across sessions $B_{BF\_vs\_nobf}$ = -5.42, 99.9% CI [-10.03, -0.43]). As the total number of BF (total of 4) and non-BF (total of 5) sessions was unequal, rather than investigating interactions we preregistered to additionally run separate analyses per BF condition. From the moment BF was introduced HRV remained high from the first to the last BF session, perhaps reflecting the experience of the instructors with HRV biofeedback. Most interestingly, the ability to upregulate HRV in action in the absence of BF developed more gradually (see Figure 3A), with a robust increase between the first and the last non-BF training session ($B_{S9\_vs\_S1}$ = 11.39, 99.9% CI [2.10, 21.08]). Overall, the in-game HRV of the experimental group went from 39.77ms in the first session to 52.67ms in the last session, a 32% increase.

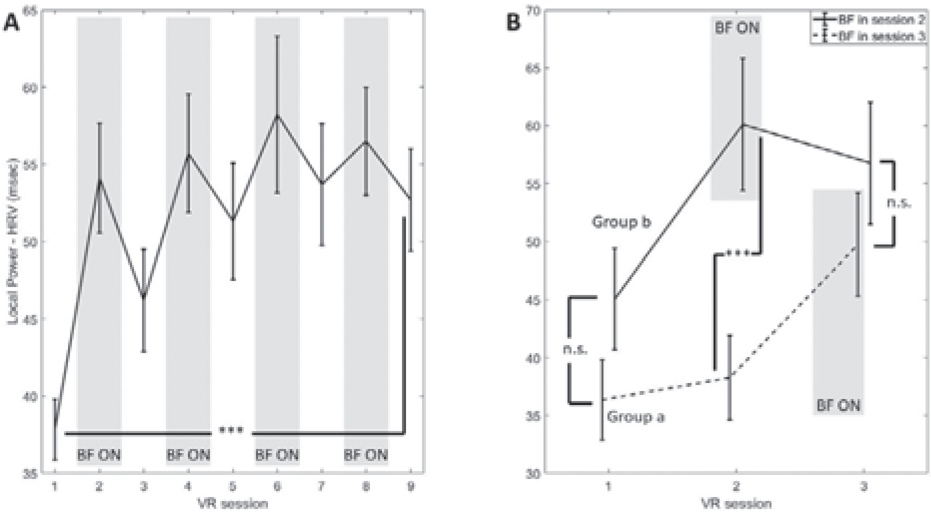
To assess the causal influence of BF on the HRV, and to rule out that the observed increases in HRV were chiefly caused by repeated exposure to the game rather than BF presentation, we subsequently compared sub-groups for whom BF was introduced at different sessions (group a: session 3 vs. group b: session 2; see Figure 3B). Confirming our preregistered hypothesis that the BF drives HRV upregulation, we found that in session 2 the introduction of the BF in group B significantly increased HRV compared to group A that played this session without BF (*M*a = 38.27, *SD*a = 19.30, *M*b = 60.08, *SD*b = 26.77; $t(49) = -3.347$, $p = 0.002$). We subsequently verified that the significant group difference was not present in session 1 where both groups played without BF ($t(42) = -1.574$, $p = 0.123$), nor in session 3, after BF was introduced also in group A ($t(50) = -1.016$, $p = 0.314$). Thus, our results suggest that changes in HRV are causally induced by BF integration in the game.

After having established the expected BF training-induced HRV-increase, we tested how the training impacted the evolution of physiological awareness. Subjective awareness of breathing increased robustly from the first to the last session (*N* = 57, $B_{S\_9vsS\_1}$ = 1.99, 99.9% CI [1.05, 2.98]). Also, awareness of heart rate increased substantially throughout the training (*N* = 57, $B_{S\_9vsS\_1}$ = 1.19, 99.9% CI [0.21, 2.16]). Although no significant increase was found for self-efficacy, which was contrary to our expectations, it was lower in BF sessions ($B_{BF\_vs\_NoBF}$ = -0.36, 99.9% CI [-0.69, -0.05]) indicating that the BF signal may have reminded police trainers that the in-game HRV self-control was challenging. However, the training was perceived positively by the police trainers as evidence by their post-training ratings of engagement, usefulness and efficacy (on 7-point scales, useful *M* = 5.72, efficacious *M* = 5.41 and engaging *M* = 4.9), all robustly above neutral (tested against preregistered reference value of 4 =

neutral anchor in the Likert scale; *t*engage(52) = 10.76, *p* < 0.001; *t*use(52) = 13.42, *p* < 0.001; *t*eff(52) = 8.49, *p* < 0.001). Illustratively, 80.76% of the police trainers responded positively to the question whether they would like to use this game in their own training program.

Figure 5.3

Evolution of the Mean Heart-Rate Variability (HRV) in the Experimental Group

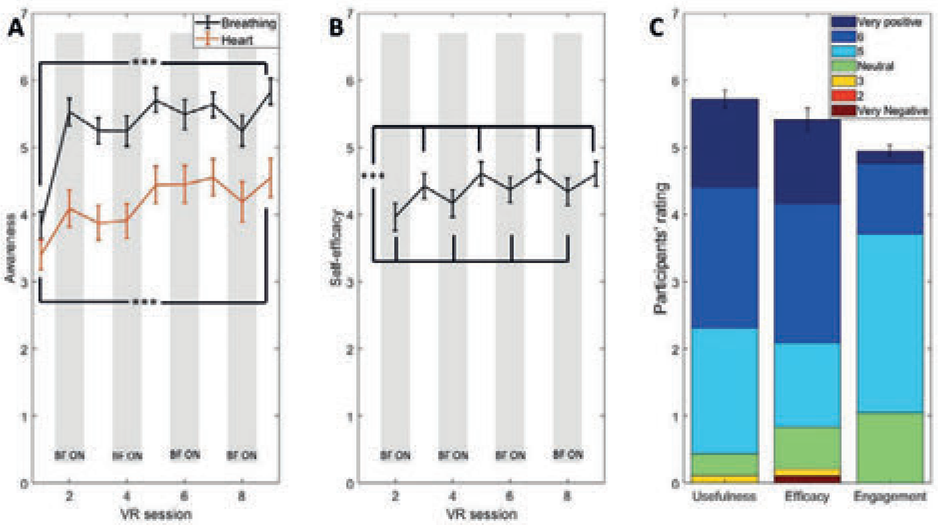


Note. (A) across time and condition for the entire experimental group; (B) for the 3 first VR sessions, where BF was introduced in the second session for half of the experimental group (labelled as group b, N = 26) and in the third session for the other half (group a, N = 31). The results show that BF consistently led to increased HRV over the BF sessions, which gradually transferred to the non-BF sessions (A), and that the HRV increase after the first session is causally related to the moment of BF introduction (B); BF ON = session with online BF presentation; Error bars represent standard error of the mean.

Next, we tested whether the effort of focusing on HRV increases did not lead instructors to neglect behavioral performance. As illustrated in Figure 5, the sustained improvements in HRV observed in Figure 3 went together with improvements in behavioral performance when comparing the first and the last VR session, both in FA reduction ($B_{S9\_vs\_S1}$ = -1.49, 99% CI [-2.83, -0.20]) and spatial awareness (unspotted targets; $B_{S9\_vs\_S1}$ = -1.09, 99% CI [-1.95, -0.20]). General shooting behavior, measured as the d’ sensitivity index from signal detection theory (McFall & Treat, 1999) did not change significantly from first session, but exhibited a positive trend ($B_{S9\_vs\_S1}$ = 0.16, 90% CI [0.01, 0.30]). While the improvements in behavior suggest that participants’ performance may have benefitted from improved HRV self-control, there was no correlation between session-by-session behavioral and physiological changes (d'; B_{HRV} = 0.04, 95% CI [-0.07, 0.15]; FA; B_{HRV}= -0.09, 95% CI [-0.68, 0.52]; unspotted targets; B_{HRV}= 0.08, 95% CI [-0.33, 0.48]).

Figure 5.4

Police trainers' appraisals and perception of the training



Note. Evolution of the (A) interoceptive awareness (breathing and HR) and (B) the self-efficacy, across time and condition for the entire experimental group; (C) Rating of the experimental group for the perceived usefulness and efficacy of the training, as well as the elicited engagement; Colors indicate the distribution of participants' responses and indicate the perception was overwhelmingly positive. BF ON = session with online BF presentation; Error bars represent standard error of the mean.

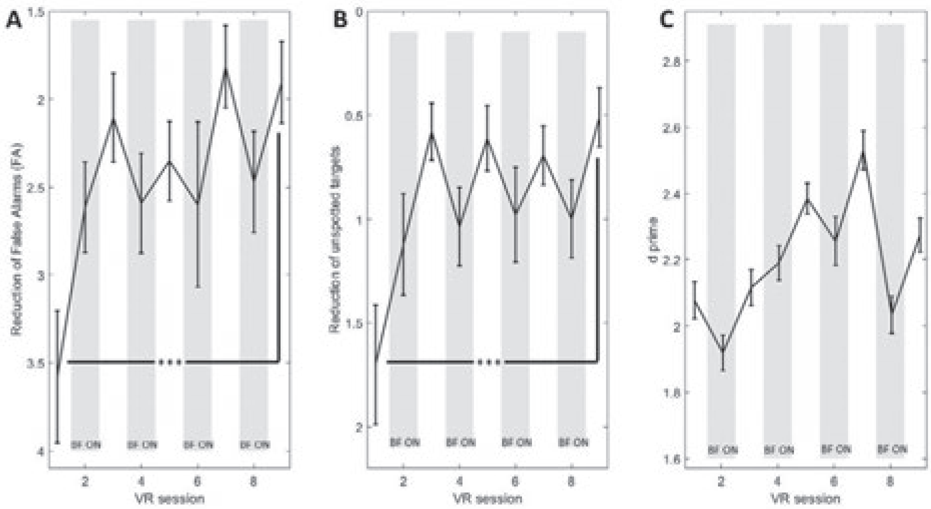
Objective 3: Skill transfer

Finally, and most critically we tested whether the ability to voluntarily upregulate HRV demonstrated in the game transferred to an independent, realistic professionally relevant assessment outside VR (see Figure 6 panel A). As only the first trial could be used for behavioral assessment (subsequent trials were distractors), this task was not optimized for evaluating the effect of the training on shooting tendencies and no difference in behavioral performance between the control and experimental groups was apparent ($X2$ (1, N = 100) = 0.09, p = 0.764; percentage Go responders per group 40.4% control group; 43.4% experimental group; for details see Supplementary Materials A). Also, during the transfer task, physiological arousal of both groups (HR) did not significantly differ (t(81) = -0.240, p = 0.811; Mexp = 87.04 bpm, SD = 13.94, Mcontr = 88.16 bmp, SD = 14.88) while staying significantly higher than baseline, thus indicating elevated arousal at test ($t(74) = -8.360$, $p < 0.001$; *M*transfer = 87.04 bpm, *SD* = 13.94, *M*base = 80.424 bmp, *SD* = 13.48). Critically, the experimental group did show significantly higher HRV already during the baseline before the transfer task ($t(94) = 2.106$, $p = 0.038$, $M_{exp} = 68.63$ ms, $SD = 36.82$, $M_{contr} = 55.36$ ms, $SD = 23.78$) and more importantly HRV was also robustly higher during the transfer task, as preregistered ($t(81) = 2.986$, $p = 0.004$; $M_{exp} = 58.77$ ms, $SD = 24.25$, $M_{contr} = 44.35$ ms, $SD = 20.33$; see Figure 6 panel B). To ascertain that this effect was not due to pre-existing group differences, we compared HRV between groups during the baseline of the first session (i.e. before the experimental group was trained). As expected, there was no significant difference between the groups at this time (day 1; $t(102) = -0.570$, $p = 0.57$, $M_{exp} = 57.94$ ms, $SD = 22.26$, $M_{contr} = 55.36$ ms, $SD = 23.78$). The critical group difference during the transfer task also remained significant when controlling for pre-training HRV levels (session 1 baseline; $t(72.88) = -3.224$, $p = 0.002$; corrected for variance inequality). Together, our results support the conclusion that the experimental group showed increased HRV in this assessment as an effect of our VR training.

5

Figure 5.5

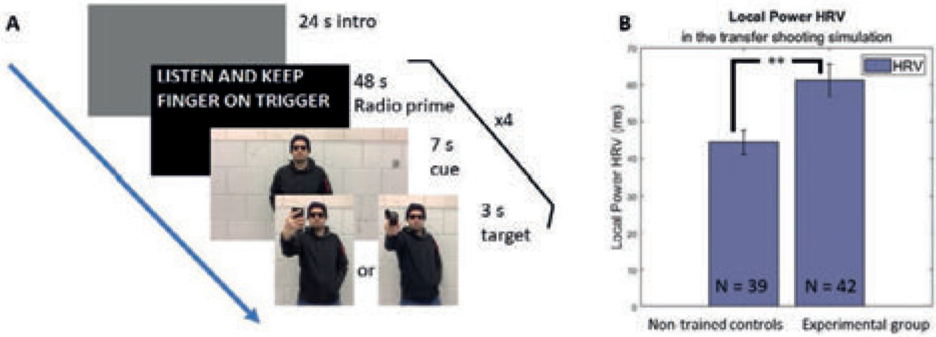
Evolution of the main behavioral metrics



Note. Evolution of the main behavioral metrics across time and condition for the experimental group; (A) False Alarms (FA); (B) Unspotted targets refers to the spatial awareness aspect of the behavioral performance, specifically the number of targets that were able to reach the player without them turning in the target direction; (C) Evolution of the player's target sensitivity (*d'*); the phasic drop in *d'* in the 8th session is due to an unexpected increased number of false negatives; BF ON = session with online BF presentation; Error bars represent standard error of the mean.

Figure 5.6

HRV local power in the independent professionally relevant (non-VR) transfer task



Note. (A) transfer task design; Intro: participants were instructed that they were in a shoot/ don't shoot decision task and asked to listen to radio dispatch; Radio prime: A realistic police radio dispatch primed the participant by describing a violent perpetrator while text asks to keep weapon pointed at screen; Cue: a target that matches the description appears; Target: target draws a phone from their pocket; Response: participant is required to withhold (No-go) on first trial (where a phone appears), or shoot in the next 3 trials where a gun appears; (B) Absolute Local Power HRV (difference in ms) between the non-trained control and the experimental groups for the critical preregistered comparison during the transfer task; Error bars represent standard error of the mean.

DISCUSSION

This preregistered randomized controlled trial among 106 police trainers provides the first evidence that HRV-BF training in an arousing VR action game can be used to (1) boost voluntary HRV upregulation in professionals with improved concurrent decision making under threat and (2) enables skill-transfer to both an in-game session without BF and an independent professionally relevant real-world testing context. By systematically varying the moment of BF introduction, we further provided evidence that the increases in HRV were causally linked to our in-game BF presentation. Furthermore, the game produced substantial increases in self-reported physiological awareness and was rated a useful, efficacious and engaging training tool by a large majority of the police trainers assessed.

Our contextualization of the biofeedback with an arousing action game format contrasts with other (VR) biofeedback approaches that have typically trained stress resilience in calming virtual environments (Blum et al., 2019; Lüddecke & Felnhofer, 2022; Weibel et al., 2023). DUST induced substantial HR increases during gameplay, in the same order of magnitude as found in established stress-induction protocols (Boesch et al., 2014; Vogel et al., 2015) and fear induction in VR (Lin, 2017). The use of real-time in-game biofeedback allowed players to recognize stress-induced reductions in HRV and at the same time motivated them to upregulate their HRV. Namely the biofeedback blurred their vision, impacting their action performance. Upregulation of HRV in this active action-decision context is more challenging than in typical HRV-BF trainings at rest, when participants are not concurrently engaged in action. The difficulty of the resulting ‘dual-tasking’ in BF conditions was reflected in the police trainers’ consistent reports of a reduced feeling of self-efficacy in BF versus non-BF sessions, even when physiological awareness steadily increased with training. The latter reflects that the police officers were in fact learning *in-action* psychophysiological awareness, a critical element to maximize the chance on skill transfer to real-life situations in which police officers are required to control their psychophysiological stress responses, particularly in stress-full situations requiring speeded action decisions. Interestingly, the increases in HRV induced by the present BF training are comparable in magnitude to the BF training-induced changes in seated and non-active setups (Bornemann et al., 2016). This is noteworthy as our participants were experienced police trainers, previously trained with more traditional passive biofeedback and therefore may have been expected to show already strong HRV regulation skills from the start. In sum, we showed that contextualizing biofeedback by adding arousal, movement, and active decision-making does not prevent the learning of voluntary HRV upregulation in action, and provides benefits even in participants experienced with HRV biofeedback.

Besides the goal of making the training enjoyable and challenging, the most important aim of the arousing action context was to promote transfer to arousing situations outside of the game. Our results showed transfer of HRV upregulation not only to a context without BF within-game (near transfer), but also to an independent non-VR task (far transfer). So far, no HRV-BF studies have reported transfer of HRV upregulation skill outside of the training environment (Lüddecke & Felnhofer, 2022). These results extend previous literature indicating that HRV biofeedback in passive and calm settings can enhance HRV control (De Witte et al., 2019; Lehrer et al., 2020; Yu et al., 2018). While studies with such non-immersive biofeedback can already lead to significant benefits beyond the training setting such as increased physiological control and improved decision-making under stress (Andersen et al., 2018; Andersen & Gustafsberg, 2016), immersive VR based biofeedback can offer advantages beyond this, by enhancing motivation (Lüddecke & Felnhofer, 2022; Weibel et al., 2023) and gamification which has been shown to aid transfer to real-world settings (Jerčić & Sundstedt, 2019; Makransky & Petersen, 2021; Parnandi & Gutierrez-Osuna, 2017, 2019). Importantly, the reported difference in HRV between the experimental and the control group cannot be attributed to a larger familiarity of the experimental group to the research setup, since the transfer task was new to both groups. A notion that was further supported by the absence of group differences in arousal (absolute HR) during the independent transfer task, despite the transferred HRV differences.

Previous studies have shown that higher levels of HRV are linked to a wide range of health and performance benefits (De Witte et al., 2019; Lehrer et al., 2020; Yu et al., 2018). While we observed improvement in in-game perforamce we could not observe any behavioral impact of the training on the transfer task, possibly resulting from the fact that our transfer task was not optimized for assessing behavioral differences and that these analyses were based on a single trial. Indeed, HRV trainings for police and military personnel that proved to be sensitive for measuring behavioral benefits were more complex simulations, also involving verbal communication (Andersen et al., 2018; Bouchard et al., 2012).

Whereas, non-game-based BF trainings can obtain negative user experiences (i.e. (van der Meulen et al., 2018), game-based approaches have been shown to enhance motivation (Schoneveld et al., 2019), a factor that has been identified as critical to foster change (Holzinger et al., 2006). In line with this notion, our VR game was still rated as engaging after nine training sessions. Indeed, standard HRV trainings in passive sitting contexts are not always considered enjoyable, particularly in police officers (van der Meulen et al., 2018). We speculate, that the relative playfulness of our training game prevented this negative reception that has been linked to difficulties federating police trainers around a common learning goal (Bennell et al., 2021; Eliasson, 2021). This notion is further supported by the finding that 80% of the trainers indicated they would adopt our VR BF game training in their own teaching.

The traditional randomized controlled trial design can be used to assess the causal influence of an intervention, but usually cannot give a mechanistic account of what, in the intervention, drives the effect (Deaton & Cartwright, 2018). Our training schedule circumvents that limitation by drawing inspiration from designs commonly used in case reports. First, we used the delayed introduction of BF for part of the experimental group, as used in the multiple baselines design (Levin et al., 2021). This design allowed us to further assess the causal role of BF presentation on HRV increases, as increases only happened after the introduction of BF. Second, we adopted an addition and withdrawal design (Kratochwill et al., 2013; Manolov et al., 2017) by alternating BF and non-BF sessions, which allowed us to further strengthen the causal claim centered around BF presentation as its presence was linked to higher levels of HRV control.

Some limitations should further be discussed when evaluating these findings. First, while our study provides important new evidence regarding the transfer of HRV training to a new context it would be important to establish also long-term effects on real-life policing outside a training context, ideally in simulated scenarios with actors (Andersen et al., 2018; Bouchard et al., 2012) or even on duty using wearables to asses psychophysiological arousal. Additionally, although the training had a significant impact at the group level, a minority of participants showed minimal improvements from the training and the current study did not assess potential individual predictors of training efficacy. As suggested by Weerdmeester et al., (2020), psychological dimensions such as growth mindset could have a major impact on the learning trajectories of participants and could therefore be included in future studies to further establish how the efficacy of in-action biofeedback training could be further improved. Particularly interesting in this respect are also computational approaches that formalize distinct aspects of biofeedback learning and could therefore provide a better understanding of the mechanistic background of individual variation in biofeedback's efficacy (Lubianiker et al., 2022).

While we have tested here the efficacy of our game in a group of police trainers, our approach of training HRV control under arousal could potentially also be relevant in different populations that suffer from negative consequences of stress. Indeed, high arousal during stressful events has been repeatedly linked to long-term trauma symptom development (Hinrichs et al., 2019; Morris et al., 2016; Schultebraucks et al., 2021), possibly reflecting hyper-activation of the amygdala and the under-activation of the prefrontal cortex (Admon et al., 2009, 2013; Kaldewaij et al., 2021; Roeckner et al., 2021). As passive forms of HRV training have already been shown to alleviate symptoms in people suffering from anxiety and depression (Goessl et al., 2017; Pizzoli et al., 2021), HRV training under arousing conditions may be especially useful for preventive efforts and our gamified biofeedback could provide motivational benefits also in other groups.

To conclude, this study presents a novel training method using a BF game in VR to help police officers cope with stressful environments while in action. This training was found to be effective in fostering HRV upregulation even during high arousal and in action. Importantly, it was also highly appreciated by the police trainers who underwent the training as participants. As police trainers are a population usually known to be critical towards innovation (Braga & Weisburd, 2019; Lingamneni, 1979), this provides promise for the adoption of this technological intervention also for other populations.

MATERIALS AND METHODS

Participants

In this pre-registered study (https://osf.io/cdsbx), participants were 109 police trainers aged between 30 and 60 years (94 males, 15 females). Most of the participants (N=64) indicated little (N=17) to no (N=47) familiarity with VR, while 29 others indicated a higher level of familiarity with VR. Participants were all trainers recruited via internal advertising from Dutch Police Training centers or IBT ("Integraal Beroepsvaardigheids Training centrum") as well as from the police academy, which consists of several geographically dispersed locations throughout the country.

We designed our study as a quasi-random pragmatic control trial (Di Nota et al., 2021): Police trainers were assigned to an experimental or a control group, based on their availability. The two groups did not differ in terms of VR experience (t(108) = -.418, p = .677) nor HRV at baseline (t(102) = -.570, p = .570). Both groups took part in an independent task at the very end of the experiment, referred to as the "transfer task". The description of the task can be found below in the Materials section.

Participation was voluntary and coordinated by the managers of the various IBT centers. According to the rules of the Dutch Police regarding research, financial compensation of the police trainers functioning as participants was not allowed. Therefore, a donation of 25 euros for each participant of the experimental group and 5 euros for each participant of the control group was allocated to a fund for the training of "PTSD dogs" (Actie ZeeHond, n.d.). The research procedures were approved by the ethical committee of the Faculty of Social Sciences of Radboud University Nijmegen (ECSW-2021-017). All participants provided informed consent in writing prior to participating in the study, in line with the guidelines of the Declaration of Helsinki (WMA, 2018).

Procedure

The full experimental procedure is illustrated in Figure 2. The experimental group took part in a nine-session training, spread over different days, spanning about one month. Each session consisted of a VR session, alternating between a session without

BF and one with BF. This alternation was used to both promote transfer (the player still received an overall biofeedback score at the end of each session) and to assess transfer effects. Furthermore, to be able to verify whether changes in physiology in the experimental group are related to the BF component of the game, in session 2 the BF component was introduced to half of the participants and only in session 3 to the other half of the participants (see Figure 2). To enhance learning of HRV upregulation skills, both sub-groups had a very short external BF training session in front of a laptop right before the in-game BF component was introduced to half of them in session 2. In line with the recommendations for neurofeedback research (Ros et al., 2020), the control participants were passive controls and only played one VR session in the same game, but without any BF, during or after the session.

Materials

Physiological Recordings

Participants' HR were measured using a Polar H10 heart-rate sensor, where the HR corresponds to the time between consecutive R-waves of the QRS complex. It has been shown before that this HR sensor reliably extracts R-R intervals, even under intense physical activity (Gilgen-Ammann et al., 2019).

The Virtual-Reality Material and Game

The VR equipment used was an HTC Vive setup, with one of the two controllers wrapped by a 3D printed case used to give the controller the approximate shape of a gun, and the exact weight of the Walter P99 QNL gun that is used by the Dutch police. The other controller was used as a dispatch-radio controller and attached to the participant's vest. A thorough description of the game and design choices can be found in our theoretical paper (Michela et al., 2019). The VR game was adapted after a first feasibility study (Michela et al., 2022). The list of adaptations can be found in the Supplementary Materials B. Briefly summarized, the VR game mechanism was designed to resemble commercially available zombie shooter VR games. At the beginning of the game, the player was teleported to the center of a large parking lot, and instructed to "protect" the location against zombies announced as aggressive by radio messages. In each VR session, 5 to 6 zombie waves approached the player, each one preceded by a radio message announcing to the player which zombie type should be shot (hostile zombies that attack the player once in range) and which one should be spared (benign zombies that dissolve after reaching the player). The hostile/benign ratio varied between VR sessions, but was kept around a 65/35% ratio as it maximizes the chances of false alarms (Young et al., 2018), and therefore mitigates risks of ceiling effects in go/no-go performance. The radio dispatch contained two pieces of information to identify hostile zombies (e.g., “Shoot only the zombies with red eyes, we expect them to be large males”). The first part of the dispatch information was the eye color of the hostile zombies (red, yellow or blue). It allowed to accurately identify hostile zombies, but was hardly visible at a large distance. The second part of the dispatch information was the body type of the hostile zombies (male/female, small/large). This information was visible from a large distance, but was less reliable, thus priming the player to shoot benign zombies. Three different variations of the VR scenario were used in the game, distinguished by a task-irrelevant stressor comprised by a loud noise that was varied to increase unpredictability (glass shattering noise, car alarm, and fire alarm). Those events were only used to add some tension and variation in the scenarios. Participants encountered each variation at least once, as described in Figure 2.

Biofeedback Parameter and Implementation

Breathing-induced fluctuations of inter-beat-interval show good resistance to movement artifacts. They were calculated by means of local-power HRV (Bornemann et al., 2019). The online calculation of this HRV score was done for each one of the nine VR sessions in which a participant took part. However, in the sessions in which BF was displayed to the player (AABBABABABA or ABABABABABA depending on the experimental subgroup; withdrawal design, A = without BF; B = with BF; see Figure 2; the two subgroups are referred to as A and B respectively in the result session, based on presence or absence of BF in session 2), higher local-power HRV was rewarded by unimpaired vision in the VR game, whereas lower local-power HRV was progressively punished by reducing the player’s field of vision (see Figure 1). In the sessions without BF, the vision of the player was not modulated based on HRV, and therefore always unimpaired.

The local-power HRV was calculated with the Python-coded “OpenHRV” program (Brammer, n.d.), which extracts peak-trough differences in a 15 second sliding window of the inter-beat RR intervals signal that were recorded by the H10 polar belt. The BF score, varying between 0 and 1, was then calculated based on the Local-Power HRV. The initial target for the HRV was set to peak-troughs differences of 100ms (which corresponds to the local-power HRV metric), but could be adapted for each participant in the game to maximize learning. With the standard target, a local-power HRV of 100ms and above would lead to a BF score of 1 (maximal visibility for the participant in the VR experience). The score would then linearly decrease to 0 when the local-power HRV was 0ms, leading to a severe visual impairment for the player in the VR game, where the tunneled vision would restrict the peripheral vision to the point where only zombies directly facing the player would be visible.

Questionnaires

All the questionnaires were administered by using the LimeSurvey platform (LimeSurvey & Schmitz, 2012) on a laptop directly before or after the VR game sessions.

Engagement questionnaire. The engagement was measured once at the end of the full training with a four items questionnaire (McAuley et al., 1989) on a 7 points Likert scale (1=Strongly disagree, 4=Neutral, 7=Strongly agree; item 2 was reverse-coded). The final score was obtained by averaging the answers on all items.

Physiological awareness questionnaire. Physiological awareness was measured after every VR session, by mean of a two items questionnaire asking how aware participants were of their breathing and of their HR during the VR task. Each question could be answered with a 7 points Likert scale (1=Strongly disagree, 4=Neutral, 7=Strongly agree).

Exit questionnaire. At the end of the full training, the experimental group received an exit questionnaire aimed at evaluating the degree of satisfaction of the training (e.g., "Would you use this VR environment in your teaching?") and the subjective experience in the game (e.g., "How stressed were you in the VR environment?"). Each question could be answered with a 7 points Likert scale, apart from open ended questions. The control group received a shortened version of the exit questionnaire. The full list of questions can be found the Supplementary Materials A. Two subsets of the questions, pertaining to the experienced usefulness and efficacy of the training, were averaged to calculate training usefulness and efficacy scores. One item for the training usefulness score was reverse coded ("The training was too long to be incorporated in the curriculum of police students").

Transfer Task

To assess if the training in the VR game would carry over to relevant policing behavior outside the game, a police-relevant task was designed drawing inspiration from the dispatch-priming paradigm of Taylor (2020). The original task consists of a single shoot-don't shoot decision made while holding a gun pointed toward a target taking an object out of their pocket (a gun or a phone), projected on a screen with a beamer. The decision moment was preceded by a radio dispatch message describing the appearing target as either an armed and violent opponent (priming the participant to shoot), or as an innocent passer-by. In the original task by Taylor (2020), participants shot an innocent target taking a phone out of their pocket more frequently (34% more) when primed for a violent perpetrator. In our version of the task, the radio message always primed the participant for a violent perpetrator, and the target always drew a phone out of their pocket in the first (critical) trial. When piloting the task, we realized that Dutch police trainers never shot the subject, as no one had their finger on the trigger, and most participants stepped aside before shooting based on their training. We hence had to modify the task by asking the participants to keep their finger on the trigger of their gun, and stay immobile (they were constrained by chairs surrounding them, so that stepping aside was not possible). Additionally, since the experiment had to be carried out in police institutions where participants could potentially communicate about the transfer task, 3 distractor trials in which a go response was asked were added after the first critical No-go trial. In those trials the subject appearing on screen took a gun instead of a phone out of his pocket, and held it for 3 seconds before the next trial. In these trials the correct response was to shoot the opponent. Data from these trials were not analyzed.

Based on the study by Taylor (Taylor, 2020), we expected the effect of our training on the transfer task to require a sample of N = 77 (df = 1, alpha=0.05, Power=0.8). This calculation was based on an effect size of w = 0.32, calculated from the effects reported in the study by Taylor (2020). In the study by Taylor et al., primed participants shot the innocent target in 62% of the cases, compared to 28% for unprimed participants. In our case, all participants were primed, but the experimental group was expected to be more aware of the adverse effects of dispatch priming, as this was also a factor in the VR game, on top of showing better behavioral control due to their more optimal physiological state, induced by the BF training. Based on that reasoning and some piloting, the effect size for the training was calculated from a contingency table assuming a 40% shooting rate for the experimental group (hence 22% lower than the original 62% for controls), hence considering that our training would have a moderate protective effect from priming.

Data Preparation

Physiological recording

The HR of the participants, measured with a Polar H10 belt throughout the sessions was analyzed separately for baselines (recording preceding the odd numbered VR sessions), in-game VR sessions (with and without BF) and during the transfer task following the last session. HR in beats per minute were used both in absolute values as well as relative changes from baseline (baseline-corrected HR), in which case the baseline value was subtracted from the in-game value. Prior to feature extraction, the data was cleaned automatically to remove artifacts with the Python software Biopeaks (Brammer, 2020), which used the artifact correction for HRV timeseries proposed by Lipponen & Tarvainen (2019). Mean and median HR were extracted for each condition. Finally, the mean and median Local Power HRV (Bornemann et al., 2019) were calculated per condition (with or without BF). For each of those sessions, frequency-based HRV metrics were not extracted if the recording was shorter than 5 minutes, and Local Power HRV was not extracted for recordings shorter than one minute (Bornemann et al., 2019).

Decision-making and monitoring

For each VR session, the behavioral decisions of the player were extracted. The number of hostile zombies killed before reaching the player represented hits, while hostiles who reached the player before being shot were the false negatives. Benign zombies reaching the player were correct rejections, and if shot by the player they were false alarms. Those four sums were then used to calculate in each VR session the sensitivity {d' = [z(Hit rate)−z(False alarm rate)]} according to signal detection theory (McFall & Treat, 1999). For the monitoring aspect of the in-game performance, an index was extracted calculating the number of "unspotted targets" (both hostile and benign), which are zombies reaching the player without ever appearing in their field of vision, therefore reaching the player unspotted. The rationale for including benign in the calculation was that not spotting an approaching individual (irrespective of their intentions) is always a tactical mistake in real life policing.

Questionnaires

The engagement and physiological awareness questionnaires were scored as described in the materials section. For the exit questionnaire, the individual items were analyzed individually.

Data Analysis

The following section describes the pre-registered hypotheses (https://osf.io/cdsbx) as well as additional exploratory analyses. The main reason that our analyses deviate from the pre-registered analyses is because the dataset was not complete due to attrition and poor data quality for some participants (see Supplementary Materials C, Table S5.1, for a detailed list of hypotheses modification). Hence, repeated-measures ANOVAs with list-wise deletion resulted in too much data loss. Therefore, all repeated measures analyses were replaced by Bayesian mixed-effects models, computed in R (Version 3.5.1; R Core team, 2016) using RStudio (Version 1.4.1717; RStudio Inc., 2009–2021) with the brms toolbox (Version 2.17.0; Bürkner, 2017; Carpenter et al., 2017). In these models the effect of time was investigated by contrasting the first and the last of the VR sessions. All categorical predictors were coded using sum-to-zero contrasts, and continuous predictors were zero-centered. As the data contains repeated measures, the models included random intercepts and slopes per participant for all relevant predictors. Interactions and full models' descriptions can be found in Supplementary Materials D. We fitted the models using 4 chains with 15000 iterations each (6000 warm-up). Statistical "significance" was derived from 95% posterior credible intervals that did not include zero. To provide more information on the robustness of a significant result, each analysis was performed with credible intervals at 90%, 95%, 99% and 99.9%. We always report the significant result with the more conservative credible interval (similar to reporting *p*-values smaller than a certain value). For all the figures of the result section, significance levels for p-values of less than 0.05 or significant values with a credible interval (CI) of 95%, are flagged with one star (*); for p-values of less than 0.01, or significant values with a CI of 99%, the flag is 2 stars (**); for p-values of less than 0.001, or significant values with a CI of 99.9%, the flag is three stars (***).

For each analysis involving multiple comparisons, if a significant result stayed significant when applying a Bonferroni correction, the *p*-value was reported uncorrected.

VR Game Validation

As one of the central design tenets and innovations of the game was to train HRV under high arousal, we compared HR for the first baseline and the subsequent VR session, across all participants (experimental group and controls). The follow up analysis investigating if the arousing effect of the VR game was maintained in the following sessions in the experimental group were carried out by a mixed-effect model where condition (with or without BF) was used as a random slope and intercept and session number was used to assess the effect of time. We reported subjective arousal with participants' rating of the stressfulness of the VR experience (answered on a 7-point Likert scale). Lastly, to test if the game successfully primed the player to shoot specific targets, a pairwise comparison was done to evaluate if participants from the control and experimental group shot significantly more benign zombies (FA) when the large identifier of these zombies (i.e., their body type) was announced by the radio dispatch information as potentially hostile (dispatch-primed FA), compared to FA happening when the body type of the zombie did not match dispatch information. As for the HR analysis, this "first session" analysis was then complemented with a mixed-effect model testing if a difference between primed and unprimed FA existed in the subsequent VR sessions played by the experimental group. Random slopes and intercepts were fitted per participant for both priming (primed vs. unprimed) and condition (with vs. without BF), as well as for the effect of time (session 1-9).

Effectiveness and appraisal of the training

To investigate the effect of time and training on the HRV of participants, a first mixed effect model was run to evaluate the effect of condition (with and without BF) on HRV. Since an equal number of sessions per condition was needed for the model to converge, and there were 5 sessions without BF and 4 sessions with BF in our design, the first session (without BF) was removed from this first model. The model had random slopes and intercepts for session (to measure effect of time), condition and subgroup (see next analysis) and was then followed by two extra smaller models looking at the effect of time on the VR sessions for each condition separately, by comparing the first and the last session of each condition. The next analysis investigated the effect of introducing BF in different sessions for the two subgroups (A and B) of the experimental group. Three *t*-tests were performed to test if there was a difference in HRV between these subgroups in the first VR session where no participant had yet received BF, in the second VR session where only group B had BF and in the third session where only group A had BF. The HRV averages and standard error of each session are reported in Figure 3. To illustrate the overall effect of BF on HRV, the first panel of Figure 3 merges together the HRV averages of subgroup "a" and "b" of the active group, although group "a" had the first BF session in session 3 only. The second panel of Figure 3 shows the two subgroups separately to avoid confusion.

Four mixed-effects models were used to investigate if the breathing and HR awareness of the participants changed throughout the training. Two complete models used all the data, with session and condition as random slopes and intercepts. Two more partial models only used the data from sessions without biofeedback, to report the evolution of interoceptive awareness in the same way behavioral and physiological data was reported (i.e., comparing the first and the last session). An additional mixed effect model was run with data from session 2 to session 9 to investigate the effect of condition and time on self-efficacy. The breathing and heart awareness averages are reported in the first panel of Figure 4, while the second panel represents the averages of self-efficacy. In both panels, subgroup “a” was merged to subgroup “b” to represent the general trend. Further, the engagement of the experimental group throughout the training was evaluated by testing the score of the engagement questionnaire against the value 4 in a single sample *t*-test. A value of 4 represents appraising the VR game as neutral. The same test was run for the scores of training usefulness and efficacy. The third panel of Figure 4 represents the average rating for the usefulness, efficacy and engagement scores. Each bar moreover represents (as a stacked bar graph) the percentages of ratings for the score, rounded to the closest integer.

Lastly, the relation between the *in-game* HRV of the participants and their behavioral performance was evaluated with a mixed-effect model. The model was evaluating if the d’ sensitivity could be predicted by HRV, with time and condition as additional predictors. The same analysis was also reproduced for FA and unspotted targets (see Figure 5). All predictors were evaluated both with a random intercept and slope per participant. Figure 5 represents the session-by-session averages for those variables. The data for subgroup “a” was rearranged and merged to subgroup “b” as explained in the previous images.

Skill transfer

A chi-squared test was performed to evaluate if the experimental group made less mistakes than the control group in the first trial of the transfer task. The two levels were shooting behavior (shot/withheld shot) and group (experimental vs. control). Two independent sample *t*-tests were done to test if a difference in shooting behavior was related to differences in HRV during the transfer task or in FA rates in the game. The latter comparison was only done on the control group as the experimental group would have had an added confound of in-game repetition (reported in Supplementary Materials A).

To ensure that the transfer task was eliciting physiological arousal and that the elicited arousal did not differ between the two groups, two independent sample t-tests were performed. The first compared the average HR of the controls and the experimental group during the transfer task, and the second the difference in HR between the baseline and the transfer task for all participants merged together.

To test the physiological effect of the BF training, the HRV of the experimental group was compared with the HRV of the controls with independent sample *t*-tests at various moments of the training: the first baseline HRV, the absolute and baseline-corrected HRV during the transfer task. The average HRV levels (with standard error) of the control and experimental groups are represented in figure 6.

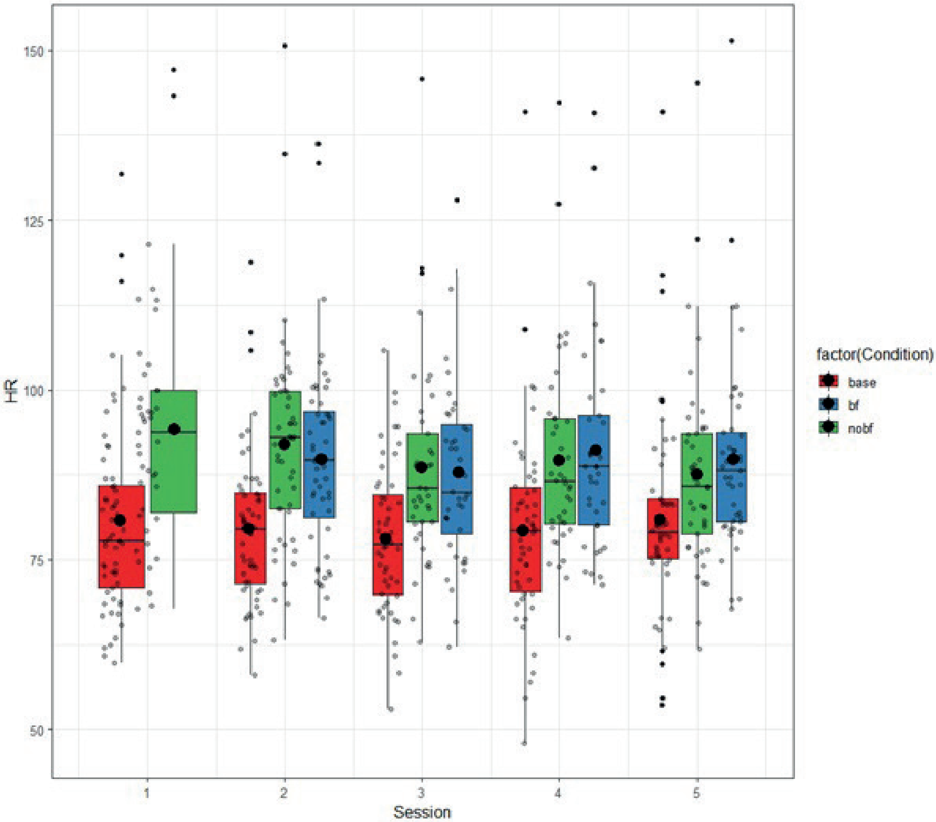
SUPPLEMENTARY MATERIALS

A: Additional results

Objective 1: Is the game environment arousing?

Figure S5.1

Heart Rate across the training sessions in the experimental group



Note. As seen in this graph, the VR sessions 2 to 9 always took place in groups of 2, with one BF and one non-BF session.

In all VR sessions, the HR was reliably higher in-game than during baseline (N = 57, B_{nobf} = 10.57, 99.9% CI [7.43, 13.70]; B_{bf} = 9.91, 99.9% CI [6.67, 12.82]). Between sessions, a reduction in HR was present between the VR session that took place after the second and the third baseline (i.e., sessions 2 and 3 compared to sessions 4 and 5; N = 57, B = -2.46, 95% CI [-4.41, -0.51]).

Objective 2: Effectiveness of the training

In-game behavioral measures. Behavioral measures were only tested in sessions without biofeedback (odd session number), as the biofeedback implementation affected the visibility, which would create a confound between behavior and physiological control.

Shooting behavior assessed with *d'* (sensitivity) improved significantly in the middle portion of the training (N = 57, $B_{S\_5vsS\_3}$ = 0.21, 99.9% CI [0.01, 0.40]; $B_{S\_7vsS\_5}$ = 0.19, 99% CI [0.02, 0.36]) but decreased at the end (N = 57, $B_{S\_9vsS\_7}$ = -0.23, 99.9% CI [-0.43, -0.02]), while false alarm rates decreased mostly at the beginning ($B_{S\_3vsS\_1}$ = -1.27, 99% CI [-2.43, -0.08]; N = 57, $B_{S\_7vsS\_5}$ = -0.81, 95% CI [-1.47, -0.13]). This indicated that participants improved mostly at the beginning in correctly shooting the "hostile" targets, and less often incorrectly shooting the "friendly" non-targets.

Likewise, the rate of unspotted approaches robustely diminished only after the first session ($B_{S\_3vsS\_1}$ = -1.00, 99% CI [-1.83, -0.19]). This indicated that in the first few sessions less zombies reached the player without appearing in their field of vision, suggesting that the spatial awareness (looking around) of players increased.

5

Exit questionnaires. The following section lists all the items of the exit questionnaire.

Table S5.2

Items of the biofeedback evaluation questionnaire, and related statistics.

Question in English (translation)	Question in Dutch (shown to participants)	Mean	Std. Deviation
1. It was easy to check my physical reactions (breathing, heart rate) during exercise.	Het was makkelijk om mijn lichamelijke reacties (ademhaling, hartslag) the controleren tijdens de training.	3.94	1.564
2. The biofeedback exercise at the beginning of each day was helpful.	De biofeedback oefening aan het begin van elke dag was nuttig.] Biofeedback Evaluation	5.44	1.474
3. It was easy to check my physical reactions (breathing, heart rate) at times with more zombies.	Het was makkelijk om mijn lichamelijke reacties (ademhaling, hartslag) te controleren op momenten met meer zombies.	3.08	1.440
4. It was easy to focus on the biofeedback during the training.	Het was makkelijk om mij op de biofeedback te concentreeren tijdens de training.	4.37	1.442
5. It was always clear how well I was doing in terms of biofeedback during training.	Het was altijd duidelijk hoe goed ik het deed m.b.t. de biofeedback tijdens de training.	5.31	1.422
6. The virtual environment clearly responded to changes in my breathing.	De virtuele omgeving reageerde duidelijk op veranderingen in mijn ademhaling.	5.31	1.336

Note. Total N = 52

Table S5.3

Items of the game evaluation questionnaire, and related statistics.

Question in English (translation)	Question in Dutch (shown to participants)	Mean	Std. Deviation
1. The "control of the self" score at the end of each session was informative.	De "control of the self" score aan het eind van elke sessie was informatief.	5.65	1.370
2. The "control of the suspect" score at the end of each session was informative.	De "control of the suspect" score aan het eind van elke sessie was informatief.	5.42	1.036
3. The "control of the situation" score at the end of each session was informative.	De "control of the situation" score aan het eind van elke sessie was informatief.	5.50	1.000
4. The game was stressful	De game was stressvol.	3.46	1.650

Note. Total *N* = 52

Table S5.4

Items of the training usefulness questionnaire, and related statistics.

Question in English (translation)	Question in Dutch (shown to participants)	Mean	Std. Deviation
1. I think Biochill adds something valuable to existing police training.	Ik denk dat Biochill iets waardevols toevoegt aan de bestaande opleiding van politieagenten.	5.90	1.107
2. I would recommend using Biochill for training new police students.	Ik zou het gebruik van Biochill aanraden voor de opleiding van nieuwe politie studenten.	5.90	1.053
3. I would use Biochill in my own training program.	Ik zou Biochill gebruiken in mijn eigen trainingsprogramma.	5.65	1.454
4. I think the Biochill training can help me teach Mental Strength to my students.	Ik denk dat de Biochill-training mij kan helpen om Mentale Kracht aan mijn studenten te leren.	5.90	1.125
5. The training was long enough for me to learn how to control my breathing.	De training was lang genoeg om te leren hoe ik mijn ademhaling onder controle kon houden.	4.63	1.692
6. I think the training is too long to be included in the curriculum of police students.	Ik denk dat de training te lang is om in het curriculum van politiestudenten te worden opgenomen.]	2.37	1.572
7. I think police students would take this training seriously.	Ik denk dat politiestudenten deze training serieus zouden nemen.	5.48	1.393
8. I took this training seriously.	Ik nam deze training serieus.	6.33	.857
9. I think more game-based training like Biochill should be developed.	Ik denk dat er meer op games gebaseerde trainingen zoals Biochill moeten worden ontwikkeld.	6.06	1.110

Note. Total *N* = 52

Table S5.5

Items of the training efficacy questionnaire, and related statistics.

Question in English (translation)	Question in Dutch (shown to participants)	Mean	Std. Deviation
1.The Biochill training program increased my focus on how you breathe under stress.	Het Biochill-trainingsprogramma verhoogde mijn aandacht voor hoe je ademt onder stress.	5.87	1.284
2. I think I am more able to control my breathing in an action context than before.	Ik denk dat ik meer in staat ben om mijn ademhaling te beheersen in een actiecontext dan voorheen.	4.96	1.508

Note. Total *N* = 52

Table S5.6

Items of the control questions questionnaire, and related statistics.

Question in English (translation)	Question in Dutch (shown to participants)	Mean	Std. Deviation
1. Did you have experience with virtual reality before this training?	Ik denk dat ik meer in staat ben om mijn ademhaling te beheersen in een actiecontext dan voorheen.	2.87	2.115
2. How much experience do you have with similar games?	Hoeveel ervaring heb je met soortgelijke spellen?	2.77	2.006

Note. Total *N* = 52

Table S5.7

Items of the engagement questionnaire, and related statistics.

Question in English (translation)	Question in Dutch (shown to participants)	Mean	Std. Deviation
1. I really enjoyed doing this training.	Ik vond deze training erg leuk om te doen.	6.48	0.7
2. This training did not hold my attention at all.	Deze training hield mijn aandacht totaal niet vast.	2.00	1.704
3. I would describe this training as very interesting.	Ik zou deze training als erg interessant omschrijven.	6.35	0.883
4. While doing this training, I often thought about how much I enjoyed it.	Terwijl ik met deze training bezig was, dacht ik er vaak aan hoe leuk ik het vond.	4.96	1.381

Note. Item 2 was reverse-coded to calculate the engagement score; Total *N* = 52

Objective 3: Skill transfer

Behavior and physiology. As explained in the manuscript, we did not find a group difference in behavior (ratio of shots vs. withheld shots; See Table S.5.8) on the transfer task. As pre-registered (H.3.1.2.2), in this case we planned a follow-up analysis, consisting of a comparison of HRV between participants (from both groups) who shot in the first trial of the transfer task vs. those who did not. However, this comparison showed no difference between shooters and non-shooters in HRV (*M*shooter = 48.51, *SD* = 18.50; *M*non-shooter = 54.02, *SD* = 26.62; *t*(76) = 75.821, *p* = 0.284). This finding supports the conclusion from the main analysis that although the transfer task did show differences in HRV between groups, there was no relation between HRV and behavior, or most likely, the task was not suited to detect such relation.

Table S5.8

Shooting behavior in the first (No-Go) trial of the transfer task.

	Withheld shot	Shot	Total
Control group	28	19	47
Experimental group	30	23	53
Total	58	42	100

We additionally investigated, in the control group, whether there was a relation between the behavior in the game and the transfer task, specifically the number of false alarms primed by radio dispatch both tasks (i.e. we tested if shooting the innocent target in the first trial of the transfer task was related to a higher number of false alarms for the primed non-targets in-game). No relation between these measures was found: shooting the innocent target in the transfer task had was not associated to a higher scores of false alarms in the first and only VR session played (*M*shooter = 1.32, *SD* = 1.376; *M*non-shooter = 1.54, *SD* = 1.835; *t*(45) = -0.444, *p* = 0.659).

B: Changes in the game design

As explained in the manuscript, the VR environment was adapted after a first feasibility study (Michela et al., 2022). Below all changes are listed in detail.

Table S5.9

List of changes between the version used in the current study and the previous

Previous game version	Current game
1. VR session duration: 15 to 17 minutes	VR session duration: 5 to 7 minutes
2. 14 Waves of zombies with up to 340 targets	5-6 waves of zombies with up to 110 targets
3. Biofeedback on compliance to breathing pace	Biofeedback on magnitude of breathing induced fluctuations of the IBI trace
4. Game events: 3 distracting game events in random order (glass breaking, car alarm, fire alarm)	Game events: one of 3 game events happening per session (glass breaking, car alarm, fire alarm)
5. 5 different session types with different order of game events and different target identifiers combinations (eye color and body type)	3 different session types with one of the 3 game events and different target identifiers combinations (eye color and body type)
6. Physiological recordings: Abdominal breathing amplitude and IBIs	Physiological recordings: IBIs
7. No offline biofeedback tutorial. The biofeedback skill had to be learned in game entirely	A short (up to 2minutes) offline biofeedback session precedes the VR sessions done on a same day

Note. In the current game version, no new *in-game* features were added.

C: Hypotheses list

As mentioned in the article, some pre-registered analyses could not be performed as the planned analysis, repeated measures ANOVA, would result in too many missing data due to listwise deletion. Therefore, these analyses were replaced with Bayesian mixed models. Additionally, exploratory analyses were added to the main article. The present list catalogs all hypotheses from the article (and pre-registration) that were modified or only reported in the supplementary materials, to clearly keep track of changes.

Table S5.10

List of pre-registered, modified and exploratory hypotheses

Hypothesis	Status
Objective 1	
H.1.1.2: In the following sessions, in-game HR is higher than baseline HR (main effect of condition)	Pre-registered, modified to fit mixed models, main result included in the article, rest in supplement
H.1.1.3: The experience will be rated as moderately stressful (higher than 2 on a 7 point scale)	Pre-registered, mean reported in the main article but the statistic was removed to avoid confusion (a significant positive result would not mean that the experience is stressful).
H.1.2.2: In the following sessions, the priming effect described in H.1.2.1 is present.	Exploratory, included in the article
Objective 2	
H.2.1.1.1: There is an increase in absolute HRV over sessions reflected in a main effect of time on absolute HRV across sessions.	Pre-registered, modified to fit mixed models, ran separately on sessions with and without biofeedback and included in the article
H.2.1.1.2: There is a main effect of condition on absolute HRV. i.e., the HRV is higher in the BF compared to the no-BF condition.	Pre-registered, modified to fit mixed models, included in the article
H.2.1.1.3: In addition to the above, our previous data (Michela et al., 2022) give reasons to expect a potential interaction effect between time and condition, as the increases in the no-BF session are expected to be steeper	Pre-registered, modified to fit mixed models, ran separately on sessions with and without biofeedback, hence interaction not calculable
H.2.2.1: Self-reported in-game breathing awareness increases both across session (main effect of time) and in BF condition when compared to the no-BF condition (main effect of condition).	Pre-registered, modified to fit mixed models, included in supplementary materials, but result mentioned in result section
H.2.2.2: Self-reported in-game HR awareness increases both across session (main effect of time) and in BF condition when compared to the no-BF condition (main effect of condition).	Pre-registered, modified to fit mixed models, included in supplementary materials, but result mentioned in result section
H.2.2.3: Self-reported in-game breathing self-efficacy awareness increases both across session (main effect of time) and in BF condition when compared to the no-BF condition (main effect of condition).	Pre-registered, modified to fit mixed models, included in supplementary materials, but result mentioned in result section
H.2.4.1: In the experimental group, HRV score in no-BF conditions is positively correlated to behavioral performance (d' sensitivity) in the same session.	Pre-registered, modified to fit mixed models, included in the article
H.2.4.2: In the experimental group, HRV score in no.BF conditions is positively correlated to the number of unspotted targets in the same session.	Exploratory, included in the article

Table S5.10 (Continued)

Hypothesis	Status
Objective 3	
H.3.1.1: While they perform the transfer task, the baseline-corrected HRV of participants from the experimental group is higher compared to the control group.	Pre-registered, included in the supplementary materials as the ambiguous formulation (baseline-corrected) was misleading.
H.3.1.1.2: While they perform the transfer task, the absolute HRV of participants from the experimental group is higher compared to the control group.	Exploratory, added to complement the above analysis, included in the article
H.3.1.1.3: When compared to baseline values, the HR of participants in the transfer task is significantly higher, yet there is no difference in HR levels between the controls and the experimental groups.	Exploratory, added to complement the above analysis, included in the article
H.3.1.2.2: If no group difference is found in HRV at point 3.1.1, then we will test this alternative hypothesis with an independent sample t-test. The previous hypothesis is then changed to: The test shows that participants who shot the target (FA) in the transfer task have significantly less HRV than the ones who did not shoot the first target.	Pre-registered, included in the supplementary materials
H.3.2.1: In the control group, the participants that do not make a (false alarm) shooting mistake in the first trial of the transfer task make significantly less “dispatch driven” false alarms in-game than the participants who do make a shooting mistake in the first trial of the transfer task.	Pre-registered, included in the supplementary materials

Note. Hypotheses that were pre-registered and reported without modifications are not included in the table; BF = Biofeedback; HR = Heart-Rate, HRV = Heart-Rate-Variability.

D: Data Analysis, details of mixed models

All categorical predictors were coded using a custom zero-sum contrast the following contrasts were used. The main results aimed at demonstrating an improvement throughout the training were done by contrasting the first session and the last. Additionally, contrasts for session were performed to estimate the effect of repetition. Those contrasts could have 4 or 5 levels depending on the test performed, as there were only 4 BF sessions (e.g., when looking at the evolution of BF sessions, only sessions 2,4,6 and 8 were considered). The following contrasts were used to contrast sessions:

Session (4 levels): Session 2 = -1/0/0, Session 3 = 1/-1/0, Session 4 = 0/1/-1, Session 5 = 0/0/1;

Session (5 levels): Session 1 = -1/0/0/0, Session 2 = 1/-1/0/0, Session 3 = 0/-1/0/0, Session 4 = 0/0/-1/1, Session 5 = 0/0/0/1;

Condition (comparison to baseline): Baseline =-1/-1, BF= 1/0, NoBF = 0/1;

Condition (comparison between BF and noBF): NoBF = 1, BF = -1.

For every model, draws were on 4 chains, each with 15000 iterations (warmup = 6000; thin = 1; total post-warmup draws = 36000).

Environment Validation

Heart rate. To ensure that the HR kept increasing from baseline as session number increases, and test if the BF condition was as arousing as the No-BF condition, the following model was run:

HR ~ Session * Condition + (1 + Condition | Subject)

Behavior. (False alarms for primed and unprimed non-targets): To evaluate if a non-target was at increased risk of being shot (compared to other non-targets) when its physical attributes partially matched radio dispatch information, the following model was run:

FA ~ session * condition * prime + (1 + condition * prime | Subject)

Effectiveness of the training

Heart Rate Variability. When evaluating the evolution of the HRV throughout the training, the original full model, incorporating both conditions (BF and NoBF) and all sessions (5 NoBF sessions and 4 BF sessions) did not converge. Hence, 3 separate models were tested: 2 models testing separately per condition (BF and NoBF) the effect of time (session number); one model for the NoBF sessions only, but excluding session 1, to get an estimate of the effect of time (session number) that is comparable with the BF sessions; one additional model was run also with the exclusion of session 1 to get an equal number of NoBF sessions as BF sessions and thus allow to compare both conditions.

For both the BF and no-BF condition (with and without session 1), the model evaluating the effect of time was:

HRV ~ session + (1 + session || subject)

Note: the model did not account for inter-subject correlation, as convergence issues prevented the calculation to converge otherwise.

For the model taking both conditions together (BF and No-BF), hence dropping the session 1 (no-BF), the model is:

HRV ~ session * condition * GROUP + (1 + condition * session || subject)

Self-report questionnaires. For the following models on self-reported physiological awareness and self-efficacy, both conditions (BF and no-BF) could be integrated together. For the first model of each condition, session 1 was not removed as the convergence issue experienced in previous models was not present here. The Self-efficacy had only 4 levels as Self-Efficacy was not measured after session 1. The condition contrast was comparing the BF and the no-BF conditions. For the two awareness conditions, a model without BF sessions was added, so that the variable's evolution could be reported in the article in the same way behavioral and physiological variables were.

Breathing awareness:

(Breathing Interoceptive Awareness) ~ Session + Condition + (1 + Condition + Session || Subject)

(Breathing Interoceptive Awareness, no-BF) ~ Session + (1 + Session || Subject)

Heart-rate awareness:

(Heart-rate Interoceptive Awareness) ~ Session + Condition + (1 + Condition + Session || Subject)

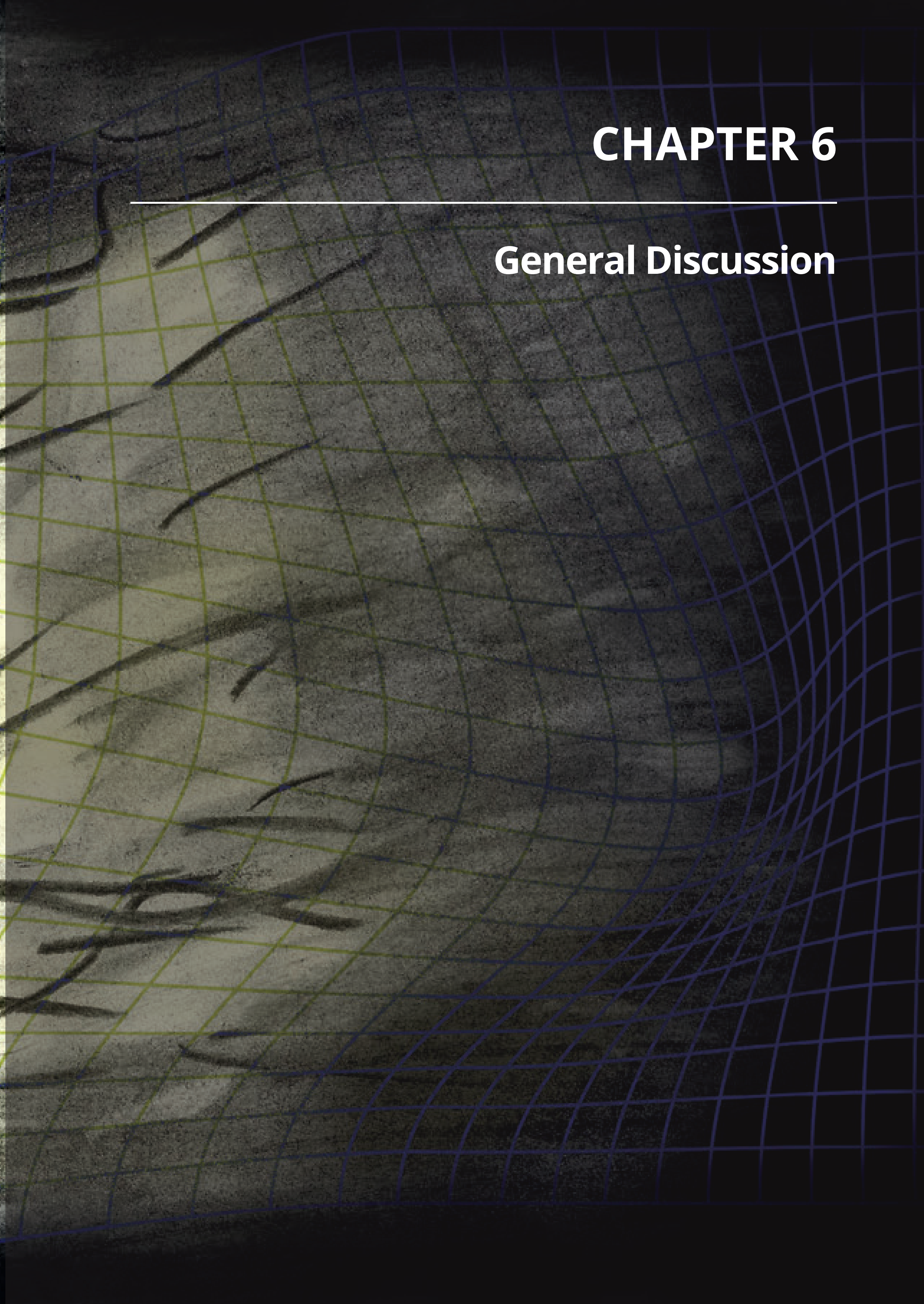
(Heart-rate Interoceptive Awareness, no-BF) ~ Session + (1 + Session || Subject)

Self-efficacy:

(Self-Efficacy) ~ Session + Condition + (1 + Condition + Session || Subject)

CHAPTER 6

General Discussion



GENERAL DISCUSSION

The current thesis describes how a VR biofeedback (BF) game intervention was designed and tested. The aim of the intervention was to train police officers to control their level of psychophysiological stress in action, indexed by Heart-Rate-Variability (HRV). The research demonstrated the trainability of HRV in an active context and the transfer of the acquired HRV upregulation skill to contexts without BF, also outside of the game environment. This game-based training was perceived as lastingly engaging by police trainers, who indicated in a large majority that they would use such a training in their own teaching practices. I will now briefly summarize the key findings from each chapter and will continue with a discussion of overarching issues pertaining to the work described throughout this thesis. I will then proceed to evaluate the success of our game in terms of design and effectiveness, review the strength and limitations of the current research, formulate suggestions for future research and conclude with a closing statement summarizing the impact that I hope this thesis had, and will continue to have, on both research and society.

Summary of results by chapter

In **chapter 2**, we discussed essential considerations on **how to design a suitable VR environment** to train police officers to manage their stress response while immersed in an action context. We highlighted that there are limited benefits to providing a realistic training environment, both in terms of representativeness and potential for behavioral transfer. We proceed to report lessons learned from our unsuccessful initial attempts at creating such an environment. Namely, the chapter carefully documents the risks of trying to reproduce all aspects of policing reality by using photorealistic simulations. First, such simulations are more at risk of feeling uncanny to the user (Mori, 1970), thus distracting them from the learning objective. Additionally, accurate re-creation of real policing situations featuring violence and human distress may produce ethical concerns due to this sensitive content and the reactions it might produce (Madary & Metzinger, 2016), as well as the risk of unwanted transfer of context-specific behaviors (e.g., overtraining a specific behavior in a training environment can lead to the same behavior being triggered automatically in real life). Lastly, just like real-life scenarios, the data of realistic training simulations are difficult to analyze and interpret: The more real and therefore complex a simulation is, the harder it is to even define what correct and incorrect behavior entails (Brehmer, 1992). Chapter 2 concludes with the suggestion that building a VR training around a game-based mechanic circumvents most of the problems of realistic VR simulations. Indeed, a game setup and narrative can be used to "suspend disbelief" (Gorini et al., 2011), thus circumventing the uncanniness problem and enhancing the engagement elicited in the players (Allcoat & von Mühlenen, 2018). Additionally, a VR game environment can reliably elicit threat reactions (Allcoat et al., 2015; Lin, 2017), while avoiding the ethical concerns of realistic setups. Lastly, defining correct and incorrect behavior in a game environment is greatly facilitated by the relative simplicity of the environment.

To help police officers to control their psychophysiological arousal in an action context, it is necessary to ensure that their psychophysiological state is measured in real-time, and in a reliable manner. **Chapter 3** detailed the necessary considerations to **choose, design and implement a biofeedback parameter** suitable for a stressful virtual action context by considering three important challenges. First is the challenge of choosing a suitable biofeedback parameter. *In-action* BF has the participant's attention divided between the BF training and the demands of the game. Hence, the BF training parameter needed to be easily controllable, atop of probing the correct psychophysiological process. We chose deep abdominal breathing as a BF training due to its ease of control (Lehrer et al., 2020; Russo et al., 2017). Additionally, breathing-induced HRV is an index of relaxation directly related to the psychophysiological process of stress management (Balzarotti et al., 2017; Hayano & Yuda, 2019; Yu et al., 2018). The second challenge was related to the implementation of the biofeedback processing. Since the participants are immersed in an active VR game, the online data-analysis needed to have good movement artifact resistance, as well as a robust way of comparing the estimated parameter to the training target. Our solution of breathing-based frequency analysis fulfilled those requirements, and allowed for a simple estimation of the biofeedback parameter. The last challenge was to create a biofeedback implementation that felt salient and meaningful to the user. The feedback had additionally to feel distinct from the action context provided by the game, so that bad behavioral performance could not be attributed to BF, as it would be the case if BF affected for instance the precision of the gun's aim. In our case we chose a visual impairment resembling a "tunneled vision" since this perceptual phenomenon is already known by police officers and linked to stress responses. Thus, using tunneled vision as a BF implementation was salient, intuitive and distinct enough from behavioral performance.

Chapter 4 reported on the **trainability** of HRV through breathing based BF in an action context. It described our first experiment testing the effects of our training in a small sample of police trainers (*n*=9). The results showed positive effects in terms of successful in-action HRV control in 8 out of the 9 participants and sustained engagement throughout the training. Hence, this study reported that training results obtained with VR BF in non-stressful passive environments (Blum et al., 2019; Rockstroh et al., 2019) can be achieved also in active environments, without impeding behavioral improvement. Moreover, this first experiment demonstrated that the training was perceived as a strong positive challenge and highly engaging which supports its use as a training environment (Jamieson et al., 2010). Lastly, the study hinted at a beneficial effect of the addition and withdrawal of BF in the training sessions.

Chapter 5 provided further support for the **training effectiveness** by showing, in a large sample, that the HRV control was causally and reliably related to the BF: Removing the BF component from a training session repeatedly led to a reduced HRV upregulation performance, and a delayed introduction of BF in the training schedule led to a delayed increase in HRV upregulation performance. As explained in detail in the strength and limitations section of this General Discussion (see below), such a causal inference is often missing in bio- and neuro-feedback literature (Thibault & Raz, 2017). Importantly, chapter 5 also provided evidence for **transfer of the physiological regulation skill:** The training also led to an increased HRV outside of the game, in a subsequent police-relevant decision-making task. Interestingly, the *in-game* improvement in HRV control could not be linked to the *in-game* **improvement in behavioral performance** (e.g., higher accuracy in shooting the correct targets). Importantly however, this study is the first to measure a transfer of BF control skill from a VR-BF training to a new non-VR condition, which was identified as an important gap in the literature by Lüddecke and Felnhofer (2022). Lastly, chapter 5 again reports on the very positive reception of the training by the participants, a fundamental prerequisite for any successful training intervention using digital innovation (Scholten & Granic, 2019).

Taken together, the results summarized above suggest that HRV BF in an engaging VR game holds extensive potential as a training tool, potentially even for more populations than police officers only. In the following section of this discussion, we will critically review each component of the training, before moving to a more general debate on the strengths and limitations of this training method and how it was used in the present thesis.

VR game validation

As stated in the introduction, there were four design goals for the VR game: (1) create an engaging experience, (2) elicit psychophysiological stress, (3) provide an active decision-making context that feels relevant for police officers and (4) have an imbedded mechanic to provide BF on the participant HRV. These four design goals can be separated in two dimensions. The first two goals (engagement and stress elicitation) relate to the desired reactions to be elicited by the game, while the last two goals (decision-making context and BF) concern the training components that had to be included in the game. I will hereby review those two design goal categories separately.

Reaction elicitation

Engagement. The game received a very positive appraisal by the participants, as it was rated as engaging for up to 10 training sessions (see chapter 4), and more than 80% of the police trainers participating in our last experiment (chapter 5) indicated that they would include it in their own teaching program. We attribute this positive reception to the use of a game as a training tool and the feedback it provides. Police proficiency is an important dimension of officers' feeling of self-worth (Walsh, 1983) and therefore the feedback provided by the game is likely to have been a powerful motivator to play the game. The feedback did not simply categorize performance into good/bad, but showed multiple dimensions of the player's performance in the form of three scores at the end of the game: a self-control score for BF performance, a behavioral control score for the shooting behavior and a situational control score for the spatial awareness. It is additionally important to acknowledge the role of experimenters in ensuring the efficacy of the training tool as they facilitated the understanding of the game's feedback and provided additional instructions to the player when needed, thus playing a positive "coaching" role of autonomy support for the participants (Trigueros et al., 2019). It should however be mentioned that the experimenters' supervision and guidance were limited to the moments where the participants were not actively engaging in the VR BF training, as evidence has shown that both passive and active experimenter presence can hinder BF performance, due to the distraction their presence creates (Borgeat et al., 1980; Bregman & McAllister, 1983). Hence, the game's success as a training tool may not only be due solely to its inherent properties of engagement elicitation but also to the supervision provided by experimenters, although as highlighted in chapter 5 by the effect that addition and removal of BF had on HRV control, the causal role of BF still remains the driving component of the witnessed HRV upregulation.

The high engagement elicited by the game hints at one of the major design successes of this game. Often, reduced engagement is witnessed as a consequence of poorly designed intervention tools. Such resistance to poor design has been documented at the hardware level (Ledger & McCaffrey, 2014) and has also been shown in game-software used for interventions (Scholten & Granic, 2019). In our game, the engagement created by the game was coupled with instantaneous feedback on the participant's physiology. Additionally, automatic analyses were implemented at the end of each game session, providing results of behavioral and physiological game-play variables (i.e., suspect control, situational control and self-control scores). With this feedback, experimenters were given tools to provide clear and constructive feedback to participants, which helped them to set their own learning goals, thus enhancing engagement. We also witnessed some examples of competition between participants, who routinely compared their scores after a training session. Finally, two qualitative observations: it occurred to us that during the testing we were occasionally actively approached by police trainers requesting enrollment in the study. Additionally, several police trainers have inquired when they could immediately start using the game in their own classes.

Psychophysiological stress. As illustrated in both chapters 4 and 5 of the present thesis, the game environment elicited arousal, as evidenced by consistent and lasting increases in HR from baseline to during the VR game. It can therefore be assumed that the HRV upregulation skill was trained in a stress-relevant context in terms of the psychophysiological state of the participants. On the self-report measures however, police trainers did not show to have experienced the training as stressful, but rather as a positive challenge (see chapter 4). The fact that the subjects did not report the game as stressful seems to suggest no stress was elicited and it should be taken into account that the HR increase could have been partially due to motion in the game. Nevertheless, widely used definitions of stress center around the concept of a challenge, and are therefore in line with the fact that the we did evoke stress even when subjects did not report it as such. As alluded to before, police may underreport negative feelings (Habersaat et al., 2021), and might consider that a situation is stressful only when it is perceived as exceeding their capacities, a common view in the stress literature (Koolhaas et al., 2011). Even when the stress as challenge definition is debated and motion would have played a role in the increase in HR, exercise is also known to reduce HRV (Michael et al., 2017; Sandercock & Brodie, 2006). Therefore, our training allows subjects to train voluntary HRV upregulation under challenging arousal conditions. This may give our implementation an advantage over traditional biofeedback setups. Additionally, a challenging environment is better than a stressful environment to learn skills, according to Jamieson et al. (2010). The authors found that students who re-appraised the task demands of a difficult examination as a challenge rather than a threat scored higher in that examination than controls.

Similarly, the appraisal of our game as a positive challenge rather than a threat suggests, following Jamieson et al. (2010), that the task demands were believed by the participant to be inferior to their available resource, which the authors suggest is a necessity to improve learning.

Training components

The last two design goals of the VR game, namely providing a relevant decision-making context and incorporating BF, were built according to the principles listed by Scholten and Granic (2019). The authors recommend including relevant experimental designs in the game intervention by using a model guiding the incorporation of evidence-based techniques (see Table 6.1), as explained in the introduction. I will hereby review the four psychological mechanisms of our VR game model (psychophysiological regulation, response inhibition, priming and spatial attention) to evaluate the success of their integration in the game.

Table 6.1

Guiding model for the incorporation of evidence-based mechanisms and techniques in the VR game environment

Psychological mechanism	**Evidence-based technique**	**Game mechanic**	**In-game outcome**	**Envisioned long term goal**
Psycho-physiological regulation	Bio-feedback	Width of the field of view linked to HRV score	Breathing/ HRV biofeedback score increase	Improved stress management in real contexts
Response inhibition	Go/No Go	Hostile (Go) vs. Benign (No Go) zombies according to dispatch information	Increased response accuracy	Improved and less biased decision making in presence of stressors
Bias resistance	Priming	Benign zombies partial matching (body type, but not eye color) with dispatch target information	Fewer primed No Go false alarms	
Spatial attention	Target detection	Targets coming from multiple sides	Fewer zombies reaching the player unspotted	Improved situational awareness

Note. The items of the last column (Envisioned long-term goal) are part of the model as they served as a rationale for the incorporation of the relative psychological mechanism in the game. Testing those effects in real-life policing contexts was however outside the scope of the present thesis.

Psychophysiological regulation. The increased capacity for psychophysiological regulation as a result of the HRV-BF intervention was not just a good result to inform theoretical models, but also an important argument for police stakeholders to further invest in the training. Indeed, the reported performance increases happened in a population of police trainers, who already received a mandatory biofeedback training (seated, without stress induction) in the years preceding our experiment, rendering the result of our training all the more valuable. Improvements were still witnessed for some participants in the last sessions of the training, although on average improvement took place mainly in the first few sessions. These results suggest that, if our training were incorporated in police officers' training, the schedule should ideally include 3 days of training, with multiple VR BF sessions per day. This suggestion, which goes against the tendency of police academies to shorten trainings to a minimum, is corroborated by findings by di Nota et al. (2021). They report in a population of active-duty police officers that reducing the number of training sessions of effective BF trainings to a single day of training led to a loss of efficiency. Additionally, the improvements still witnessed in the last VR training sessions for some participants leads me to suggest that implementations in the officers' curricula should consider (1) tailoring the number of sessions to the specific needs of each participant and (2) modifying the game to make it more independent from experimenter output, to alleviate the required manpower to operate the training.

The implementation of the BF as visual impairment that mimics the phenomenon of tunneled vision was suggested by the stakeholders of the project. To understand how this visual implementation of biofeedback was appraised by the officers, self-efficacy and physiological awareness were measured during the training, as described in chapter 5. The presence of BF decreased the reported feeling of self-efficacy in participants. This reduced self-efficacy hints at the punitive aspect of the BF presentation: When biofeedback is present, it calls participants out on forgetting to upregulate HRV. Moreover, the feeling of respiratory and cardiac awareness increased over sessions. Thus, it could be inferred that being repeatedly called out on forgetting to upregulate HRV trains awareness on physiological control performance. While our setup does not allow to claim any causal relation between appraisal and performance, these findings are broadly in line with the model theorized by Weerdmeester et al. (2020), suggesting that increases in BF control require changes in appraisal of self-efficacy and awareness.

Response inhibition. As explained in the introduction, we needed to create a decision-making context assessing a psychological mechanism that is frequently impaired due to stress responses and relevant to police behavior in the training game. This mechanism allowed us to monitor the improvement of police officer's performance, and the potentially beneficial effects of BF on behavior. Therefore, we introduced a Go/No-go task based on a zombie-shooter dynamic. The player had to decide for each incoming zombie if it was a target to be shot at (Go, to prevent being attacked) or a benign zombie to spare (No Go). False alarms were punished with a loud unpleasant noise. The number of benign zombies shot by the player (False Alarms; FA) was quite low. As a result, in contrast to conventional Go/Nogo tasks, FAs in our game cannot be considered a reliable indicator of response inhibition due to potential floor effects. However, the variable can still be used to monitor compliance to the task as a high number of FA could be indicative of not understanding/caring for the given task. Nevertheless, to use it as a behavioral performance measure and study its relation to physiology, I recommend to find ways to make the game more difficult, especially when dealing with police officers as participants. The current version of the game already used optimal go/no-go ratios (3:1 hostile to benign; Young et al., 2018) to maximize false alarms, but other suggested methods to increase false alarms, such as multiple types of go stimuli or shorter time intervals for decision (e.g., by increasing the speed of targets) were not implemented. Since changing elements in the game can have unforeseen effects in terms of training outcomes, I however advocate caution in implementing drastic game-play changes. For instance, following Jamieson et al., (2010, 2013), it could be imagined that a more difficult go/no-go decision task could increase the stressfulness of the game and elicit a feeling of helplessness in the participant, which would reduce the feeling that their resources meet the task's demand and hence negatively impact learning.

Beside the methodological limitation of insufficient FA, a theoretical consideration has to be discussed as well: the benefit of our HRV training may be dependent on the behavioral domain assessed. As reported in chapter 5, the inclusion of the go/no go mechanic in the game stemmed from a need to incorporate behaviors that are thought to be affected by the presence of stressors in real-life policing situations. Additionally, risky decision-making and HRV have been reported to correlate (Forte et al., 2022) and previous work has suggested increased parasympathetic dominance is linked to more optimal decisions (Hashemi et al., 2019; Klaassen et al., 2021). However, the relation with HRV is mostly found in other types of risky decision-making tasks, as for example the Iowa Gambling Task and the Balloon Analog Risk Task (Forte et al., 2021; Ramírez et al., 2015). Both tasks include an element of risk taking and uncertainty that were not explicitly included in our game model, elements that I would recommend future research to include in their BF training contexts.

Bias resistance. The same methodological and theoretical considerations described above also apply to the priming task: It was included for its police-relevance but might lack a strong enough relation with HRV. Indeed, it could be argued that priming is a relevant dimension to train since stress reactions increase potentiated biases (Yu, 2016). Additionally, HRV has been directly linked to less biased behavior by Di Palma et al. (2019). However, their task did not induce the time pressure of shooting an incoming zombie, but rather focused on evaluating the pain experienced by actors with varying skin color. It can therefore be suggested that the type of priming used in our task was too automatic and fast paced to reproduce those effects. This hypothesis is further strengthened by the fact that HRV level has been linked to altered processing only for positive stimuli when stressors are present (Macatee et al., 2017), and would therefore not apply to incoming zombies. I therefore recommend further research to focus on higher order priming tasks involving emotional and verbal dimensions when investigating the benefits of HRV on bias reduction.

Interestingly, as reported in the Supplementary Materials of chapter 5, FA occurred more frequently when the shot zombie had a body type matching the dispatch prime announced by radio before each zombie wave. Hence, the bias induction manipulation was effective, and did indeed increase FA rates, thus potentially mimicking effects of excessive use of force induced by radio dispatch priming reported in police officers in more realistic setups (Taylor, 2020). Interestingly, while the overall rate of FA (see the response inhibition section) dropped significantly throughout the training, this improvement could not be seen in the FA caused by priming. In other words, while the number of benign zombies being shot by mistake *as a whole* reduced throughout the training, the number of benign zombies incorrectly shot because they matched the body type announced in the radio dispatch stayed constant. Thus, as seen in chapter 5, bias resistance as we implemented it is not affected by the HRV training in-game. This negative result could explain why no behavioral effect of the HRV training was found in the transfer condition, that also tested bias resistance through dispatch induced priming.

Spatial Attention. In real life policing situations, attention often has to be allocated to multiple directions. At the same time, the stress response alters attention by enhancing focus to coarse visual features (Lojowska et al., 2015) and creating attentional biases (Aue & Okon-Singer, 2015), thus impacting the capacity an officer has to monitor their surroundings. We took advantage of the 3D characteristics of VR and the zombie shooter dynamic to measure the level of attention that players had toward monitoring their surroundings. This metric allowed us to provide participants with a police-relevant measure of spatial attention. Since zombie targets could reach the player from any direction, we measured the time that each zombie spent in the field of view of the player before reaching them. We then counted the number of zombies (both benign and hostile) reaching the player without ever being spotted (never entering their field of view) as an inverse marker of spatial attention. Even though this number of unspotted zombies decreased slightly throughout the training, it already started fairly low (less that one unspotted zombie per play session on average; see chapter 5). The monitoring task might therefore have been too easy for police trainers. For that reason, just as the behavioral indices described above, the number of unspotted zombies in a game session was not related to the *in-game* HRV score. Therefore, I again recommend to find ways to make this aspect of the game more difficult (e.g., by speeding up the incoming zombies) in order to study the relationship between spatial attention and physiological control.

Strengths, limitations and suggestions for future research

Strengths and limitations of the studies design

The first strength of the research presented in this thesis lies in the studies' design. To assess the efficacy of a training, a Randomized Controlled Trial (RCT) as used in Chapter 5, is considered the gold standard, as it allows to experimentally link a change in outcomes to the performed intervention. Critiques have however been raised on the fact that RCTs are not capable, per se, to give a mechanistic account of how or why an intervention works (Deaton & Cartwright, 2018). For instance, the observed higher HRV at transfer for the experimental group compared to controls in chapter 5, would in itself not enable us to determine what exactly drives this effect (Deaton & Cartwright, 2018), as an increase of HRV at the end of a training could be due to confounds, such as habituation to the stressor and experimental context (Feda & Roemmich, 2016). Our training schedule in chapter 5 was therefore designed to allow us to go further in the identification of causal mechanisms of change. Specifically, the delayed introduction of BF in half of the experimental group, inspired by multiple baseline designs used in small-N designs (Levin et al., 2021), worked as an RCT nested within the main RCT. It allowed to causally attribute the increase in HRV scores to the BF presentation, with minimal perturbations in the training schedule. Moreover, the addition and withdrawal of BF through the different VR sessions (in both chapters 4 and 5) was accompanied by increases and reductions in HRV upregulation performance, thus further strengthening the claim that BF causally induces the changes in HRV.

The addition and withdrawal design was crucially introduced in chapter 4, to investigate in a very small sample (*N* =9) the influence of BF on HRV upregulation performances. While it allowed us to gather enough information on BF effectiveness to design the larger study in chapter 5, this design also presented another advantage: It potentially allowed to increase the effectiveness of the training by making control of HRV a more voluntary action. Specifically, our results on BF appraisal showed that participants' interoceptive awareness increased steadily across training, thus indicating that participants were more and more aware of their breathing and HR. However, self-efficacy varied with the presence/absence of BF, as it was lower after sessions with BF than when BF was withdrawn. This result suggests that BF acted as a "reality check": in sessions without BF participant had the impression to be more in control of their HRV, but when BF was introduced again, they were reminded how hard it was. Thus, BF's main function could have been to remind participants to *voluntarily* engage in HRV upregulation (Bornemann et al., 2019), a skill they might otherwise forget to apply due to the action context.

Despite its efficacy to provide causal evidence of the BF mechanism, our experimental design did not test long term outcomes of the training, which could be expected to last more than 12 months (Andersen et al., 2018). This information would be critical to determine the frequency and need for "booster sessions" in the case of training implementation in the police curricula (Andersen et al., 2018). Ideally, future research should link the training to long-term evaluation in real life using ecological momentary assessment (EMA) and ecological psychophysiological assessment. EMA has been shown to produce higher quality results than standard assessments using a single questionnaire-based measurement (Moore et al., 2016) and can even be paired with interventions (McDevitt-Murphy et al., 2018). Due to the intensity of police work, we however expect that the interruptive nature of EMA might lower response rates if the assessment isn't fast and intuitive enough (Kini, 2013; Ponnada et al., 2017). We therefore recommend any EMA affordances to study long terms effects of BF training on police officers to be designed to fit the specific needs of this population (e.g., pairing EMA triggers with moments of lower intensity police work and after-action debriefing moments).

Strengths and limitations of using police trainers as a sample

Our game was validated in a population of police trainers, which differs from the *on-duty* police officer population originally intended in the grant proposal, especially in terms of use of force: Expert police officers have been shown to use less verbal and physical force than novices (R. R. Johnson et al., 2014; Paoline & Terrill, 2007).

This population's proficiency might have produced a ceiling effect on behavioral performance, which could explain the lack of behavioral results in our transfer task, as mentioned in chapter 5. Although choosing trainers was a decision related to COVID-lockdowns at the police academy, this population has allowed us to conclude that even experts benefit from HRV upregulation training. Additionally, while trainers might have been more inclined to review a new teaching tool positively, this group is also known to be highly critical about curricular changes (Lingamneni, 1979; J. Smith, 2022). Their positive rating of the training was therefore not granted, and is an encouraging result for the future of this intervention. However, trainers might have been a more motivated group than younger police recruits might be, hence further research on engagement of police recruits is needed.

Regarding the representativity of the sample, the number of recruited police trainers (N = 109) participating in the study described in chapter 5 represent nearly a quarter of the entire population of police trainers in the Netherlands. While this sample is large enough to be deemed representative of the Dutch police trainers' population, research on individual differences - relevant for personalized intervention development - would require larger numbers (Brydges, 2019).

Limitations concerning behavioral outcomes

Despite the positive training effects in terms of HRV upregulation and transfer of the HRV upregulation skill, a limitation of our study concerns *behavioral* transfer outside the VR game. Our study was not able to demonstrate the behavioral benefits of the training in new contexts in terms of decision-making. We argue in chapter 5 that this negative result could be attributed to the measurement reliability (only a single trial) of our transfer task. It was inspired by the dispatch priming task from Taylor (2020) and primarily designed to measure the susceptibility of police officers to be primed by a realistic radio-dispatch to shoot an innocent subject matching the dispatch information. This task matched behaviors trained *in-game*, as explained in chapters 1, 4 and 5. It additionally held the advantage of being practical for physiological regulation in an arousing and active decision-making police scenario without the need for confederates and external assessors. However, the task lacked the complexity of simulated situations, which have been shown to be sensitive to BF-induced changes in Canadian military and police personnel (Andersen et al., 2018; Andersen & Gustafsberg, 2016; Bouchard, Bernier, et al., 2012). In line with those results, we suggest future research to test the benefit of our training on HRV and behavior in real policing settings.

The search for an optimal "transfer task" evaluating the behavioral benefits of our BF training raises concerns regarding a theoretical limitation inherent to all research on police performance. What constitutes good policing? This interrogation has often been raised (J. M. Brown, 2013; Norris & Norris, 1993) and has been debated in terms of management, political agendas and societal impact. From a management perspective, the use of "performance indicators" of good policing have been both praised for their clarity and efficacy (Prenzler & Lewis, 2005) and criticized for their flaws in focusing on less meaningful data (Freiberg, 2005), as for example focusing on the number of arrests instead of preventive policing actions. Reliance on strict performance indicators is additionally thought to be too subject to political agendas. Since there is a political necessity to show tangible results, it is often requested to the police hierarchy to provide simple metrics proving police efficiency (e.g., number of arrests), which often results in removing from the police organization the required independence (e.g., resources allocation) to adopt policing strategies yielding actual results (Rogerson, 1995). Consequentially, most research on police, including ours, focuses on scales inspired by the societal impact of policing, focusing mainly on the excessive use of force (Andersen & Gustafsberg, 2016; Bennell et al., 2021; Cyr, 2016; Hine et al., 2018; Marenin, 2016; Rajakaruna et al., 2017; Staller et al., 2018; Taylor, 2020) and often relying on appraisal scores used in police academies. While those solutions are suitable to assess the benefit of a wide variety of training interventions, more research is needed to establish meaningful performance criteria of policing, especially when it comes to understanding the processes at play in police interactions, not merely its outcomes (Willis & Mastrofski, 2016). In our research, we measured behavioral transfer by using a task in which a radio dispatch priming was followed by a shoot/don't shoot decision. Thus, it falls in the category of outcome-oriented performance metric where participants have to perform a "split-second decision", which is deemed non-representative of real policing by many researchers (Garrett & Stoughton, 2017; Garrison, 2018; Murphy, 2014; White, 2016). Hence, as suggested before, I recommend for future research to use simulated scenarios with actors as a transfer task, which is both more representative of policing work and more sensitive to BF training (Andersen et al., 2018; Bouchard, Bernier, et al., 2012).

Considerations while selecting a biofeedback target

In addition to the difficulties of defining the proper outcome measures of good policing, the difficulties of translating findings from fundamental research on physiological mechanisms into actionable BF targets should be stressed. As mentioned in the introduction, our BF training was aimed at increasing vagally mediated parasympathetic activity through breathing-induced HRV. In the process of defining HRV as our BF target, we had to discard alternative candidates, such as attempting to train the freezing response itself. Freezing is characterized by an event-related anticipatory HR deceleration which is thought to aid information collection prior to action, a phenomenon of notable interest for police-relevant decision-making (Hashemi et al., 2019), as it is also suggested to improve perception (Lojowska et al., 2015). We additionally decided not to focus in our first efforts on electrophysiological cortical signals as frontal theta waves, a reliable event-related index of affective control (Cavanagh, 2014). Both measures are event-related biological markers, and hence most commonly measured through the averaging of several events, which precludes the possibility of providing close-to-instantaneous feedback, since several trials would have to be recorded to identify a pattern on which to provide feedback. Additionally, event-related markers are usually measured in a timespan that ranges a few seconds only. Therefore, even if through some innovative technique feedback could be provided on single trials, the event would still be very short lived. It would hence be impractical to provide this feedback to the user in a timeframe that allows regulation to take place, especially if an additional delay is produced by the need for artifact removal (see chapter 3). Hence, it is important to consider, when designing evidence-based interventions, that not all identified scientific measures are equally translatable into a training target. In our project, the evidence gathered in research focusing on event-related markers was still useful, but could only be indirectly used to determine a sound training target.

On the other hand, while BF parameters that are not event-related are easier to train, these parameters have not yet been thoroughly investigated in dynamic setups. For example, while a higher HRV is generally beneficial in many aspects of policing (Thayer et al., 2009), research still has to establish whether this holds for all aspects of policing, especially moments where use of force is actually required. Indeed, reduced HRV has been linked with temporary increases in force (Jerath et al., 2006), an advantage when handling violent perpetrators. HRV upregulation could however also be beneficial in the build-up moments leading to use of force (Andersen et al., 2018), in order to prevent the situation from escalating to an excessive degree of violence.

Finally, it has been shown that non-event-related electrophysiological cortical signals could be used as a neurofeedback target in contexts where movements are limited enough to prevent movement artifacts (Schoneveld et al., 2016). For instance, theta waves could be an interesting candidate to train attention and working memory

(Wang & Hsieh, 2013), as would parietal alpha power (Ros et al., 2017) as well as the EEG-based protocol from Keynan et al., (2019), which targets the EEG signature of the amygdala measured at scalp level to increase stress resilience. Police-specific neurofeedback trainings could additionally be aimed at reducing the burden of work-related mental injuries as post-traumatic stress disorder by combining effective VR protocols (Rizzo & Shilling, 2017) with effective neurofeedback protocols (Kluetsch et al., 2014). For the purpose of our research focusing on more dynamic *in-action* BF, however, I still consider our breathing-based HRV upregulation the most viable option as simplicity, freedom of movement, controllability and signal quality were all part of the training's requirements.

Future Directions for our Game

Thanks to the encouraging results from this thesis, and the interest and support from the Dutch police academy, a new version of our Decision Under Stress Training (DUST) will be developed to be operated and tested within the curricula of large numbers of police students. In the following section the main requirements for further development of the DUST project are shortly presented, along with a list of more explorative suggestions for other researchers aiming to develop *in-action* biofeedback trainings.

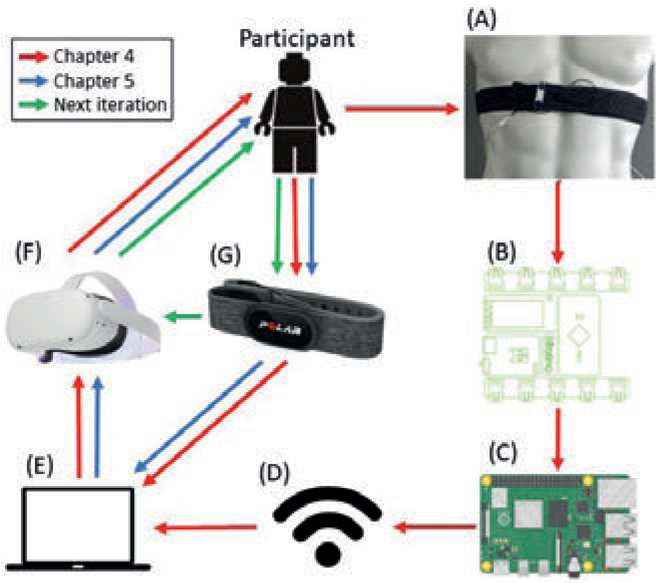
Simplification of the setup

Hardware. The hardware requirements for the training should be simplified, as simplicity is one of the leading factors of technological adoption (Holden & Karsh, 2010). In the first experiment (chapter 4) the hardware requirements were an important limitation: a breathing-belt sensor, an HR-belt sensor, two laptops (one of which with high graphic capacities) and a VR unit with external trackers were needed. In the following experiment (chapter 5) the setup was reduced by removing one laptop and the breathing belt, but remained bulky. As illustrated in Figure 6.1, thanks to recent advances in VR technology and the availability of new user grade products, the setup can be reduced considerably for future versions of the game. The next game iteration will be run directly on the VR goggles' hardware, thus reducing the setup to an HR-belt sensor and a single VR unit, without the need for external trackers. This set of improvements would also reduce the cost of a single setup from the 4000 Euros required in the first experiment to a moderate 300 euros, thus enhancing the deployment capacities for the system, as cost is often a hindrance for VR intervention scalability (Bush, 2008).

Software. The various programs required to run the training (real-time physiological analysis, the game and the information transfer software) will be merged together in a single program to make the training more plug-and-play. While separate programs were required in the first iterations (chapters 4 and 5) to reduce development costs and maintain the capacity to carefully monitor physiological recordings and data saving, those affordances came at the cost of simplicity and required extensive training of the experimenter to run the setup (see below). The new version of the game should have the capacity to automatically detect sensor misplacements and should feature enhanced artifact correction capacities to relieve the trainers from understanding the details of the physiological recordings. The tutorial at the beginning of the game should additionally be expanded to improve the "onboarding" of the trainee by explaining the scope of the training and the feedback scores at the end of the training.

Figure 6.1

Overview of the hardware components involved in the BF closed-loop system.



Note. Three successions of implementation of the biofeedback (BF), in consecutive studies. In chapter 4 the biofeedback was directed at breathing pace and involved the path in red: (A) a respiratory inductance plethysmography belt sent the analog breathing signal to (B) a BITalino (r)evolution board. This signal was then analyzed in (C) a Raspberry Pi 4 Model B, and the processed BF signal provided via (D) Wi-Fi to (E) a laptop that ran the game displayed to the player through (F) VR goggles. Simultaneously, the HR was recorded through (G) a polar H10 belt, which was connected to the gaming laptop via Bluetooth. The biofeedback procedure in chapter 5 only used the blue path, as BF was based directly on the HR signal. The next iteration of our game will be implemented on the VR goggles directly, thus simplifying the setup further, as illustrated by the green path.

Trainers' preparation

The current setup required the experimenters to be present throughout the whole training, provide extensive information to participants, and carefully monitor the quality and saving of the incoming physiological data. As explained earlier, the experimenter also played the role of a positive coach between sessions. The new version of the software would alleviate most of those burdens. Importantly, the participants in the experiments of this thesis were police trainers, who already had prior knowledge of the effect of stressors on physiology and behavior (van der. Velden, 2014). Hence, our training was accompanied with little-to-no psycho-educational material when compared to other BF trainings aimed at police officers-in-training (Andersen et al., 2018; Di Nota et al., 2021). Such materials notably include information about the effect of stress responses on physiology, the consequences of stress induction on behavior, or also the physiological base of HRV and its relation to police work. The upcoming implementation trials are expected to take place in police academies with police students as the primary target population for the training. Therefore, the tutorial of the next version of the game should compensate for the lack of knowledge of this new type of participants, by introducing necessary notions while immersed in the VR environment.

Using the game for other populations

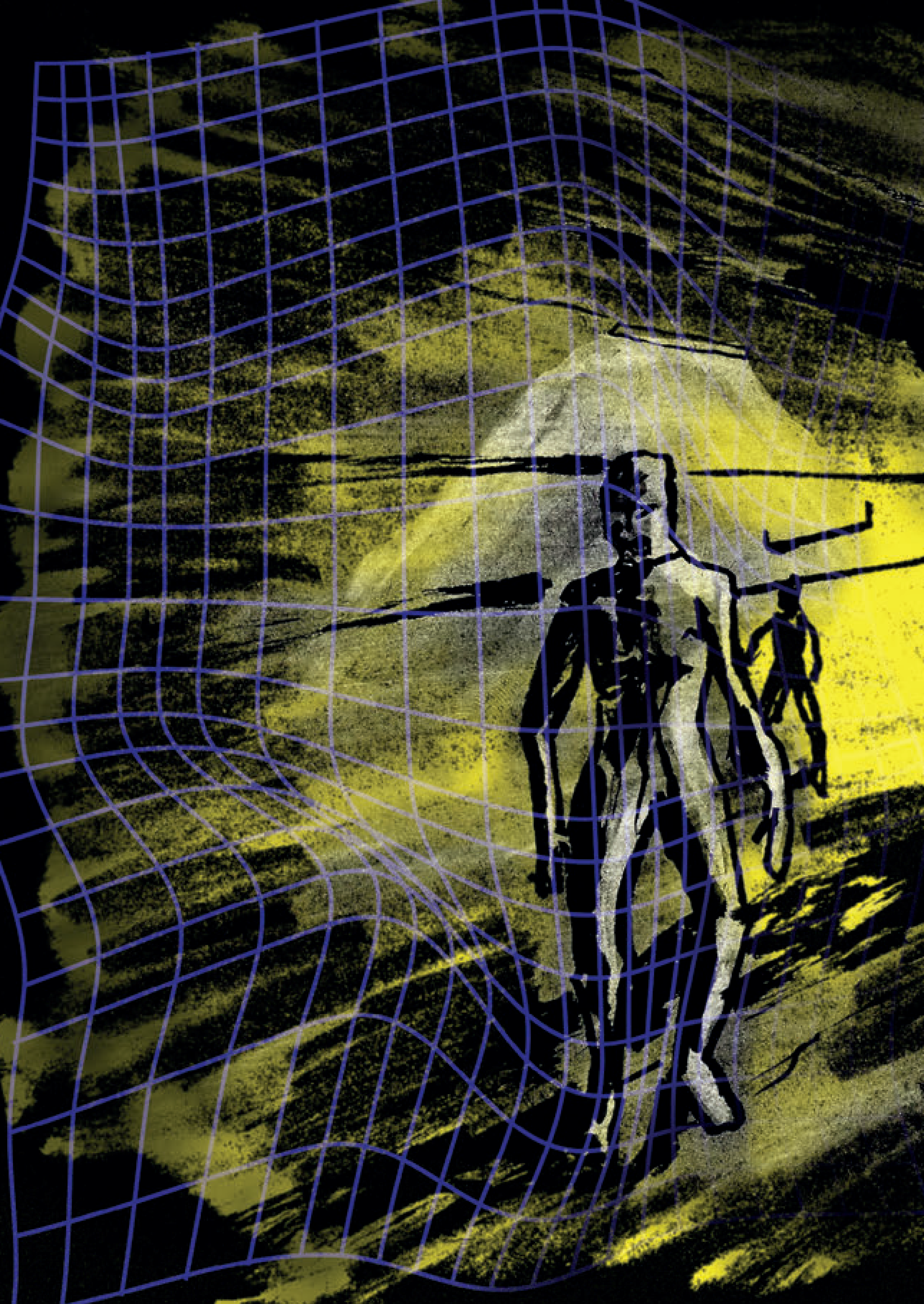
The VR game presented in this thesis was designed for police officers. While many elements of the game are designed to maximize the transfer of the trained HRV skill to policing contexts, the core BF dynamic could easily be used for different purposes. Indeed, the work-related stress as experienced by police officers is also experienced by a wider range of first-responders (Kleim & Westphal, 2011). Paramedics, military personnel, firefighters, and many others could therefore benefit from our VR training game. Most of the VR-based trainings for first responders so far mainly focus on scenario-based trainings mimicking real situations (Xie et al., 2021), which have a harder time eliciting engagement and arousal. Since our game already proved its efficacy in eliciting arousal, game content modifications should aim to adapt the content to a variety of first responders and only replace the police specific elements disseminated in the game (radio dispatch format, reason for shooting at zombies, replacing a gun with a firefighting nozzle, and so on). Additionally, the game could potentially also be used for improving mental health in the general population, to reduce the high incidence of stress-related disorders (Gradus, 2017), as suggested by the encouraging results of prior work on the benefits of game-based biofeedback interventions for mental health in youth (Bossenbroek et al., 2020; Lobel et al., 2016; Schoneveld et al., 2016; Weerdmeester, 2021).

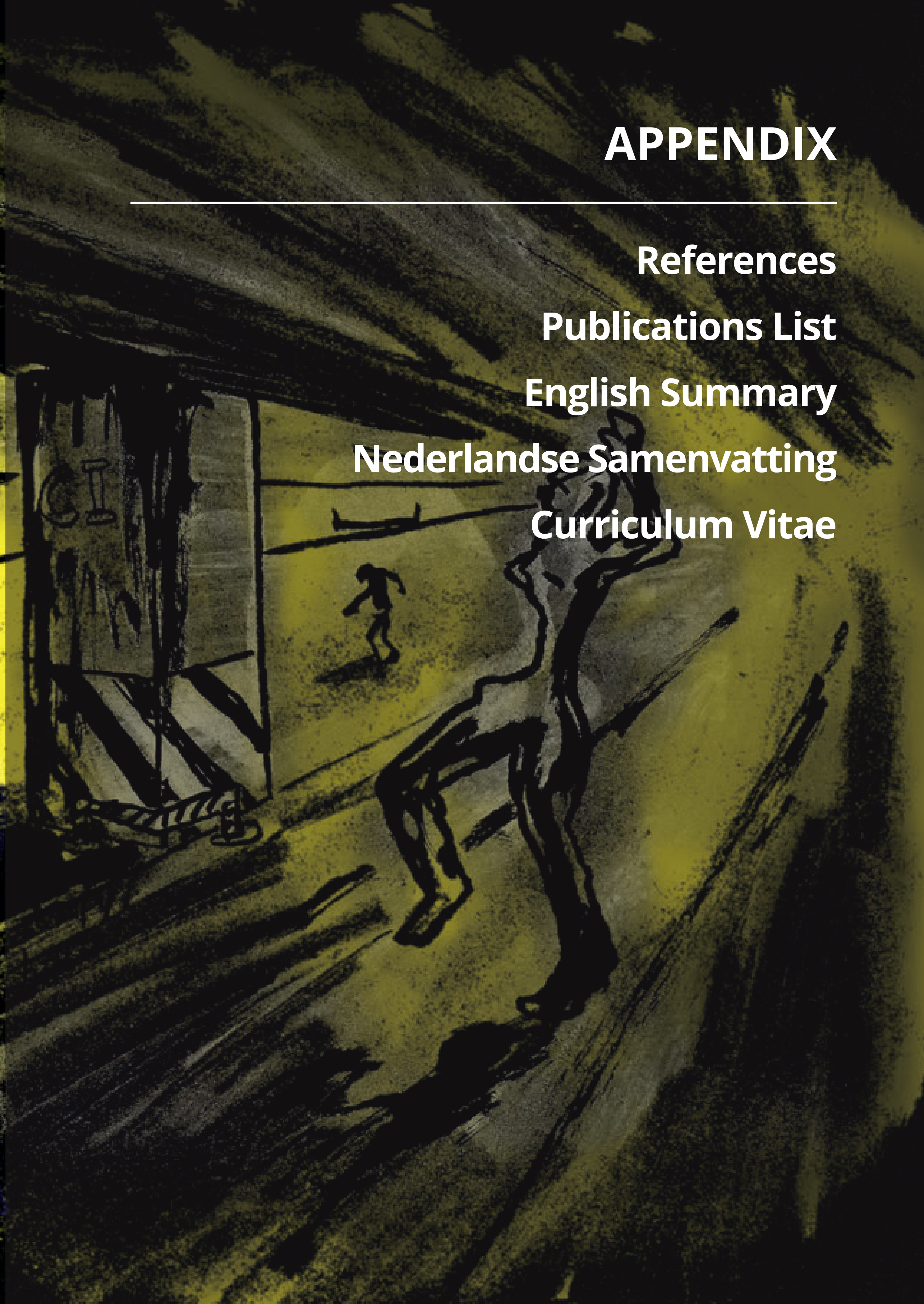
Closing statements

The current thesis presented the design, implementation, and testing of an *in-action* VR biofeedback training game, aimed at helping police officers to upregulate their HRV. We assessed the efficacy of the training to increase HRV both within and outside of the VR game training environment (chapter 4 and 5). Our results in chapter 5 demonstrated the trainability of HRV upregulation in an action context, as well as skill-transfer, and provided insights into the causal role of biofeedback in improving HRV upregulation.

Still, further research is needed to specifically assess the behavioral benefits of our training—especially on the long term—as well as its efficacy in a police education context. I additionally encourage future researchers investigating bio- and neurofeedback to consider the use of an addition and withdrawal design to both accurately test and increase the transfer effects, and a multiple baselines design to strengthen the experimental capacity for causal inferences, as discussed above and in chapter 5. Regarding *in-action* BF, I recommend researchers to carefully consider which training parameter to select by reviewing our considerations stated in chapter 3. Moreover, I recommend researchers using VR to carefully consider the level of realism that their experimental setup requires, to avoid disengaging the participants with uncanny experiences, as thoroughly explained in chapter 2. Lastly, I encourage working in multidisciplinary teams including game designers and stakeholders when planning to embed a training procedure within a game context, to maximize the impact of the intervention for a greater societal benefit.

To conclude, the positive reception of the game by the police trainers who participated in the studies has ultimately led the Dutch Police Academy to further support the development of the game, to facilitate its deployment in police education and training. Thus, with the methods and the approach described in this thesis I hope to have contributed towards the resilience and competence of Dutch police and, potentially, other groups that are faced with serious stressors in the line of duty.





APPENDIX

REFERENCES

1. *Actie ZeeHond*. (n.d.). Retrieved July 30, 2021, from https://www.actievoorgeleidehonden.nl/campagne/actie-zeehond
2. Adang, O., & Timmer, J. (2005). Beheersing van Gevaar; Praktijkboek voor de opleiding van vuurwapendragenden en de toetsing van geweldstoepassing. In *Reed Business, Reed elsevier*.
3. Admon, R., Lubin, G., Stern, O., Rosenberg, K., Sela, L., Ben-Ami, H., & Hendler, T. (2009). Human vulnerability to stress depends on amygdala's predisposition and hippocampal plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(33), 14120–14125. https://doi.org/10.1073/pnas.0903183106
4. Admon, R., Milad, M. R., & Hendler, T. (2013). A causal model of post-traumatic stress disorder: Disentangling predisposed from acquired neural abnormalities. In *Trends in Cognitive Sciences* (Vol. 17, Issue 7, pp. 337–347). Elsevier. https://doi.org/10.1016/j.tics.2013.05.005
5. Allcoat, D., Greville, W. J., Newton, P. M., & Dymond, S. (2015). Frozen with fear: Conditioned suppression in a virtual reality model of human anxiety. *Behavioural Processes*, *118*, 98–101. https://doi.org/10.1016/j.beproc.2015.06.011
6. Allcoat, D., & von Mühlenen, A. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology*, *26*(1063519), 1–13. https://doi.org/10.25304/rlt.v26.2140
7. Andersen, J. P., Di Nota, P. M., Beston, B., Boychuk, E. C., Gustafsberg, H., Poplawski, S., & Arpaia, J. (2018). Reducing lethal force errors by modulating police physiology. *Journal of Occupational and Environmental Medicine*, *60*(10), 867–874. https://doi.org/10.1097/JOM.0000000000001401
8. Andersen, J. P., & Gustafsberg, H. (2016). A Training Method to Improve Police Use of Force Decision Making: A Randomized Controlled Trial. *SAGE Open*, *6*(2). https://doi.org/10.1177/2158244016638708
9. Andersen, J. P., Papazoglou, K., Koskelainen, M., Nyman, M., Gustafsberg, H., & Arnetz, B. B. (2015). Applying Resilience Promotion Training Among Special Forces Police Officers. *SAGE Open*, *5*(2). https://doi.org/10.1177/2158244015590446
10. Arnsten, A. F. T. (2015). Stress weakens prefrontal networks: Molecular insults to higher cognition. In *Nature Neuroscience* (Vol. 18, Issue 10, pp. 1376–1385). Nat Neurosci. https://doi.org/10.1038/nn.4087
11. Attig, C., Rauh, N., Franke, T., & Krems, J. F. (2017). System latency guidelines then and now – Is zero latency really considered necessary? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10276 LNAI*, 3–14. https://doi.org/10.1007/978-3-319-58475-1_1/TABLES/1
12. Aue, T., & Okon-Singer, H. (2015). Expectancy biases in fear and anxiety and their link to biases in attention. *Clinical Psychology Review*, *42*, 83–95. https://doi.org/10.1016/j.cpr.2015.08.005
13. Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*(MAR), 456. https://doi.org/10.3389/fpsyg.2017.00456
14. Bakdash, J. Z., & Marusich, L. R. (2021). *CRAN - Package rmcorr*.
15. Baldwin, S., Bennell, C., Andersen, J. P., Semple, T., & Jenkins, B. (2019). Stress-Activity Mapping: Physiological Responses During General Duty Police Encounters. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02216
16. Baldwin, S., Bennell, C., Blaskovits, B., Brown, A., Jenkins, B., Lawrence, C., McGale, H., Semple, T., & Andersen, J. P. (2022). A Reasonable Officer: Examining the Relationships Among Stress, Training, and Performance in a Highly Realistic Lethal Force Scenario. *Frontiers in Psychology*, *12*, 5629. https://doi.org/10.3389/fpsyg.2021.759132
17. Balzarotti, S., Biassoni, F., Colombo, B., & Ciceri, M. R. (2017). Cardiac vagal control as a marker of emotion regulation in healthy adults: A review. *Biological Psychology*, *130*(October), 54–66. https://doi.org/10.1016/j.biopsycho.2017.10.008

18. Baños, R. M., Botella, C., & Perpiñá, C. (1999). Virtual reality and psychopathology. *Cyberpsychology and Behavior*, *2*(4), 283–292. https://doi.org/10.1089/cpb.1999.2.283
19. Baños, R. M., Liaño, V., Botella, C., Alcañiz, M., Guerrero, B., & Rey, B. (2006). Changing induced moods via virtual reality. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3962 LNCS*, 7–15. https://doi.org/10.1007/11755494_3
20. Baranowski, T., Blumberg, F., Buday, R., DeSmet, A., Fiellin, L. E., Green, C. S., Kato, P. M., Lu, A. S., Maloney, A. E., Mellecker, R., Morrill, B. A., Peng, W., Shegog, R., Simons, M., Staiano, A. E., Thompson, D., & Young, K. (2016). Games for Health for Children - Current Status and Needed Research. *Games for Health Journal*, *5*(1), 1–12. https://doi.org/10.1089/g4h.2015.0026
21. Batista, D., Silva, H., & Fred, A. (2017). Experimental characterization and analysis of the BITalino platforms against a reference device. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2418–2421. https://doi.org/10.1109/EMBC.2017.8037344
22. Ben-Menachem, E., Revesz, D., Simon, B. J., & Silberstein, S. (2015). Surgically implanted and non-invasive vagus nerve stimulation: a review of efficacy, safety and tolerability. *European Journal of Neurology*, *22*(9), 1260–1268. https://doi.org/10.1111/ENE.12629
23. Bennell, C., Alpert, G., Andersen, J. P., Arpaia, J., Huhta, J.-M. M., Kahn, K. B., Khanizadeh, A.-J. J., McCarthy, M., McLean, K., Mitchell, R. J., Nieuwenhuys, A., Palmer, A., & White, M. D. (2021). Advancing police use of force research and practice: urgent issues and prospects. *Legal and Criminological Psychology*, *26*(2), 121–144. https://doi.org/10.1111/lcrp.12191
24. Ben-Tal, A., Shamailov, S. S., & Paton, J. F. R. (2014). Central regulation of heart rate and the appearance of respiratory sinus arrhythmia: New insights from mathematical modeling. *Mathematical Biosciences*, *255*(1), 71–82. https://doi.org/10.1016/j.mbs.2014.06.015
25. Binder, A., & Scharf, P. (1980). Violent Police-citizen Encounter. *Annals of the American Academy of Political and Social Science*, *452*, 111–121.
26. Blanchard, D. C., Griebel, G., Pobbe, R., & Blanchard, R. J. (2011). Risk assessment as an evolved threat detection and analysis process. *Neuroscience and Biobehavioral Reviews*, *35*(4), 991–998. https://doi.org/10.1016/j.neubiorev.2010.10.016
27. Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology: Just another methodological tool for social psychology? *Psychological Inquiry*, *13*(2), 146–149. https://doi.org/10.1207/S15327965PLI1302_03
28. Blum, J., Rockstroh, C., & Göritz, A. S. (2019). Heart rate variability biofeedback based on slow-paced breathing with immersive virtual reality nature scenery. *Frontiers in Psychology*, *10*(SEP). https://doi.org/10.3389/fpsyg.2019.02172
29. Boesch, M., Sefidan, S., Ehlert, U., Annen, H., Wyss, T., Steptoe, A., & La Marca, R. (2014). Mood and autonomic responses to repeated exposure to the Trier Social Stress Test for Groups (TSST-G). *Psychoneuroendocrinology*, *43*, 41–51. https://doi.org/10.1016/j.psyneuen.2014.02.003
30. Borgeat, F., Hade, B., Larouche, L. M., & Bedwani, C. N. (1980). Effect of therapist's active presence on EMG biofeedback training of headache patients. *Biofeedback and Self-Regulation*, *5*(2), 275–282. https://doi.org/10.1007/BF00998603
31. Bornemann, B., Kok, B. E., Böckler, A., & Singer, T. (2016). Helping from the heart: Voluntary upregulation of heart rate variability predicts altruistic behavior. *Biological Psychology*, *119*, 54–63. https://doi.org/10.1016/j.biopsycho.2016.07.004
32. Bornemann, B., Kovacs, P., & Singer, T. (2019). Voluntary upregulation of heart rate variability through biofeedback is improved by mental contemplative training. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-44201-7
33. Bossenbroek, R., Wols, A., Weerdmeester, J., Lichtwarck-Aschoff, A., Granic, I., & van Rooij, M. M. J. W. (2020). Efficacy of a Virtual Reality Biofeedback Game (DEEP) to Reduce Anxiety and Disruptive Classroom Behavior: Single-Case Study. *JMIR Mental Health*, *7*(3), e16066. https://doi.org/10.2196/16066

34. Botella, C., Osma, J., Palacios, A. G., Guillén, V., & Baños, R. (2008). Treatment of complicated grief using virtual reality: a case report. *Death Studies*, *32*(7), 674–692. https://doi.org/10.1080/07481180802231319
35. Bouchard, S., Bernier, F., Boivin, É., Morin, B., & Robillard, G. (2012). Using biofeedback while immersed in a stressful videogame increases the effectiveness of stress management skills in soldiers. *PLoS ONE*, *7*(4). https://doi.org/10.1371/journal.pone.0036169
36. Bouchard, S., Guitard, T., Laforest, M., Dumoulin, S., Boulanger, J., & Bernier, F. (2012). The Potential of Stress Management Training as a Coping Strategy for Stressors Experienced in Theater of Operation : A Systematic Review. *Post Traumatic Stress Disorders in a Global Context*, 271–286.
37. Braga, A. A., & Weisburd, D. (2019). Conclusion: Police Innovation and the Future of Policing. In *Police Innovation: Contrasting Perspectives* (pp. 544–563). Cambridge University Press. https://doi.org/10.1017/9781108278423.024
38. Brammer, J. C. (n.d.). *GitHub - JanCBrammer/OpenHRV: Heart rate variability (HRV) biofeedback with Polar ECG chest straps.* Retrieved April 20, 2022, from https://github.com/JanCBrammer/OpenHRV
39. Brammer, J. C. (2020a). biopeaks: a graphical user interface for feature extraction from heart- and breathing biosignals. *Journal of Open Source Software*, *5*(54), 2621. https://doi.org/10.21105/joss.02621
40. Brammer, J. C. (2020b). *biofeedback_analyses/hrv_utils.py at c9b2ffd067d2e9197c5750f0ae7749ebb7918366 · JanCBrammer/biofeedback_analyses · GitHub*. https://github.com/JanCBrammer/biofeedback_analyses/blob/c9b2ffd067d2e9197c5750f0ae7749ebb7918366/biofeedback_analyses/analysis_utils/hrv_utils.py#L112
41. Brammer, J. C., van Peer, J. M., Michela, A., van Rooij, M. M. J. W., Oostenveld, R., Klumpers, F., Dorrestijn, W., Granic, I., & Roelofs, K. (2021). Breathing Biofeedback for Police Officers in a Stressful Virtual Environment: Challenges and Opportunities. *Frontiers in Psychology*, *12*, 586553. https://doi.org/10.3389/fpsyg.2021.586553
42. Bregman, N. J., & McAllister, H. A. (1983). Voluntary control of skin temperature: Role of experimenter presence versus absence. *Biofeedback and Self-Regulation*, *8*(4), 543–546. https://doi.org/10.1007/BF00998761
43. Brehmer, B. (1992). Dynamic decision making: human control of complex systems. *Acta Psychologica*, *81*(3), 211–241. https://doi.org/10.1016/0001-6918(92)90019-A
44. Brown, J. M. (2013). The Future of Policing. *The Future of Policing*, 1–522. https://doi.org/10.4324/9780203435946
45. Brown, T. E., Beightol, L. A., Koh, J., & Eckberg, D. L. (1993). Important influence of respiration on human R-R interval power spectra is largely ignored. *Journal of Applied Physiology*, *75*(5), 2310–2317. https://doi.org/10.1152/jappl.1993.75.5.2310
46. Brydges, C. R. (2019). Effect Size Guidelines, Sample Size Calculations, and Statistical Power in Gerontology. *Innovation in Aging*, *3*(4), 1–8. https://doi.org/10.1093/GERONI/IGZ036
47. Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. M.-A., Dawson, D. R., Anderson, N. D., Gilbert, S. J., Dumontheil, I., & Channon, S. (2006). *The case for the development and use of ecologically valid measures of executive function*. *2006*, 194–209. https://doi.org/10.10170S1355617706060310
48. Burish, T. G., & Schwartz, D. P. (1980). EMG biofeedback training, transfer of training, and coping with stress. *Journal of Psychosomatic Research*, *24*(2), 85–96. https://doi.org/10.1016/0022-3999(80)90058-6
49. Bürkner, P. C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/JSS.V080.I01
50. Bush, J. (2008). Viability of virtual reality exposure therapy as a treatment alternative. *Computers in Human Behavior*, *24*(3), 1032–1040. https://doi.org/10.1016/j.chb.2007.03.006

51. Caballero Sánchez, C., Barbado Murillo, D., Davids, K., & Moreno Hernández, F. J. (2016). Variations in task constraints shape emergent performance outcomes and complexity levels in balancing. *Experimental Brain Research*, *234*(6), 1611–1622. https://doi.org/10.1007/s00221-016-4563-2
52. Carleton, R. N., Afifi, T. O., Turner, S., Taillieu, T., Duranceau, S., LeBouthillier, D. M., Sareen, J., Ricciardelli, R., MacPhee, R. S., Groll, D., Hozempa, K., Brunet, A., Weekes, J. R., Griffiths, C. T., Abrams, K. J., Jones, N. A., Beshai, S., Cramm, H. A., Dobson, K. S., … Asmundson, G. J. G. (2018). Mental Disorder Symptoms among Public Safety Personnel in Canada. *Canadian Journal of Psychiatry*, *63*(1), 54–64. https://doi.org/10.1177/0706743717723825
53. Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/JSS.V076.I01
54. Cavanagh, J. F. (2014). Frontal Theta as a Mechanism for Affective and Effective Control. *Psychophysiology*, *51*(8), S8–S8. https://doi.org/10.1016/j.tics.2014.04.012.Frontal
55. Ceci, S. J., Kahan, D. M., & Braman, D. (2010). The WEIRD are even weirder than you think: Diversifying contexts is as important as diversifying samples. *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X
56. Cheetham, M. (2017). Editorial: The uncanny valley hypothesis and beyond. *Frontiers in Psychology*, *8*(OCT), 1–3. https://doi.org/10.3389/fpsyg.2017.01738
57. Chen, H.-C., Chou, F. H.-C., Chen, M.-C., Su, S.-F., Wang, S.-Y., Feng, W.-W., Chen, P.-C., Lai, J.-Y., Chao, S.-S., Yang, S.-L., Tsai, T.-C., Tsai, K.-Y., Lin, K.-S., Lee, C.-Y., & Wu, H.-C. (2013). A Survey of Quality of Life and Depression for Police Officers in Kaohsiung, Taiwan. *Quality of Life Research 2006 15:5*, *15*(5), 925–932. https://doi.org/10.1007/S11136-005-4829-9
58. Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, *9*(NOV), 1–20. https://doi.org/10.3389/fpsyg.2018.02086
59. Cokelaer, T., & Hasch, J. (2017). “Spectrum”: Spectral Analysis in Python. *The Journal of Open Source Software*, *2*(18), 348. https://doi.org/10.21105/JOSS.00348
60. Cornwell, B. R., Garrido, M. I., Overstreet, C., Pine, D. S., & Grillon, C. (2017). The Unpredictive Brain Under Threat: A Neurocomputational Account of Anxious Hypervigilance. *Biological Psychiatry*, *82*(6), 447–454. https://doi.org/10.1016/j.biopsych.2017.06.031
61. Cummings, J. J., & Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, *19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740
62. Cyr, K. (2016). Police Use of Force: Assessing Necessity and Proportionality. *Alberta Law Review*, 663. https://doi.org/10.29173/alr424
63. Davenport, D. R. (1999). Environmental constraints and organizational outcomes: Modeling communities of municipal police departments. *Police Quarterly*, *2*(2), 174–200. https://doi.org/10.1177/109861119900200203
64. Davies, A. J. (2017). Shoot/do not shoot–what are the influences? The police recruit perspective. *Policing and Society*, *27*(5), 494–507. https://doi.org/10.1080/10439463.2015.1077835
65. de Carvalho, M. R., Dias, T. R. de S., Duchesne, M., Nardi, A. E., & Appolinario, J. C. (2017). Virtual Reality as a Promising Strategy in the Assessment and Treatment of Bulimia Nervosa and Binge Eating Disorder: A Systematic Review. *Behavioral Sciences (Basel, Switzerland)*, *7*(3). https://doi.org/10.3390/bs7030043
66. de la Rosa, S., & Breidt, M. (2018). Virtual reality: A new track in psychological research. *British Journal of Psychology*, *109*(3), 427–430. https://doi.org/10.1111/bjop.12302

67. de Voogd, L. D., Hagenberg, E., Zhou, Y. J., de Lange, F. P., & Roelofs, K. (2022). Acute threat enhances perceptual sensitivity without affecting the decision criterion. *Scientific Reports 2022 12:1*, *12*(1), 1–11. https://doi.org/10.1038/s41598-022-11664-0
68. De Witte, N. A. J., Buyck, I., & Van Daele, T. (2019). Combining Biofeedback with Stress Management Interventions: A Systematic Review of Physiological and Psychological Effects. *Applied Psychophysiology and Biofeedback 2019 44:2*, *44*(2), 71–82. https://doi.org/10.1007/S10484-018-09427-7
69. Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*, *210*(October 2017), 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005
70. Di Nota, P. M., Anderson, G. S., Ricciardelli, R., Carleton, R. N., & Groll, D. (2020). Mental disorders, suicidal ideation, plans and attempts among Canadian police. *Occupational Medicine*, *70*(3), 183–190. https://doi.org/10.1093/OCCMED/KQAA026
71. Di Nota, P. M., Arpaia, J., Boychuk, E. C., Collins, P. I., & Andersen, J. P. (2021). Testing the Efficacy of a 1-Day Police Decision-Making and Autonomic Modulation Intervention: A Pragmatic Randomized Control Trial. *Frontiers in Psychology*, *this issue*, 3490. https://doi.org/10.3389/FPSYG.2021.719046
72. Di Nota, P. M., & Huhta, J. M. (2019). Complex motor learning and police training: Applied, cognitive, and clinical perspectives. In *Frontiers in Psychology* (Vol. 10, Issue JULY, p. 1797). Frontiers. https://doi.org/10.3389/fpsyg.2019.01797
73. Di Palma, M., Arcangeli, E., Lattanzi, D., Gabbiadini, A., Gallucci, M., Cuppini, R., Minelli, A., & Berlingeri, M. (2019). Heart Rate Variability reveals the fight between racially biased and politically correct behaviour. *Scientific Reports 2019 9:1*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-47888-w
74. Dirkin, G. R. (1983). Cognitive Tunneling: Use of Visual Infromation Under Stress. *Perceptual and Motor Skills*, *56*, 191–198.
75. Driskell, J. E., & Johnston, J. H. (1998). Stress exposure training. In *Stress Exposure Training. In: Cannon-Bowers, J.A. and Salas, E., Eds., Making Decisions under Stress: Implications for Individual and Team Training, American Psychological Association, Washington DC,* (pp. 191–217). American Psychological Association. https://doi.org/10.1037/10278-007
76. Driskell, J. E., Johnston, J. H., & Salas, E. (2001). Does stress training generalize to novel settings? *Human Factors*, *43*(1), 99–110. https://doi.org/10.1518/001872001775992471
77. Ebbinghaus, B. (2005). When less is more selection problems in large-N and small-N cross-national comparisons. *International Sociology*, *20*(2), 133–152. https://doi.org/10.1177/0268580905052366
78. Eliasson, M. (2021). Experience, Seniority and Gut Feeling—A Qualitative Examination of How Swedish Police Officers Perceive They Value, Evaluate and Manage Knowledge When Making Decisions. *Frontiers in Education*, *6*(November), 1–11. https://doi.org/10.3389/feduc.2021.731320
79. Ellrich, K., Baier, D., & Pfeiffer, C. (2011). Gewalt gegen Polizeibeamte. Befunde zu Einsatzbeamten, Situationsmerkmalen und Folgen von Gewaltübergriffen [Violence against police officers. Findings on officers, situational characteristics and consequences of violent assaults]. *KfN Forschungsbericht Nr.3*, *3*(3), 44–45.
80. Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *56*(2), 171–184. https://doi.org/10.1016/J.IJPSYCHO.2004.11.003
81. Feda, D. M., & Roemmich, J. N. (2016). Effect of Interpersonal and Cognitive Stressors on Habituation and the Utility of Heart Rate Variability to Measure Habituation. *Stress and Health*, *32*(4), 320–327. https://doi.org/10.1002/SMI.2618
82. Fekedulegn, D., Burchfiel, C. M., Charles, L. E., Hartley, T. A., Andrew, M. E., & Violanti, J. M. (2016). Shift work and sleep quality among urban police officers. *Journal of Occupational and Environmental Medicine*, *58*(3), e66–e71. https://doi.org/10.1097/JOM.0000000000000620

83. Flake, J. K., Pek, J., & Hehman, E. (2017). Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063
84. Föcker, J., Cole, D., Beer, A. L., & Bavelier, D. (2018). Neural bases of enhanced attentional control: Lessons from action video game players. *Brain and Behavior*, *8*(7). https://doi.org/10.1002/brb3.1019
85. Forte, G., Morelli, M., & Casagrande, M. (2021). Heart rate variability and decision-making: Autonomic responses in making decisions. *Brain Sciences*, *11*(2), 1–12. https://doi.org/10.3390/brainsci11020243
86. Forte, G., Morelli, M., Grässler, B., & Casagrande, M. (2022). Decision making and heart rate variability: A systematic review. *Applied Cognitive Psychology*, *36*(1), 100–110. https://doi.org/10.1002/acp.3901
87. Fragkaki, I., Roelofs, K., Stins, J., Jongedijk, R. A., & Hagenaars, M. A. (2017). Reduced freezing in posttraumatic stress disorder patients while watching affective pictures. *Frontiers in Psychiatry*, *8*(MAR), 1–9. https://doi.org/10.3389/fpsyt.2017.00039
88. Franzen, M. D., & Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychological assessment. In *Ecological validity of neuropsychological testing.* (pp. 91–112). Gr Press/St Lucie Press, Inc.
89. Freiberg, A. (2005). Managerialism in Australian criminal justice: RIP for KPIs? *Monash University Law Review*, *31*(1), 12–36. https://search.informit.org/doi/abs/10.3316/ielapa.200802263
90. Fromberger, P., Jordan, K., & Müller, J. L. (2018). Virtual reality applications for diagnosis, risk assessment and therapy of child abusers. *Behavioral Sciences and the Law*, *36*(2), 235–244. https://doi.org/10.1002/bsl.2332
91. Gamble, K. R., Vettel, J. M., Patton, D. J., Eddy, M. D., Caroline Davis, F., Garcia, J. O., Spangler, D. P., Thayer, J. F., & Brooks, J. R. (2018). Different profiles of decision making and physiology under varying levels of stress in trained military personnel. *International Journal of Psychophysiology*, *131*(March), 73–80. https://doi.org/10.1016/j.ijpsycho.2018.03.017
92. Garrett, B., & Stoughton, S. (2017). A tactical fourth amendment. In *Virginia Law Review* (Vol. 103, Issue 2, pp. 211–307). https://scholarship.law.duke.edu/faculty_scholarship/3840
93. Garrison, A. H. (2018). Criminal Culpability, Civil Liability, and Police Created Danger: Why and How the Fourth Amendment Provides Very Limited Protection From Police Use of Deadly Force. *Geo. Mason U. Civ. Rts. L. J.*, *28*, 1–57. https://papers.ssrn.com/abstract=4394546
94. Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2019). RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. *European Journal of Applied Physiology*, *119*(7), 1525–1532. https://doi.org/10.1007/s00421-019-04142-5
95. Gladwin, T. E., Hashemi, M. M., van Ast, V., & Roelofs, K. (2016). Ready and waiting: Freezing as active action preparation under threat. *Neuroscience Letters*, *619*, 182–188. https://doi.org/10.1016/j.neulet.2016.03.027
96. Goessl, V. C., Curtiss, J. E., & Hofmann, S. G. (2017). The effect of heart rate variability biofeedback training on stress and anxiety: A meta-analysis. *Psychological Medicine*, *47*(15), 2578–2586. https://doi.org/10.1017/S0033291717001003
97. Gorini, A., Capideville, C. S., De Leo, G., Mantovani, F., & Riva, G. (2011). The role of immersion and narrative in mediated presence: the virtual hospital experience. *Cyberpsychology, Behavior and Social Networking*, *14*(3), 99–105. https://doi.org/10.1089/cyber.2010.0100
98. Gradus, J. L. (2017). Prevalence and prognosis of stress disorders: A review of the epidemiologic literature. *Clinical Epidemiology*, *9*, 251–260. https://doi.org/10.2147/CLEP.S106250
99. Graef, Roger. (1990). *Talking Blues: Police in their Own Words* (02.08.1990). HarperCollins Publishers.

100. Gummidela, V. N. C., Silva, D. R. da C., & Gutierrez-Osuna, R. (2021). Evaluating the Role of Breathing Guidance on Game-Based Interventions for Relaxation Training. *Frontiers in Digital Health*, *3*, 188. https://doi.org/10.3389/fdgth.2021.760268
101. Habersaat, S., Abdellaoui, S. H., & Wolf, J. M. (2021). Social desirability, stress and health in police officers: preliminary results. *Policing*, *44*(2), 213–229. https://doi.org/10.1108/PIJPSM-08-2020-0133
102. Hagenaars, M. A., Oitzl, M., & Roelofs, K. (2014). Updating freeze: Aligning animal and human research. *Neuroscience and Biobehavioral Reviews*, *47*, 165–176. https://doi.org/10.1016/j.neubiorev.2014.07.021
103. Hansen, A. L., Johnsen, B. H., & Thayer, J. F. (2009). Relationship between heart rate variability and cognitive function during threat of shock. *Anxiety, Stress and Coping*, *22*(1), 77–89. https://doi.org/10.1080/10615800802272251
104. Harrison, L. A., Ahn, C., & Adolphs, R. (2015). Exploring the structure of human defensive responses from judgments of threat scenarios. *PLoS ONE*, *10*(8), 1–24. https://doi.org/10.1371/journal.pone.0133682
105. Hashemi, M. M., Gladwin, T. E., de Valk, N. M., Zhang, W., Kaldewaij, R., van Ast, V., Koch, S. B. J., Klumpers, F., & Roelofs, K. (2019). Neural Dynamics of Shooting Decisions and the Switch from Freeze to Fight. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-40917-8
106. Hashemi, M. M., Zhang, W., Kaldewaij, R., Koch, S. B. J., Smit, A., Figner, B., Jonker, R., Klumpers, F., & Roelofs, K. (2021). Human defensive freezing: associations with hair cortisol and trait anxiety. *Psychoneuroendocrinology*, *133*, 105417. https://doi.org/10.1016/j.psyneuen.2021.105417
107. Hassan, U., & Anwar, M. S. (2010). Reducing noise by repetition: Introduction to signal averaging. *European Journal of Physics*, *31*(3), 453–465. https://doi.org/10.1088/0143-0807/31/3/003
108. Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619
109. Hayano, J., & Yuda, E. (2019). Pitfalls of assessment of autonomic function by heart rate variability. In *Journal of Physiological Anthropology* (Vol. 38, Issue 1, pp. 1–8). BioMed Central. https://doi.org/10.1186/s40101-019-0193-2
110. Hermans, E. J., Henckens, M. J. A. G., Joels, M., & Fernandez, G. (2014). Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends in Neurosciences*, *37*(6), 304–314. https://doi.org/10.1016/j.tins.2014.03.006
111. Hermans, E. J., Van Marle, H. J. F., Ossewaarde, L., Henckens, M. J. A. G., Qin, S., Van Kesteren, M. T. R., Schoots, V. C., Cousijn, H., Rijpkema, M., Oostenveld, R., & Fernández, G. (2011). Stress-related noradrenergic activity prompts large-scale neural network reconfiguration. *Science*, *334*(6059), 1151–1153. https://doi.org/10.1126/science.1209603
112. Herrero, R., García-Palacios, A., Castilla, D., Molinari, G., & Botella, C. (2014). Virtual reality for the induction of positive emotions in the treatment of fibromyalgia: A pilot study over acceptability, satisfaction, and the effect of virtual reality on mood. *Cyberpsychology, Behavior, and Social Networking*, *17*(6), 379–384. https://doi.org/10.1089/cyber.2014.0052
113. Hidalgo-Muñoz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., Tattegrain, H., & Gabaude, C. (2019). Respiration and Heart Rate Modulation Due to Competing Cognitive Tasks While Driving. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/FNHUM.2018.00525
114. Hine, K. A., Porter, L. E., Westera, N. J., Alpert, G. P., & Allen, A. (2018). Exploring Police Use of Force Decision-Making Processes and Impairments Using a Naturalistic Decision-Making Approach. *Criminal Justice and Behavior*, *45*(11), 1782–1801. https://doi.org/10.1177/0093854818789726
115. Hinrichs, R., van Rooij, S. J. H., Michopoulos, V., Schultebraucks, K., Winters, S., Maples-Keller, J., Rothbaum, A. O., Stevens, J. S., Galatzer-Levy, I., Rothbaum, B. O., Ressler, K. J., & Jovanovic, T. (2019). Increased Skin Conductance Response in the Immediate Aftermath of Trauma Predicts PTSD Risk. *Chronic Stress*, *3*. https://doi.org/10.1177/2470547019844441

116. Hinte, F. (1971). Should Desensitization Techniques be a Part of Police Recruit Training? *Group Psychotherapy and Psychodrama*, *24*(3-- 4), 107–110. https://psycnet.apa.org/record/1972-21114-001
117. Hirsch, J. A., & Bishop, B. (1981). Respiratory sinus arrhythmia in humans: How breathing pattern modulates heart rate. *American Journal of Physiology - Heart and Circulatory Physiology*, *10*(4), 620–629. https://doi.org/10.1152/ajpheart.1981.241.4.h620
118. Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: data analysis with estimation graphics. *Nature Methods 2019 16:7*, *16*(7), 565–566. https://doi.org/10.1038/s41592-019-0470-3
119. Holden, R. J., & Karsh, B. T. (2010). The Technology Acceptance Model: Its past and its future in health care. *Journal of Biomedical Informatics*, *43*(1), 159–172. https://doi.org/10.1016/j.jbi.2009.07.002
120. Holzinger, A., Pichler, A., & Maurer, H. (2006). Multi Media e-learning software TRIANGLE case-study: experimental results and lessons learned. *Journal of Universal Science and Technology of Learning*, *1*(1), 61–92.
121. Horner, R. H., Carr, E. G., Halle, J., Mcgee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. In *Exceptional Children* (Vol. 71, Issue 2, pp. 165–179). SAGE PublicationsSage CA: Los Angeles, CA. https://doi.org/10.1177/001440290507100203
122. Huhta, J. M., Di Nota, P. M., Nyman, M., Pietilä, E., & Ropo, E. (2021). Universal police behaviours during critical incidents and their connection to personality: A preliminary study. *Nordic Journal of Studies in Policing*, *8*(1), 1–18. https://doi.org/10.18261/ISSN.2703-7045-2021-01-04/ASSET/0D2CB913-088A-46A3-A744-F23D162E705A/ASSETS/GRAPHIC/NJSP-2020-1-4-1.JPG
123. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55
124. Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal of Experimental Social Psychology*, *46*(1), 208–212. https://doi.org/10.1016/j.jesp.2009.08.015
125. Jamieson, J. P., Mendes, W. B., & Nock, M. K. (2013). Improving Acute Stress Responses: The Power of Reappraisal. In *Current Directions in Psychological Science* (Vol. 22, Issue 1, pp. 51–56). SAGE PublicationsSage CA: Los Angeles, CA. https://doi.org/10.1177/0963721412461500
126. Jerath, R., Edry, J. W., Barnes, V. A., & Jerath, V. (2006). Physiology of long pranayamic breathing: neural respiratory elements may provide a mechanism that explains how slow deep breathing shifts the autonomic nervous system. *Medical Hypotheses*, *67*(3), 566–571. https://doi.org/10.1016/J.MEHY.2006.02.042
127. Jerčić, P., & Sundstedt, V. (2019). Practicing emotion-regulation through biofeedback on the decision-making performance in the context of serious games: A systematic review. *Entertainment Computing*, *29*, 75–86. https://doi.org/10.1016/J.ENTCOM.2019.01.001
128. Jiménez Morgan, S., & Molina Mora, J. A. (2017). Effect of Heart Rate Variability Biofeedback on Sport Performance, a Systematic Review. In *Applied Psychophysiology Biofeedback* (Vol. 42, Issue 3, pp. 235–245). Springer US. https://doi.org/10.1007/s10484-017-9364-2
129. Johnsen, B. H., Hansen, A. L., Murison, R., Eid, J., & Thayer, J. F. (2012). Heart rate variability and cortisol responses during attentional and working memory tasks in naval cadets. *International Maritime Health*, *63*(4), 181–187.
130. Johnson, D. J., Cesario, J., & Pleskac, T. J. (2018). How prior information and police experience impact decisions to shoot. *Journal of Personality and Social Psychology*, *115*(4), 601–623. https://doi.org/10.1037/pspa0000130
131. Johnson, R. R., Stone, B. T., Miranda, C. M., Vila, B., James, L., James, S. M., Rubio, R. F., & Berka, C. (2014). Identifying psychophysiological indices of expert vs. novice performance in deadly force judgment and decision making. *Frontiers in Human Neuroscience*, *8*(JULY). https://doi.org/10.3389/fnhum.2014.00512

132. Kaldewaij, R., Koch, S. B. J., Hashemi, M. M., Zhang, W., Klumpers, F., & Roelofs, K. (2021). Anterior prefrontal brain activity during emotion control predicts resilience to post-traumatic stress symptoms. *Nature Human Behaviour*, *5*(8), 1055–1064. https://doi.org/10.1038/s41562-021-01055-2
133. Kazdin, A. E. (2011). Evidence-Based Treatment Research: Advances, Limitations, and Next Steps. *American Psychologist*, *66*(8), 685–698. https://doi.org/10.1037/A0024975
134. Kendall, M. G. (1938). A New Measure of Rank Correlation. *Biometrika*, *30*(1/2), 81. https://doi.org/10.2307/2332226
135. Keynan, J. N., Cohen, A., Jackont, G., Green, N., Goldway, N., Davidov, A., Meir-Hasson, Y., Raz, G., Intrator, N., Fruchter, E., Ginat, K., Laska, E., Cavazza, M., & Hendler, T. (2019). Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience. In *Nature Human Behaviour* (Vol. 3, Issue 1, pp. 63–73). Nature Publishing Group. https://doi.org/10.1038/s41562-018-0484-3
136. Keynan, J. N., Meir-Hasson, Y., Gilam, G., Cohen, A., Jackont, G., Kinreich, S., Ikar, L., Or-Borichev, A., Etkin, A., Gyurak, A., Klovatch, I., Intrator, N., & Hendler, T. (2016). Limbic Activity Modulation Guided by Functional Magnetic Resonance Imaging–Inspired Electroencephalography Improves Implicit Emotion Regulation. *Biological Psychiatry*, *80*(6), 490–496. https://doi.org/10.1016/j.biopsych.2015.12.024
137. Kini, S. (2013). Please Take My Survey: Compliance with smartphone-based EMA / ESM studies. In *Dartmouth College Undergraduate Theses*. https://digitalcommons.dartmouth.edu/senior_theses/83
138. Klaassen, F. H., Held, L., Figner, B., O'Reilly, J. X., Klumpers, F., de Voogd, L. D., & Roelofs, K. (2021). Defensive freezing and its relation to approach-avoidance decision-making under threat. *Scientific Reports*, *11*(1). https://doi.org/10.1038/S41598-021-90968-Z
139. Kleim, B., & Westphal, M. (2011). Mental Health in First Responders: A Review and Recommendation for Prevention and Intervention Strategies. *Traumatology*, *17*(4), 17–24. https://doi.org/10.1177/1534765611429079
140. Klinger, D. A., & Brunson, R. K. (2009). Police officers' perceptual distortions during lethal force situations: Informing the reasonableness standard. *Criminology & Public Policy*, *8*(1), 117–140. https://doi.org/10.1111/j.1745-9133.2009.00537.x
141. Kluetsch, R. C., Ros, T., Théberge, J., Frewen, P. a, Calhoun, V. D., Schmahl, C., Jetly, R., & Lanius, R. a. (2014). Plastic modulation of PTSD resting-state networks and subjective wellbeing by EEG neurofeedback. *Acta Psychiatrica Scandinavica*, *130*(2), 123–136. https://doi.org/10.1111/acps.12229
142. Koolhaas, J. M., Bartolomucci, A., Buwalda, B., de Boer, S. F., Flügge, G., Korte, S. M., Meerlo, P., Murison, R., Olivier, B., Palanza, P., Richter-Levin, G., Sgoifo, A., Steimer, T., Stiedl, O., van Dijk, G., Wöhr, M., & Fuchs, E. (2011). Stress revisited: a critical evaluation of the stress concept. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1291–1301. https://doi.org/10.1016/J.NEUBIOREV.2011.02.003
143. Kothgassner, O. D., & Felnhofer, A. (2020). Does virtual reality help to cut the Gordian knot between ecological validity and experimental control? *Annals of the International Communication Association*, *0*(0), 210–218. https://doi.org/10.1080/23808985.2020.1792790
144. Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-Case Intervention Research Design Standards. *Remedial and Special Education*, *34*(1), 26–38. https://doi.org/10.1177/0741932512452794
145. Kromenacker, B. W., Sanova, A. A., Marcus, F. I., Allen, J. J. B., & Lane, R. D. (2018). Vagal Mediation of Low-Frequency Heart Rate Variability during Slow Yogic Breathing. *Psychosomatic Medicine*, *80*(6), 581–587. https://doi.org/10.1097/PSY.0000000000000603
146. Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin and Review*, *20*(1), 54–72. https://doi.org/10.3758/s13423-012-0341-8
147. Laborde, S. (2019). Slow-paced breathing and cardiac vagal activity: influence on stress regulation, sleep, and cognitive executive performance. In *Psychology. Normandie Université, English, NNT : 2019NORMC044*.

148. Laborde, S., Allen, M. S., Borges, U., Hosang, T. J., Furley, P., Mosley, E., & Dosseville, F. (2021). The Influence of Slow-Paced Breathing on Executive Function. *Journal of Psychophysiology*. https://doi.org/10.1027/0269-8803/a000279
149. Lagos, L., Vaschillo, E., Vaschillo, B., Lehrer, P., Bates, M., & Pandina, R. (2011). Virtual Reality–Assisted Heart Rate Variability Biofeedback as a Strategy to Improve Golf Performance: A Case Study. *Biofeedback*, *39*(1), 15–20. https://doi.org/10.5298/1081-5937-39.1.11
150. Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. In *Neuropsychological Rehabilitation* (Vol. 24, Issues 3–4, pp. 445–463). Neuropsychol Rehabil. https://doi.org/10.1080/09602011.2013.815636
151. Lay, S., Brace, N., Pike, G., & Pollick, F. (2016). Circling around the uncanny valley: Design principles for research into the relation between human likeness and eeriness. *I-Perception*, *7*(6). https://doi.org/10.1177/2041669516681309
152. Ledger, D., & McCaffrey, D. (2014). *Inside Wearables Part 1: How behavior change unlocks long-term engagement | by Endeavour Partners | Medium*. Medium. https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3
153. Lehrer, P., & Gevirtz, R. (2014). Heart rate variability biofeedback: How and why does it work? *Frontiers in Psychology*, *5*(JUL), 1–9. https://doi.org/10.3389/fpsyg.2014.00756
154. Lehrer, P., Kaur, K., Sharma, A., Shah, K., Huseby, R., Bhavsar, J., & Zhang, Y. (2020). Heart Rate Variability Biofeedback Improves Emotional and Physical Health and Performance: A Systematic Review and Meta Analysis. *Applied Psychophysiology Biofeedback*, *45*(3), 109–129. https://doi.org/10.1007/s10484-020-09466-z
155. Leppma, M., Mnatsakanova, A., Sarkisian, K., Scott, O., Adjeroh, L., Andrew, M. E., Violanti, J. M., & Mccanlies, E. C. (2017). Stressful life events and posttraumatic growth among police officers: A cross-sectional study. *Stress and Health*, *October 2016*, 1–12. https://doi.org/10.1002/smi.2772
156. Levin, J. R., Ferron, J. M., & Gafurov, B. S. (2021). Investigation of Single-Case Multiple-Baseline Randomization Tests of Trend and Variability. *Educational Psychology Review*, *33*(2), 713–737. https://doi.org/10.1007/s10648-020-09549-7
157. Lightner, A. D., Barclay, P., Hagen, E. H., & Hagen, E. H. (2017). *Registered report Radical framing effects in the ultimatum game : the impact of explicit culturally transmitted frames on economic decision-making Author for correspondence :* https://doi.org/10.1098/rsos.170543
158. Limerick, H., Coyle, D., & Moore, J. W. (2014). The experience of agency in human-computer interactions: a review. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/FNHUM.2014.00643
159. LimeSurvey, P. T., & Schmitz, C. (2012). *LimeSurvey: An Open Source survey tool*.
160. Lin, J. H. T. (2017). Fear in virtual reality (VR): Fear elements, coping reactions, immediate and next-day fright responses toward a survival horror zombie virtual reality game. *Computers in Human Behavior*, *72*, 350–361. https://doi.org/10.1016/j.chb.2017.02.057
161. Lingamneni, J. R. (1979). Resistance to change in police organizations: The diffusion paradigm. *Criminal Justice Review*, *4*(2), 17–26. https://doi.org/10.1177/073401687900400204
162. Lipponen, J. A., & Tarvainen, M. P. (2019). A robust algorithm for heart rate variability time series artefact correction using novel beat classification. *Journal of Medical Engineering and Technology*, *43*(3), 173–181. https://doi.org/10.1080/03091902.2019.1640306
163. Lobel, A., Gotsis, M., Reynolds, E., Annetta, M., Engels, R. C. M. E., & Granic, I. (2016). Designing and utilizing biofeedback games for emotion regulation: The case of Nevermind. *Conference on Human Factors in Computing Systems - Proceedings*, *07-12-May-2016*, 1945–1951. https://doi.org/10.1145/2851581.2892521

A

164. Lojowska, M., Gladwin, T. E., Hermans, E. J., & Roelofs, K. (2015). Freezing promotes perception of coarse visual features. *Journal of Experimental Psychology: General*, *144*(6), 1080–1088. https://doi.org/10.1037/xge0000117
165. Love, K., & Singer, M. S. (1988). Self-Efficacy, Psychological Well-Being, Job Satisfaction and Job Involvement: A Comparison of Male and Female Police Officers. *Police Studies*, *11*(2), 98–102. https://heinonline.org/HOL/Page?handle=hein.journals/polic11&id=108&div=&collection=
166. Lucas, G. M., Gratch, J., King, A., & Morency, L. P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043
167. Lüddecke, R., & Felnhofer, A. (2022). Virtual Reality Biofeedback in Health: A Scoping Review. *Applied Psychophysiology Biofeedback*, *47*(1), 1–15. https://doi.org/10.1007/s10484-021-09529-9
168. Macatee, R. J., Albanese, B. J., Schmidt, N. B., & Cougle, J. R. (2017). The Moderating Influence of Heart Rate Variability on Stressor-Elicited Change in Pupillary and Attentional Indices of Emotional Processing: An Eye-Tracking Study. *Biological Psychology*, *123*, 83. https://doi.org/10.1016/J.BIOPSYCHO.2016.11.013
169. Madary, M., & Metzinger, T. K. (2016). Real virtuality: A code of ethical conduct. Recommendations for good scientific practice and the consumers of VR-technology. In *Frontiers Robotics AI* (Vol. 3, Issue FEB). Frontiers Media S.A. https://doi.org/10.3389/frobt.2016.00003
170. Maguen, S., Metzler, T. J., McCaslin, S. E., Inslicht, S. S., Henn-Haase, C., Neylan, T. C., & Marmar, C. R. (2009). Routine Work Environment Stress and PTSD Symptoms in Police Officers. *Journal of Nervous & Mental Disease*, *197*(10), 754–760. https://doi.org/10.1097/NMD.0B013E3181B975F8
171. Maier, S. U., Makwana, A. B., & Hare, T. A. (2015). Acute Stress Impairs Self-Control in Goal-Directed Choice by Altering Multiple Functional Connections within the Brain's Decision Circuits. *Neuron*, *87*(3), 621–631. https://doi.org/10.1016/j.neuron.2015.07.005
172. Manolov, R., Moeyaert, M., & Evans, J. J. (2015). Resources and guidelines for analysing SCED data "Further developments in summarising and meta-analysing single-case data: An illustration with neurobehavioural interventions in acquired brain injury" (Manuscript submitted for publication) Exten. *Neuropsychological Rehabilitation*, *24*(3–4), 1–238. https://doi.org/https://doi.org/10.1080/09602011.2014.903199
173. Marenin, O. (2016). Cheapening Death: Danger, Police Street Culture, and the Use of Deadly Force. *Police Quarterly*, *19*(4), 461–487. https://doi.org/10.1177/1098611116652983
174. Mather, M., & Thayer, J. F. (2018). How heart rate variability affects emotion regulation brain networks. *Current Opinion in Behavioral Sciences*, *19*, 98–104. https://doi.org/10.1016/j.cobeha.2017.12.017
175. McAuley, E. D., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventoiy in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48–58. https://doi.org/10.1080/02701367.1989.10607413
176. McCall, C., Hildebrandt, L. K., Bornemann, B., & Singer, T. (2015). Physiophenomenology in retrospect: Memory reliably reflects physiological arousal during a prior threatening experience. *Consciousness and Cognition*, *38*, 60–70. https://doi.org/10.1016/j.concog.2015.09.011
177. Mccanlies, E. C., Mnatsakanova, A., Andrew, M. E., Burchfiel, C. M., & Violanti, J. M. (2014). Positive psychological factors are associated with lower PTSD symptoms among police officers: Post Hurricane Katrina. *Stress and Health*, *30*(5), 405–415. https://doi.org/10.1002/smi.2615
178. Mccraty, R., & Atkinson, M. (2012). Resilience Training Program Reduces Physiological and Psychological Stress in Police Officers. *Global Advances in Health and Medicine*, *1*(5), 44–66. https://doi.org/10.7453/gahmj.2012.1.5.013

179. McCraty, R., Atkinson, M., Lipsenthal, L., & Arguelles, L. (2009). New Hope for Correctional Officers: An Innovative Program for Reducing Stress and Health Risks. *Applied Psychophysiology and Biofeedback 2009 34:4*, *34*(4), 251–272. https://doi.org/10.1007/S10484-009-9087-0
180. McDevitt-Murphy, M. E., Luciano, M. T., & Zakarian, R. J. (2018). Use of Ecological Momentary Assessment and Intervention in Treatment With Adults. *FOCUS*, *16*(4), 370–375. https://doi.org/10.1176/appi.focus.20180017
181. McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology*, *50*, 215–241. https://doi.org/10.1146/annurev.psych.50.1.215
182. Mendes, W. B., Gray, H. M., Mendoza-Denton, R., Major, B., & Epel, E. S. (2007). Why egalitarianism might be good for your health: Physiological thriving during stressful intergroup encounters. *Psychological Science*, *18*(11), 991–998. https://doi.org/10.1111/j.1467-9280.2007.02014.x
183. Michael, S., Graham, K. S., & Oam, G. M. D. (2017). Cardiac autonomic responses during exercise and post-exercise recovery using heart rate variability and systolic time intervals-a review. *Frontiers in Physiology*, *8*(MAY), 259883. https://doi.org/10.3389/FPHYS.2017.00301/BIBTEX
184. Michela, A., van Peer, J. M., Brammer, J. C., Nies, A., van Rooij, M. M. J. W., Oostenveld, R., Dorrestijn, W., Smit, A. S., Roelofs, K., Klumpers, F., & Granic, I. (2022). Deep-Breathing Biofeedback Trainability in a Virtual-Reality Action Game: A Single-Case Design Study With Police Trainers. *Frontiers in Psychology*, *13*, 29. https://doi.org/10.3389/fpsyg.2022.806163
185. Michela, A., van Rooij, M. M. J. W., Klumpers, F., van Peer, J. M., Roelofs, K., & Granic, I. (2019). Reducing the Noise of Reality. *Psychological Inquiry*, *30*(4), 203–210. https://doi.org/10.1080/1047840X.2019.1693872
186. Miethe, T. D., Venger, O., & Lieberman, J. D. (2019). Police use of force and its video coverage: An experimental study of the impact of media source and content on public perceptions. *Journal of Criminal Justice*, *60*, 35–46. https://doi.org/10.1016/J.JCRIMJUS.2018.10.006
187. Miller, L. (2015). Why cops kill: The psychology of police deadly force encounters. In *Aggression and Violent Behavior* (Vol. 22, pp. 97–111). https://doi.org/10.1016/j.avb.2015.04.007
188. Miller, L. C., Shaikh, S. J., Jeong, D. C., Wang, L., Gillig, T. K., Godoy, C. G., Appleby, P. R., Corsbie-Massay, C. L., Marsella, S., Christensen, J. L., & Read, S. J. (2019). Causal Inference in Generalizable Environments: Systematic Representative Design. *Psychological Inquiry*, *30*(4), 173–202. https://doi.org/10.1080/1047840X.2019.1693866
189. Moore, R. C., Depp, C. A., Wetherell, J. L., & Lenze, E. J. (2016). Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. *Journal of Psychiatric Research*, *75*, 116–123. https://doi.org/10.1016/j.jpsychires.2016.01.011
190. Morgado, P., Sousa, N., & Cerqueira, J. J. (2015). The impact of stress in decision making in the context of uncertainty. *Journal of Neuroscience Research*, *93*(6), 839–847. https://doi.org/10.1002/JNR.23521
191. Mori, M. (1970). Bukimi no tani [The uncanny valley]. In *Energy* (Vol. 7, pp. 33–35).
192. Morris, M. C., Hellman, N., Abelson, J. L., & Rao, U. (2016). Cortisol, Heart Rate, and Blood Pressure as Early Markers of PTSD Risk: A Systematic Review and Meta-Analysis. *Clin Psychol Rev*, *49*, 79–91. https://doi.org/10.1016/j.cpr.2016.09.001.Cortisol
193. Murphy, J. (2014). *Beyond a Split-Second: An Exploratory Study of Police Use of Force and Use of Force Training in Canada.* [Simon Fraser University]. https://summit.sfu.ca/item/14482
194. Nacke, L. E., Kalyn, M., Lough, C., & Mandryk, R. L. (2011). Biofeedback game design: Using direct and indirect physiological control to enhance game interaction. *Conference on Human Factors in Computing Systems - Proceedings*, 103–112. https://doi.org/10.1145/1978942.1978958

A

195. Nicolò, A., Massaroni, C., & Passfield, L. (2017). Respiratory frequency during exercise: The neglected physiological measure. *Frontiers in Physiology*, *8*(DEC), 922. https://doi.org/10.3389/FPHYS.2017.00922/BIBTEX
196. Niermann, H. C. M., Tyborowska, A., Cillessen, A. H. N., van Donkelaar, M. M., Lammertink, F., Gunnar, M. R., Franke, B., Figner, B., & Roelofs, K. (2019). The relation between infant freezing and the development of internalizing symptoms in adolescence: A prospective longitudinal study. *Developmental Science*, *22*(3), e12763. https://doi.org/10.1111/DESC.12763
197. Nieuwenhuys, A., Caljouw, S. R., Leijsen, M. R., Schmeits, B. A. J., & Oudejans, R. R. D. (2009). Quantifying police officers' arrest and self-defence skills: Does performance decrease under pressure? *Ergonomics*, *52*(12), 1460–1468. https://doi.org/10.1080/00140130903287981
198. Nieuwenhuys, A., & Oudejans, R. R. D. (2010). Effects of anxiety on handgun shooting behavior of police officers: A pilot study. *Anxiety, Stress and Coping*, *23*(2), 225–233. https://doi.org/10.1080/10615800902977494
199. Nieuwenhuys, A., & Oudejans, R. R. D. (2011). Training with anxiety: Short-and long-term effects on police officers' shooting behavior under pressure. *Cognitive Processing*, *12*(3), 277–288. https://doi.org/10.1007/s10339-011-0396-x
200. Nieuwenhuys, A., Savelsbergh, G. J. P., & Oudejans, R. R. D. (2015). Persistence of threat-induced errors in police officers' shooting decisions. *Applied Ergonomics*, *48*, 263–272. https://doi.org/10.1016/j.apergo.2014.12.006
201. Norris, C., & Norris, N. (1993). Defining good policing: The instrumental and moral in approaches to good practice and competence. *Policing and Society*, *3*(3), 205–221. https://doi.org/10.1080/10439463.1993.9964669
202. Oostenveld, R. (n.d.). *GitHub - eegsynth/eegsynth: Converting real-time EEG into sounds, music and visual effects*. Retrieved October 10, 2021, from https://github.com/eegsynth/eegsynth
203. Oudejans, R. R. D. (2008). Reality-based practice under pressure improves handgun shooting performance of police officers. *Ergonomics*, *51*(3), 261–273. https://doi.org/10.1080/00140130701577435
204. Oudejans, R. R. D., & Pijpers, J. R. (2009). Training with anxiety has a positive effect on expert perceptual-motor performance under pressure. *Quarterly Journal of Experimental Psychology (2006)*, *62*(8), 1631–1647. https://doi.org/10.1080/17470210802557702
205. Padmanaban, N., Ruban, T., Sitzmann, V., Norcia, A. M., & Wetzstein, G. (2018). Towards a machine-learning approach for sickness prediction in 360° stereoscopic videos. *IEEE Transactions on Visualization and Computer Graphics*, *24*(4), 1594–1603. https://doi.org/10.1109/TVCG.2018.2793560
206. Pallavicini, F., Argenton, L., Toniazzi, N., Aceti, L., & Mantovani, F. (2016). Virtual Reality Applications for Stress Management Training in the Military. *Aerospace Medicine and Human Performance*, *87*(12), 1021–1030. https://doi.org/10.3357/AMHP.4596.2016
207. Pan, X., & Hamilton, A. F. de C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290
208. Paoline, E. A., & Terrill, W. (2007). Police education, experience, and the use of force. *Criminal Justice and Behavior*, *34*(2), 179–196. https://doi.org/10.1177/0093854806290239
209. Parnandi, A., & Gutierrez-Osuna, R. (2017). Physiological Modalities for Relaxation Skill Transfer in Biofeedback Games. *IEEE Journal of Biomedical and Health Informatics*, *21*(2), 361–371. https://doi.org/10.1109/JBHI.2015.2511665
210. Parnandi, A., & Gutierrez-Osuna, R. (2019). Visual biofeedback and game adaptation in relaxation skill transfer. *IEEE Transactions on Affective Computing*, *10*(2), 276–289. https://doi.org/10.1109/TAFFC.2017.2705088
211. Parsons, T. D. (2015). Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00660

212. Pascucci, D., Hervais-Adelman, A., & Plomp, G. (2018). Gating by induced A-Γ asynchrony in selective attention. *Human Brain Mapping*, *39*(10), 3854–3870. https://doi.org/10.1002/hbm.24216
213. Peira, N., Fredrikson, M., & Pourtois, G. (2014). Controlling the emotional heart: Heart rate biofeedback improves cardiac control during emotional reactions. *International Journal of Psychophysiology*, *91*(3), 225–231. https://doi.org/10.1016/j.ijpsycho.2013.12.008
214. Pizzoli, S. F. M., Marzorati, C., Gatti, D., Monzani, D., Mazzocco, K., & Pravettoni, G. (2021). A meta-analysis on heart rate variability biofeedback and depressive symptoms. *Scientific Reports 2021 11:1*, *11*(1), 1–10. https://doi.org/10.1038/s41598-021-86149-7
215. Ponnada, A., Haynes, C., Maniar, D., Manjourides, J., & Intille, S. (2017). Microinteraction Ecological Momentary Assessment Response Rates. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1–16. https://doi.org/10.1145/3130957
216. Porcelli, A. J., & Delgado, M. R. (2017). Stress and decision making: effects on valuation, learning, and risk-taking. *Current Opinion in Behavioral Sciences*, *14*, 33–39. https://doi.org/10.1016/j.cobeha.2016.11.015
217. Prenzler, T., & Lewis, C. (2005). Performance indicators for police oversight agencies. *Australian Journal of Public Administration*, *64*(2), 77–83. https://doi.org/10.1111/j.1467-8500.2005.00443.x
218. Price, C. J., & Hooven, C. (2018). Interoceptive awareness skills for emotion regulation: Theory and approach of mindful awareness in body-oriented therapy (MABT). *Frontiers in Psychology*, *9*(MAY), 798. https://doi.org/10.3389/FPSYG.2018.00798/BIBTEX
219. Pulopulos, M. M., Baeken, C., & De Raedt, R. (2020). Cortisol response to stress: The role of expectancy and anticipatory stress regulation. *Hormones and Behavior*, *117*(October 2019), 104587. https://doi.org/10.1016/j.yhbeh.2019.104587
220. Pulopulos, M. M., Vanderhasselt, M. A., & De Raedt, R. (2018). Association between changes in heart rate variability during the anticipation of a stressful situation and the stress-induced cortisol response. *Psychoneuroendocrinology*, *94*(May), 63–71. https://doi.org/10.1016/j.psyneuen.2018.05.004
221. PyQtGraph. (2020). *PyQtGraph - Scientific Graphics and GUI Library for Python*. https://www.pyqtgraph.org/
222. Rajakaruna, N., Henry, P. J., Cutler, A., & Fairman, G. (2017). Ensuring the validity of police use of force training. *Police Practice and Research*, *18*(5), 507–521. https://doi.org/10.1080/15614263.2016.1268959
223. Ramírez, E., Ortega, A. R., & Reyes Del Paso, G. A. (2015). Anxiety, attention, and decision making: The moderating role of heart rate variability. *International Journal of Psychophysiology*, *98*(3), 490–496. https://doi.org/10.1016/j.ijpsycho.2015.10.007
224. Reichenberger, J., Porsch, S., Wittmann, J., Zimmermann, V., & Shiban, Y. (2017). Social fear conditioning paradigm in virtual reality: Social vs. electrical aversive conditioning. *Frontiers in Psychology*, *8*(NOV), 1–15. https://doi.org/10.3389/fpsyg.2017.01979
225. Rihs, T. A., Michel, C. M., & Thut, G. (2007). Mechanisms of selective inhibition in visual spatial attention are indexed by alpha-band EEG synchronization. *The European Journal of Neuroscience*, *25*(2), 603–610. https://doi.org/10.1111/j.1460-9568.2007.05278.x
226. Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., & Alcañiz, M. (2007). Affective interactions using virtual reality: The link between presence and emotions. *Cyberpsychology and Behavior*, *10*(1), 45–56. https://doi.org/10.1089/cpb.2006.9993
227. Riva, G., Wiederhold, B. K., & Mantovani, F. (2019). Neuroscience of Virtual Reality: From Virtual Exposure to Embodied Medicine. *Cyberpsychology, Behavior, and Social Networking*, *22*(1), 82–96. https://doi.org/10.1089/cyber.2017.29099.gri
228. Riverbank-Computing. (2020). *Riverbank Computing | Introduction*. https://riverbankcomputing.com/software/pyqt/

A

229. Rizzo, A., Cukor, J., Gerardi, M., Alley, S., Reist, C., Roy, M., Rothbaum, B. O., & Difede, J. A. (2015). Virtual Reality Exposure for PTSD Due to Military Combat and Terrorist Attacks. *Journal of Contemporary Psychotherapy*, *45*(4), 255–264. https://doi.org/10.1007/s10879-015-9306-3
230. Rizzo, A., & Shilling, R. (2017). Clinical Virtual Reality tools to advance the prevention, assessment, and treatment of PTSD. *European Journal of Psychotraumatology*, *8*(sup5), 1414560. https://doi.org/10.1080/20008198.2017.1414560
231. Rockstroh, C., Blum, J., & Göritz, A. S. (2019). Virtual reality in the application of heart rate variability biofeedback. *International Journal of Human Computer Studies*, *130*, 209–220. https://doi.org/10.1016/j.ijhcs.2019.06.011
232. Roeckner, A. R., Oliver, K. I., Lebois, L. A. M., van Rooij, S. J. H., & Stevens, J. S. (2021). Neural contributors to trauma resilience: a review of longitudinal neuroimaging studies. *Translational Psychiatry*, *11*(1), 1–17. https://doi.org/10.1038/s41398-021-01633-y
233. Roelofs, K. (2017). Freeze for action: neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718), 20160206. https://doi.org/10.1098/rstb.2016.0206
234. Roelofs, K., & Dayan, P. (2022). Freezing revisited: coordinated autonomic and central optimization of threat coping. *Nature Reviews Neuroscience*, *23*(9), 568–580. https://doi.org/10.1038/s41583-022-00608-2
235. Roelofs, K., Klaassen, F. H., & Dayan, P. (2023). Reply to 'Post-encounter freezing during approach–avoidance conflict: the role of the hippocampus.' *Nature Reviews Neuroscience 2023*, 1–2. https://doi.org/10.1038/s41583-023-00704-x
236. Rogerson, P. (1995). Performance measurement and policing: Police service or law enforcement agency? *Public Money and Management*, *15*(4), 25–30. https://doi.org/10.1080/09540969509387891
237. Ros, T., Baars, B. J., Lanius, R. A., & Vuilleumier, P. (2014). Tuning pathological brain oscillations with neurofeedback: a systems neuroscience framework. *Frontiers in Human Neuroscience*, *8*, 1008. https://doi.org/10.3389/fnhum.2014.01008
238. Ros, T., Enriquez-Geppert, S., Zotev, V., Young, K. D., Wood, G., Whitfield-Gabrieli, S., Wan, F., Vuilleumier, P., Vialatte, F., Van De Ville, D., Todder, D., Surmeli, T., Sulzer, J. S., Strehl, U., Sterman, M. B., Steiner, N. J., Sorger, B., Soekadar, S. R., Sitaram, R., … Thibault, R. T. (2020). Consensus on the reporting and experimental design of clinical and cognitive-behavioural neurofeedback studies (CRED-nf checklist). *Brain*, *143*(6), 1674–1685. https://doi.org/10.1093/brain/awaa009
239. Ros, T., Frewen, P., Théberge, J., Michela, A., Kluetsch, R., Mueller, A., Candrian, G., Jetly, R., Vuilleumier, P., & Lanius, R. A. (2016). Neurofeedback Tunes Scale-Free Dynamics in Spontaneous Brain Activity. *Cerebral Cortex*, *September 2016*, 1–12. https://doi.org/10.1093/cercor/bhw285
240. Ros, T., Michela, A., Bellman, A., Vuadens, P., Saj, A., & Vuilleumier, P. (2017). Increased Alpha-Rhythm Dynamic Range Promotes Recovery from Visuospatial Neglect: A Neurofeedback Study. *Neural Plasticity*, *2017*. https://doi.org/10.1155/2017/7407241
241. Ros, T., Munneke, M. A. M., Ruge, D., Gruzelier, J., & Rothwell, J. C. (2010). Endogenous control of waking brain rhythms induces neuroplasticity in humans. *European Journal of Neuroscience*, *31*(4), 770–778.
242. Rösler, L., & Gamer, M. (2019). Freezing of gaze during action preparation under threat imminence. *Scientific Reports 2019 9:1*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-53683-4
243. Rued, H. A., Hilmert, C. J., Strahm, A. M., & Thomas, L. E. (2019). The influence of stress on attentional bias to threat: An angry face and a noisy crowd. *Psychonomic Bulletin and Review*, *26*(3), 943–950. https://doi.org/10.3758/s13423-018-1538-2
244. Russo, M. A., Santarelli, D. M., & O'Rourke, D. (2017). The physiological effects of slow breathing in the healthy human. In *Breathe* (Vol. 13, Issue 4, pp. 298–309). European Respiratory Society. https://doi.org/10.1183/20734735.009817

245. Sandercock, G. R. H., & Brodie, D. A. (2006). The use of heart rate variability measures to assess autonomic control during exercise. *Scandinavian Journal of Medicine & Science in Sports*, *16*(5), 302–313. https://doi.org/10.1111/J.1600-0838.2006.00556.X
246. Scholten, H., & Granic, I. (2019). Use of the principles of design thinking to address limitations of digital mental health interventions for youth: Viewpoint. *Journal of Medical Internet Research*, *21*(1), 1–14. https://doi.org/10.2196/11528
247. Schoneveld, E. A., Lichtwarck-Aschoff, A., & Granic, I. (2019). What keeps them motivated? Children's views on an applied game for anxiety. *Entertainment Computing*, *29*, 69–74. https://doi.org/10.1016/j.entcom.2018.12.003
248. Schoneveld, E. A., Malmberg, M., Lichtwarck-Aschoff, A., Verheijen, G. P., Engels, R. C. M. E., & Granic, I. (2016). A neurofeedback video game (MindLight) to prevent anxiety in children: A randomized controlled trial. *Computers in Human Behavior*, *63*, 321–333. https://doi.org/10.1016/j.chb.2016.05.005
249. Schubert, D. (2013). Do we always have to strive for "realism"? *Gamasutra*.
250. Schultebraucks, K., Sijbrandij, M., Galatzer-Levy, I., Mouthaan, J., Olff, M., & van Zuiden, M. (2021). Forecasting individual risk for long-term Posttraumatic Stress Disorder in emergency medical settings using biomedical data: A machine learning multicenter cohort study. *Neurobiology of Stress*, *14*(December 2020), 100297. https://doi.org/10.1016/j.ynstr.2021.100297
251. Schwerdtfeger, A. R., Schwarz, G., Pfurtscheller, K., Thayer, J. F., Jarczok, M. N., & Pfurtscheller, G. (2020). Heart rate variability (HRV): From brain death to resonance breathing at 6 breaths per minute. *Clinical Neurophysiology*, *131*(3), 676–693. https://doi.org/10.1016/j.clinph.2019.11.013
252. SciCrunch. (2020). *SciCrunch | Browse RRID Content*. https://scicrunch.org/resources/about/registry/SCR_018732
253. Segovia, K. Y., & Bailenson, J. N. (2009). Virtually True: Children's Acquisition of False Memories in Virtual Reality. *Media Psychology*, *12*(4), 371–393. https://doi.org/10.1080/15213260903287267
254. Seifert, L., Papet, V., Strafford, B. W., Coughlan, E. K., & Davids, K. (2018). Skil transfer, expertise and talent development: An ecological dynamics perspective. *Movement and Sports Sciences - Science et Motricite*, *2018*(102), 39–49. https://doi.org/10.1051/sm/2019010
255. Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *0*, 258. https://doi.org/10.3389/FPUBH.2017.00258
256. Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in Psychology*, *5*, 1040. https://doi.org/10.3389/fpsyg.2014.01040
257. Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Weiskopf, N., Blefari, M. L., Rana, M., Oblak, E., Birbaumer, N., & Sulzer, J. (2017). Closed-loop brain training: The science of neurofeedback. In *Nature Reviews Neuroscience* (Vol. 18, Issue 2, pp. 86–100). Nature Publishing Group. https://doi.org/10.1038/nrn.2016.164
258. Skolnick, J. H., & McCoy, C. (1984). Police Accountability and the Media. *Law & Social Inquiry*, *9*(3), 521–557. https://doi.org/10.1111/j.1747-4469.1984.tb00019.x
259. Slater, M. (2018). Immersion and the illusion of presence in virtual reality. In *British Journal of Psychology* (Vol. 109, Issue 3, pp. 431–433). John Wiley and Sons Ltd. https://doi.org/10.1111/bjop.12305
260. Smith, J. (2022). *Innovation Resistance? Understanding Officer Attitudes Toward Police Innovation*. https://stars.library.ucf.edu/etd2020/1291/
261. Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, *17*(4), 510–550. https://doi.org/10.1037/a0029312
262. Smith, P. L., & Little, D. R. (2018). In defense of the small-N design. *Psychonomic Bulletin and Review*, *25*, 2083–2101. https://doi.org/10.3758/s13423-018-1451-8
263. Srinivasan, K., Sucharita, S., & Vaz, M. (2002). Effect of standing on short term heart rate variability across age. *Clinical Physiology and Functional Imaging*, *22*(6), 404–408. https://doi.org/10.1046/j.1475-097X.2002.00450.x

264. Staller, M. S., Christiansen, P., Zaiser, B., Körner, S., & Cole, J. C. (2018). Do they Aggress Earlier? Investigating the Effects of Ego Depletion on Police Officers' Use of Force Behavior. *Journal of Police and Criminal Psychology*, *33*(4), 332–344. https://doi.org/10.1007/s11896-017-9249-6
265. Staller, M. S., & Körner, S. (2019). Commentary: Complex Motor Learning and Police Training: Applied, Cognitive, and Clinical Perspectives. *Frontiers in Psychology*, *10*, 2444. https://doi.org/10.3389/fpsyg.2019.02444
266. Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *3*(I), 37–149.
267. Stanney, K. M., Hale, K. S., Nahmens, I., & Kennedy, R. S. (2003). What to Expect from Immersive Virtual Environment Exposure: Influences of Gender, Body Mass Index, and Past Experience. *Human Factors*, *45*(3), 504–520. https://doi.org/10.1518/hfes.45.3.504.27254
268. Stanney, K. M., & Hash, P. (1998). Locus of user-initiated control in virtual environments: Influences on cybersickness. *Presence: Teleoperators and Virtual Environments*, *7*(5), 447–459. https://doi.org/10.1162/105474698565848
269. Stetz, M. C., Thomas, M. L., Russo, M. B., Stetz, T. A., Wildzunas, R. M., McDonald, J. J., Wiederhold, B. K., & Romano, J. A. (2007). Stress, Mental Health, and Cognition: A Brief Review of Relationships and Countermeasures. *Aviation, Space, and Environmental Medicine*, *78*(5,suppl), B252–B260.
270. Stone, J. A., Strafford, B. W., North, J. S., Toner, C., & Davids, K. (2018). Effectiveness and efficiency of virtual reality designs to enhance athlete development: An ecological dynamics perspective. *Movement and Sports Sciences - Science et Motricite*, *2018*(102), 51–60. https://doi.org/10.1051/sm/2018031
271. Sun, Z., Cao, N., & Ma, X. (2017). Attention, comprehension, execution: Effects of different designs of biofeedback display. *Conference on Human Factors in Computing Systems - Proceedings*, *Part F127655*, 2132–2139. https://doi.org/10.1145/3027063.3053082
272. Tajadura-Jiménez, A., Banakou, D., Bianchi-Berthouze, N., & Slater, M. (2017). Embodiment in a Child-Like Talking Virtual Body Influences Object Size Perception, Self-Identification, and Subsequent Real Speaking. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-09497-3
273. Tarlow, K. R. (2017a). An Improved Rank Correlation Effect Size Statistic for Single-Case Designs: Baseline Corrected Tau. In *Behavior Modification* (Vol. 41, Issue 4). https://doi.org/10.1177/0145445516676750
274. Tarlow, K. R. (2017b). *Baseline Corrected Tau for single-case research (R code). Retrieved from http://ktarlow.com/stats/*. http://ktarlow.com/stats/r/bctau.txt
275. Tarlow, K. R. (2018). *R functions for meta-analysis (R code). Retrieved from http://ktarlow.com/stats/*. http://ktarlow.com/stats/r/meta.txt
276. Taylor, P. L. (2020). Dispatch Priming and the Police Decision to Use Deadly Force. *Police Quarterly*, *23*(3), 311–332. https://doi.org/10.1177/1098611119896653
277. Thayer, J. F., Åhs, F., Fredrikson, M., Sollers, J. J., & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience and Biobehavioral Reviews*, *36*(2), 747–756. https://doi.org/10.1016/j.neubiorev.2011.11.009
278. Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function, and cognitive performance: The neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine*, *37*(2), 141–153. https://doi.org/10.1007/s12160-009-9101-z
279. Thibault, R. T., & Raz, A. (2017). The psychology of neurofeedback: Clinical intervention even if applied placebo. *American Psychologist*, *72*(7), 679–688. https://doi.org/10.1037/amp0000118
280. Tolin, D. F., Davies, C. D., Moskow, D. M., & Hofmann, S. G. (2020). Biofeedback and neurofeedback for anxiety disorders: A quantitative and qualitative systematic review. *Advances in Experimental Medicine and Biology*, *1191*, 265–289. https://doi.org/10.1007/978-981-32-9705-0_16

281. Trigueros, R., Aguilar-Parra, J. M., Cangas-Díaz, A. J., Fernández-Batanero, J. M., Mañas, M. A., Arias, V. B., & López-Liria, R. (2019). The influence of the trainer on the motivation and resilience of sportspeople: A study from the perspective of self-determination theory. *PLOS ONE*, *14*(8), e0221461. https://doi.org/10.1371/JOURNAL.PONE.0221461
282. van der Meulen, E., Bosmans, M. W. G., Lens, K. M. E., Lahlah, E., & van der Velden, P. G. (2018). Effects of Mental Strength Training for Police Officers: a Three-Wave Quasi-experimental Study. *Journal of Police and Criminal Psychology*, *33*(4), 385–397. https://doi.org/10.1007/s11896-017-9247-8
283. van der Ploeg, H. M. (1984). The development and validation of the Dutch form of the Test Anxiety Inventory. *Applied Psychology*, *33*(2), 243–254. https://doi.org/10.1111/J.1464-0597.1984.TB01432.X
284. Van Der Velden, P. G., Kleber, R. J., Grievink, L., & Yzermans, J. C. (2010). Confrontations with aggression and mental health problems in police officers: The role of organizational stressors, life-events and previous mental health problems. *Psychological Trauma: Theory, Research, Practice, and Policy*, *2*(2), 135–144. https://doi.org/10.1037/a0019158
285. van der Velden, P., Lens, K., Hoffenkamp, H., Bosmans, M., & van der Meulen, E. (2014). Samenvatting Rapport Evaluatie training Mentale Kracht Samenvatting. *Extern Wetenschappelijke Betrekkingen*, 1–10.
286. Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37
287. van der Zwan, J. E., de Vente, W., Huizink, A. C., Bögels, S. M., & de Bruin, E. I. (2015). Physical Activity, Mindfulness Meditation, or Heart Rate Variability Biofeedback for Stress Reduction: A Randomized Controlled Trial. *Applied Psychophysiology Biofeedback*, *40*(4), 257–268. https://doi.org/10.1007/s10484-015-9293-x
288. Van Rooij, M., Lobel, A., Harris, O., Smit, N., & Granic, I. (2016). DEEP: A biofeedback virtual reality game for children at-risk for anxiety. *Conference on Human Factors in Computing Systems - Proceedings*, *07-12-May-2016*, 1989–1997. https://doi.org/10.1145/2851581.2892452
289. Van Rossum, G. Van, & Drake, F. (2009). *Python reference manual* (Scotts Valley, Ed.). CreateSpace.
290. Velden, P. van der. (2014). *Evaluatie, training, mentale kracht : een plan-, proces-, en een effectevaluatie van de training mentale kracht voor politiemedewerkers*.
291. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods 2020 17:3*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2
292. Vogel, S., Klumpers, F., Kroes, M. C. W., Oplaat, K. T., Krugers, H. J., Oitzl, M. S., Joëls, M., & Fernández, G. (2015). A Stress-Induced Shift from Trace to Delay Conditioning Depends on the Mineralocorticoid Receptor. *Biological Psychiatry*, *78*(12), 830–839. https://doi.org/10.1016/j.biopsych.2015.02.014
293. Vourvopoulos, A., Pardo, O. M., Lefebvre, S., Neureither, M., Saldana, D., Jahng, E., & Liew, S.-L. (2019). Effects of a Brain-Computer Interface With Virtual Reality (VR) Neurofeedback: A Pilot Study in Chronic Stroke Patients. *Frontiers in Human Neuroscience*, *13*(June), 1–17. https://doi.org/10.3389/fnhum.2019.00210
294. Walker, G. H., Stanton, N. A., Kazi, T. A., Salmon, P. M., & Jenkins, D. P. (2009). Does advanced driver training improve situational awareness? *Applied Ergonomics*, *40*(4), 678–687. https://doi.org/10.1016/J.APERGO.2008.06.002
295. Walsh, W. F. (1983). Leadership - A Police Perspective . *Police Chief*, *50*(11), 26–29. https://www.ojp.gov/ncjrs/virtual-library/abstracts/leadership-police-perspective
296. Wang, J. R., & Hsieh, S. (2013). Neurofeedback training improves attention and working memory performance. *Clinical Neurophysiology*, *124*(12), 2406–2420. https://doi.org/10.1016/j.clinph.2013.05.020

297. Wang, S., & Rochat, P. (2017). Human Perception of Animacy in Light of the Uncanny Valley Phenomenon. *Perception*, *46*(12), 1386–1411. https://doi.org/10.1177/0301006617722742
298. Waskom, M., Botvinnik, O., Gelbart, M., Ostblom, J., Hobson, P., & Lukauskas, S. (2020). *mwaskom/seaborn: v0.11.0 (September 2020)*. https://doi.org/10.5281/ZENODO.7495530
299. Weerdmeester, J. (2021). Take a DEEP Breath: Exploring the Potential of Game-based Biofeedback Interventions for Anxiety Regulation. In *Doctoral dissertation Radboud university Nijmegen*.
300. Weerdmeester, J., van Rooij, M., Harris, O., Smit, N., Engels, R. C. M. E., & Granic, I. (2017). Exploring the role of self-efficacy in biofeedback video games. *CHI PLAY 2017 Extended Abstracts - Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*, 453–461. https://doi.org/10.1145/3130859.3131299
301. Weerdmeester, J., Van Rooij, M. M. J. M., Engels, R. C. M. E. C., & Granic, I. (2020). An integrative model for the effectiveness of biofeedback interventions for anxiety regulation: Viewpoint. *Journal of Medical Internet Research*, *22*(7), e14958. https://doi.org/10.2196/14958
302. Weibel, R. P., Kerr, J. I., Naegelin, M., Ferrario, A., Schinazi, V. R., La Marca, R., Hoelscher, C., Nater, U. M., & von Wangenheim, F. (2023). Virtual reality-supported biofeedback for stress management: Beneficial effects on heart rate variability and user experience. *Computers in Human Behavior*, *141*(September 2022), 107607. https://doi.org/10.1016/j.chb.2022.107607
303. West, C., Fekedulegn, D., Andrew, M., Burchfiel, C. M., Harlow, S., Bingham, C. R., McCullagh, M., Park, S. K., & Violanti, J. (2017). On-Duty Nonfatal Injury that Lead to Work Absences Among Police Officers and Level of Perceived Stress. *Journal of Occupational and Environmental Medicine*, *XX*(X), 1. https://doi.org/10.1097/JOM.0000000000001137
304. White, M. D. (2016). Transactional Encounters, Crisis-Driven Reform, and the Potential for a National Police Deadly Force Database. *Criminology and Public Policy*, *15*(1), 223–235. https://doi.org/10.1111/1745-9133.12180
305. Wickens, C. D. (2010). Multiple resources and performance prediction. *Https://Doi.Org/10.1080/14639220210123806*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806
306. Williams, S. N., & Clarke, A. V. (2019). How the Desensitization of Police Violence, Stereotyped Language, and Racial Bias Impact Black Communities. *Psychology and Cognitive Sciences – Open Journal*, *5*(2), 62–67. https://doi.org/10.17140/pcsoj-5-151
307. Willis, J. J., & Mastrofski, S. D. (2016). Improving policing by integrating craft and science: what can patrol officers teach us about good police work? *Https://Doi.Org/10.1080/10439463.2015.1135921*, *28*(1), 27–44. https://doi.org/10.1080/10439463.2015.1135921
308. Wilson, C. J., & Soranzo, A. (2015). The Use of Virtual Reality in Psychology: A Case Study in Visual Perception. *Computational and Mathematical Methods in Medicine*, *2015*. https://doi.org/10.1155/2015/151702
309. Wirz, L., Bogdanov, M., & Schwabe, L. (2018). Habits under stress: mechanistic insights across different types of learning. *Current Opinion in Behavioral Sciences*, *20*, 9–16. https://doi.org/10.1016/j.cobeha.2017.08.009
310. Wixon, D. (2003). Evaluating usability methods. *Interactions*, *10*(4), 28. https://doi.org/10.1145/838830.838870
311. WMA. (2018). WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects – WMA – The World Medical Association. In *WMA Declaration of Helsinki – Ethical principles for medical research involving human subjects*.
312. Worrall, J. L. (1999). Public perceptions of police efficacy and image: The "fuzziness" of support for the police. *American Journal of Criminal Justice*, *24*(1), 47–66. https://doi.org/10.1007/bf02887617

313. Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., Li, C., Li, W., Huang, H., Akdere, M., Mousas, C., & Yu, L. F. (2021). A Review on Virtual Reality Skill Training Applications. In *Frontiers in Virtual Reality* (Vol. 2, p. 49). Frontiers Media S.A. https://doi.org/10.3389/frvir.2021.645153
314. Yasuma, F., & Hayano, J. I. (2004). Respiratory Sinus Arrhythmia: Why Does the Heartbeat Synchronize with Respiratory Rhythm? *Chest*, *125*(2), 683–690. https://doi.org/10.1378/chest.125.2.683
315. Young, M. E., Sutherland, S. C., & McCoy, A. W. (2018). Optimal go/no-go ratios to maximize false alarms. *Behavior Research Methods*, *50*(3), 1020–1029. https://doi.org/10.3758/s13428-017-0923-5
316. Yu, B., Funk, M., Hu, J., Wang, Q., & Feijs, L. (2018). Biofeedback for everyday stress management: A systematic review. *Frontiers in ICT*, *5*(SEP). https://doi.org/10.3389/fict.2018.00023
317. Yu, R. (2016). Stress potentiates decision biases: A stress induced deliberation-to-intuition (SIDI) model. *Neurobiology of Stress*, *3*, 83. https://doi.org/10.1016/J.YNSTR.2015.12.006
318. Zechner, O., Kleygrewe, L., Jaspaert, E., Schrom-feiertag, H., Hutter, R. I. V., & Tscheligi, M. (2023). Enhancing Operational Police Training in High Stress Situations with Virtual Reality : Experiences , Tools and Guidelines. *Multimodal Technologies and Interaction*, *7*(14).
319. Zhang, W., Kaldewaij, R., Hashemi, M. M., Koch, S. B. J., Smit, A., van Ast, V. A., Beckmann, C. F., Klumpers, F., & Roelofs, K. (2022). Acute-stress-induced change in salience network coupling prospectively predicts post-trauma symptom development. *Translational Psychiatry*, *12*(1), 1–8. https://doi.org/10.1038/s41398-022-01798-0
320. Zhao, K., Wulder, M. A., Hu, T., Bright, R., Wu, Q., Qin, H., Li, Y., Toman, E., Mallick, B., Zhang, X., & Brown, M. (2019). Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sensing of Environment*, *232*, 111181.
321. Złotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00883

PUBLICATIONS LIST

Ros, T., Frewen, P., Théberge, J., **Michela, A.**, Kluetsch, R., Mueller, A., Candrian, G., Jetly, R., Vuilleumier, P., & Lanius, R. A. (2016). Neurofeedback Tunes Scale-Free Dynamics in Spontaneous Brain Activity Neurofeedback Tunes Scale-Free Dynamics in Spontaneous Brain Activity. Cerebral Cortex, September 2016, 1–12. https://doi.org/10.1093/cercor/bhw285

Saj, A., Ros, T., **Michela, A.**, & Vuilleumier, P. (2017). Effect of a single early EEG neurofeedback training on remediation of spatial neglect in the acute phase. Annals of Physical and Rehabilitation Medicine. https://doi.org/10.1016/j.rehab.2017.11.001

Ros, T., **Michela, A.**, Bellman, A., Vuadens, P., Saj, A., & Vuilleumier, P. (2017). Increased Alpha-Rhythm Dynamic Range Promotes Recovery from Visuospatial Neglect: A Neurofeedback Study. Neural Plasticity, 2017. https://doi.org/10.1155/2017/7407241

Michela, A., van Rooij, M. M. J. W., Klumpers, F., van Peer, J. M. J. M., Roelofs, K., & Granic, I. (2019). Reducing the Noise of Reality. Psychological Inquiry, 30(4), 203–210. https://doi.org/10.1080/1047840X.2019.1693872

Ros, T., **Michela, A.**, Mayer, A., Bellmann, A., Vuadens, P., Zermatten, V., Saj, A., Vuilleumier, P., Meyer, A., Bellman, A., Vuadens, P., Zermatten, V., Saj, A., & Vuilleumier, P. (2019). Disruption of large-scale electrophysiological networks in stroke patients with visuospatial neglect. *Network Neuroscience*, *6*(1), 602318. https://doi.org/10.1101/602318

Ros, T., Kwiek, J., Andriot, T., **Michela, A.**, Vuilleumier, P., Garibotto, V., & Ginovart, N. (2021). PET Imaging of Dopamine Neurotransmission During EEG Neurofeedback. Frontiers in Physiology, 11, 1749. https://doi.org/10.3389/FPHYS.2020.590503/BIBTEX

Brammer, J. C., van Peer, J. M., Michela, A., van Rooij, M. M. J. W., Oostenveld, R., Klumpers, F., Dorrestijn, W., Granic, I., & Roelofs, K. (2021). Breathing Biofeedback for Police Officers in a Stressful Virtual Environment: Challenges and Opportunities. Frontiers in Psychology, 12, 586553. https://doi.org/10.3389/fpsyg.2021.586553

Michela, A., van Peer, J. M., Brammer, J. C., Nies, A., van Rooij, M. M. J. W., Oostenveld, R., Dorrestijn, W., Smit, A. S., Roelofs, K., Klumpers, F., & Granic, I. (2022). Deep-Breathing Biofeedback Trainability in a Virtual-Reality Action Game: A Single-Case Design Study With Police Trainers. Frontiers in Psychology, 13, 29. https://doi.org/10.3389/fpsyg.2022.806163

Michela, A., van Peer, J. M., Oostenveld, R., Dorrestijn, W., Smit, A. S., Granic, I., Roelofs, K. and Klumpers, F. *(in prep).* Preparing The Heart for Duty: Virtual Reality Biofeedback in an Arousing Action Game Improves in-action Voluntary Heart Rate Variability Control in Experienced Police

A

RESEARCH DATA MANAGEMENT STATEMENT

This dissertation followed the laws and ethical guidelines applicable. To further increase the usefulness of the data collected, research data management was conducted by following the FAIR principles (Findable, Accessible, Interoperable, Reusable).

All the research presented in this thesis was conducted in accordance with the principles of the Declaration of Helsinki and were evaluated by the Ethical Committee of the Faculty of Social Sciences (ECSS). The Dean of the Faculty then formally approved the presented research with the positive advice of the ECSS. This work was funded by the Netherlands Organization of Scientific Research (NOW), specifically the Creative Industry grant # 055.16.139 2795 awarded to Karin Roelofs, Isabela Granic, and Floris Klumpers. Additional funds were provided by CbusineZ after approval by the PolitieZorgPolis (PZP).

The studies described in chapters 3 and 4 are approved in the ECSS approval number ECSW-2020-112. The large Randomized Controlled Trial of chapter 5 was approved by two amendments on the previous ethical approval, number ECSW-2021-017 and ECSW-2021-082.

Data availability

The full datasets and research documentation of the different studies presented in this thesis are stored on the project network drive of the Radboud University (FSW-BSI-DP-PhD_Abele_Michela.lnk).

The datasets presented in chapter 3 and 4 are not readily available online due to privacy concerns, as the small number of participants facilitates their identification. As the dataset contains performance indexes of police officers, we preferred to make the data available to motivated requests only. Requests to access those datasets should be directed to the Abele Michela (abele.michela@gmail.com).

The dataset presented in chapter 5 has been released online in the Radboud Data Repository (RDR) under the following link: https://doi.org/10.34973/jh1r-5q76. Datasets shared as a Data Sharing Collection of the Radboud University remain available for at least 10 years after the termination of the studies.

The informed consent forms signed by the participants were collected on paper according to the ECSS procedure and are stored in the central archive of the Radboud University for 10 years after the termination of the studies.

Privacy

To ensure the anonymity of the participants taking part in the experiments described in the chapters 3 to 5, participants were assigned an individual subject code. The pseudonymization key allowing to link the individual subject code with the participants' identity was stored separately from the relative data, and destroyed one month after the project finalization by the manager of the dataset (Abele Michela).

ENGLISH SUMMARY

The present thesis investigated how a VR game-based biofeedback training could be used to help police officers train control over their psychophysiological stress response, and ultimately improve their decision-making performance and resilience when facing stressful situations. The thesis starts in Chapter 1 & 2 by laying the theoretical foundations of this new training by highlighting the boundaries and limitations of VR and biofeedback when used for scientific and training applications. Indeed, as most novel technologies, VR biofeedback comes with many promises and expectations. However, it is often hard to fulfill those promises when one considers the complexity of developing realistic VR scenarios as well the challenges of interpreting data when subjects can freely move and explore. Hence, the theoretical foundation laid out in this thesis advocates the use of simpler, game-like, environments to effectively train police officers on specific tasks as stress management. The thesis then continues in chapter 3 by highlighting the technical challenges and solutions related to creating effective physiological markers for in-action biofeedback targeted at increasing calm during stress, for example when dealing with participant movement and players' sense of controllability.

The first proof-of-concept study described in chapter 4 investigated the effect of training a small group of 9 police trainers for multiple days with our VR game. The aim of this study was to carefully document the evolution of in-game breathing training effects over a 1-month period, comprising 10 training sessions. The results demonstrated the feasibility and promise of biofeedback training in an active context. Many features of the experimental schedule and discoveries made in this explorative study were used to design the following study, a large-scale multi-day randomized control trial, that included 109 police trainers from police centers throughout the Netherlands. In this last study described in chapter 5, we demonstrated the efficacy of training Heart Rate Variability (HRV) in a condensed and fast-paced schedule, as well as the transfer of the trained skill to a new police-relevant task outside VR. This last study also demonstrated physiological awareness of the participants increased throughout the training and a high sense of engagement and willingness to use our training.

This thesis demonstrates the effectiveness of our VR environment to train HRV control skills in a way that transfers to real life situations outside VR. It additionally proposes a study design able to investigate causal relationships between different elements of a training intervention. Finally, the present study provides ground for

further research, as well as for the implementation of VR biofeedback trainings in the curricula of police students and other first responders.

NEDERLANDSE SAMENVATTING (DUTCH SUMMARY)

Het huidige proefschrift onderzocht hoe een op VR-games gebaseerde biofeedbacktraining kan worden gebruikt om politieagenten te helpen de controle over hun psychofysiologische stressreactie te trainen, en uiteindelijk hun besluitvormingsprestaties en veerkracht te verbeteren wanneer ze met stressvolle situaties worden geconfronteerd. Het proefschrift begint in hoofdstuk 1 en 2 met het leggen van de theoretische basis voor deze nieuwe training door de grenzen en beperkingen van VR en biofeedback te benadrukken wanneer gebruikt voor wetenschappelijke en trainingstoepassingen. Zoals de meeste nieuwe technologieën brengt VR biofeedback inderdaad veel beloftes en verwachtingen met zich mee. Het is echter vaak moeilijk om deze beloften waar te maken als men kijkt naar de complexiteit van het ontwikkelen van realistische VR-scenario's en de uitdagingen van het interpreteren van gegevens wanneer proefpersonen zich vrij kunnen bewegen en verkennen. Daarom pleit de theoretische basis die in dit proefschrift wordt uiteengezet voor het gebruik van eenvoudigere, spelachtige omgevingen om politieagenten effectief te trainen in specifieke taken, zoals stressmanagement. Het proefschrift gaat vervolgens verder in hoofdstuk 3 door de nadruk te leggen op de technische uitdagingen en oplossingen die verband houden met het creëren van effectieve fysiologische markers voor in-action biofeedback gericht op het vergroten van de kalmte tijdens stress, bijvoorbeeld bij het omgaan met de bewegingen van deelnemers en het gevoel van beheersbaarheid van spelers.

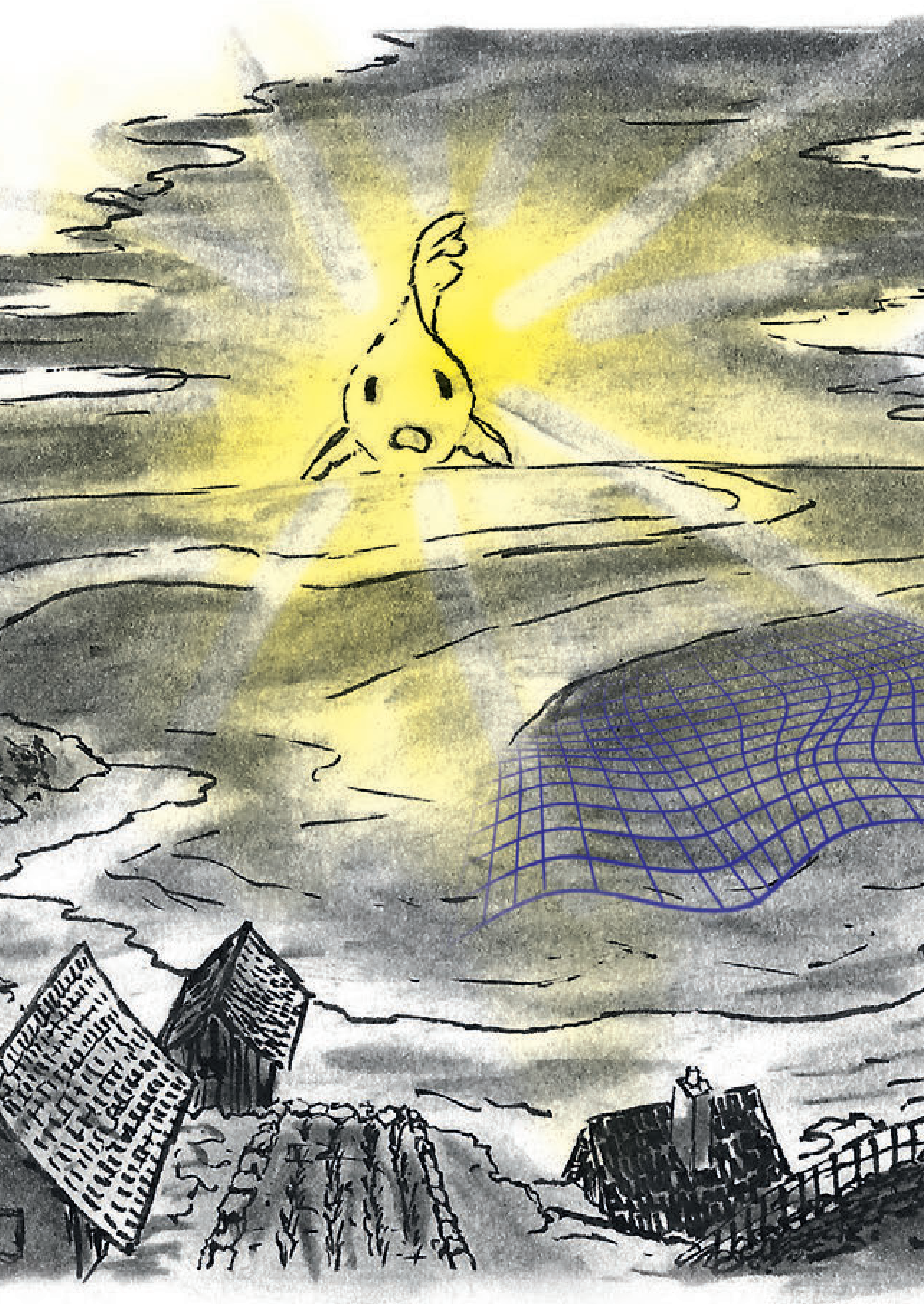
De eerste proof-of-concept studie beschreven in hoofdstuk 4 onderzocht het effect van het meerdere dagen trainen van een kleine groep van 9 politietrainers met onze VR-game. Het doel van deze studie was om de evolutie van de effecten van ademhalingstraining tijdens een periode van 1 maand zorgvuldig te documenteren, bestaande uit 10 trainingssessies. De resultaten toonden de haalbaarheid en belofte aan van biofeedbacktraining in een actieve context. Veel kenmerken van het experimentele schema en de ontdekkingen die in dit verkennende onderzoek zijn gedaan, zijn gebruikt om het volgende onderzoek op te zetten, een grootschalig meerdaags gerandomiseerd controleonderzoek, waaraan 109 politietrainers van politiecentra door heel Nederland deelnamen. In dit laatste onderzoek, beschreven in hoofdstuk 5, hebben we de effectiviteit aangetoond van het trainen van hartslagvariabiliteit (HRV) in een beknopt en snel schema, evenals de overdracht van de getrainde vaardigheid naar een nieuwe politie-relevante taak buiten VR. Dit laatste onderzoek toonde ook aan dat het fysiologische bewustzijn van de deelnemers tijdens de training toenam en dat er een hoog gevoel van betrokkenheid en bereidheid was om onze training te gebruiken.

Dit proefschrift demonstreert de effectiviteit van onze VR-omgeving om HRV-controlevaardigheden te trainen op een manier die wordt overgedragen naar situaties in het echte leven buiten VR. Het stelt bovendien een onderzoeksontwerp voor dat causale relaties tussen verschillende elementen van een trainingsinterventie kan onderzoeken. Ten slotte biedt de huidige studie aanleiding voor verder onderzoek, evenals voor de implementatie van VR-biofeedbacktraining in de curricula van politiestudenten en andere eerstehulpverleners.

CURRICULUM VITAE

Abele Michela was born on May 24th 1988 in Sierre (Switzerland). He obtained a Bachelor of Sciences in Life Sciences and Technologies from École Polytechnique Fédérale de Lausanne (EPFL) in 2012, and a Master in Neurosciences from the University of Geneva in 2015. During his master, he designed several studies to investigate the use of neurofeedback training to enhance attention in healthy participants and stroke survivors. Aside of his master, he worked for several years as an actor for the Police Academy of Savatan (Switzerland) and several other Swiss police institutions. His desire to improve police training by applying his scientific knowledge of neurofeedback was made possible in 2017 by the opening of a PhD position at the Radboud University. There, he received a lot of insights and inspiration from the fundamental knowledge of physiological stress and decision making from the EPAN group (Experimental Psychopathology and Affective Neurosciences lab) and the applied knowledge of game-based interventions from the GemH Lab (Gaming for Emotional Mental Health lab). In order to design a state-of-the-art training for police officers, Abele and his colleagues additionally worked in synergy with game designers and police researchers. Their work resulted in DUST (Decision Under Stress Training), a Virtual-Reality game-based biofeedback application. It was designed to training police officers to improve their capacity for decision making in stressful contexts by down-regulation of their stress-induced physiological arousal.

Thanks to the strong theoretical anchorage and the essential design values of this project, DUST has been received very positively by the police trainers that participated in the data collection that took place amidst the Covid-19 pandemic. Abele's role in creating strong bonds between academic research and applied police realities is reflected by the continuation of the implementation effort of DUST in the police academy, where a new version is being developed to incorporate DUST training in the police curricula.



ACKNOWLEDGEMENTS (DANKWOORD)

ACKNOWLEDGEMENTS (DANKWOORD)

1. Funders

Acknowledgements are usually more personal in PhD theses. I however felt that it was important to thank the **Behavioural Science Institute**, then directed by **Toon Cillessen**, for allowing me to move to Switzerland during the pandemic, while remaining a fully enrolled PhD student. The institute further supported me financially with a 6-months paid contract extension to compensate for delays induced by the COVID-19 pandemic. The institute supported all the extra financial costs related to this extension, as well as the costs related to the special need I had of living close to my family. I think that this generosity has largely contributed to preserve my mental health throughout the difficult times that have surrounded the last years of my thesis.

On a slightly more ironical note, I would also like to thank my own savings as well as the right of unemployment I could benefit from. Indeed, as many PhD students worldwide, the last months of my PhD have been worked without financial compensation as my PhD contract ended in March 2023. I feel it is my duty to give visibility to this phenomenon, extremely common in Academia, that would however sound eerie in any other work context.

2. The Colleagues

2.1 The DUST Team

Isabela, you have been essential to my survival in the Netherlands. Your scientific guidance, deeply rooted in values of excellence, curiosity and (above all) care are among of the main reasons for this project's success. Your vision of the world, and how it translated in our work, are things that I'll take along, whatever the next journey is about. But I also can't thank you enough for your hospitality and empathy in some of the most difficult times I faced. I hope I'll be able to pay it forward one day, as paying back is probably impossible. Thank you for all the wind you put in my sails!

Karin, your "to the point" attitude has kept the project on track in many occasions. While you deserve many thanks for all the help and guidance, I would like to thank you even more for believing in this project so much as to taking it further, with new PhDs continuing what we all built together. Our agreement on the need to make a difference "out there" with this project has been a foundation for the whole project. I'm very thankful that your lab is keeping expanding its activities in that direction.

Floris, when you took over supervision half way through my PhD, we knew it wouldn't be easy times ahead, as all the fun designing part was over. Yet, with your unshakable positivity you managed to always lighten the mood. Working with you has made most of the rough times a PhD student experiences surprisingly smooth. Thank you for your guidance!

Jacobien, the project would have had a hard time keeping internal consistency without you. Just as its important to set a course for a project, making sure the boat holds until reaching destination is essential. I've seen you pointing out problems and spotting inconsistencies in our framework before anyone else so many times! Thank you for your patient and careful guidance!

Wendy, some of the most important insights of this project came from you. You've kept us safe from drifting too much into theories and abstractions by always bringing back the policing realities we were dealing with, both from a scientific and human perspective. You've been THE person I always consulted when I needed to find meaning in my academic meanderings. It fueled me throughout these years. This project would have had no way to be so deeply rooted in the police realities and institution if it wasn't for you.

Ken, you gigantic piece of… genius! DUST wouldn't have been able to go anywhere without you, and your patient capacity for translating scientific gibberish into actionable ideas. I miss the 6th floor extravaganza that you were orchestrating daily, the crazy parties at your old place, and in any dubiously commendable place in Nijmegen. Thinking of you on the shore of the ocean, living and starting a family with **Floortje** and **Orlando** within the walls of the tiniest capital possible reminds me how unpredictable and wonderful life can be. I'm looking forward working with you again in the future!

Jan, there is a lot I'd like to say to you, half of which is probably too cheesy for you bear reading, let alone be read by anyone looking at my thesis. So, I'll try writing it your way: You've been a pillar in difficult moments, your contribution has been vital to the project, thank you for not killing me when we had to share a flat in den Haag.

Erik, **Thijs**, both of you have been operating the crucible of DUST. I can't thank you enough for giving this crazy project flesh, and tolerating my sudden outbursts of unrealistic ideation. You are also at the origin of the first "pirate prototype" of DUST, a creepy balloon-shooter that convinced everyone on the team that a switch for zombies was needed. Thank you!

Marieke, our paths went separate ways at a moment where the boat was already sailing rather smoothly, which is also thanks to your help and guidance in getting this complex project to work.

Annika, I miss our recurrent stumbling upon each other in the streets of Nijmegen! Having you in our team for counseling and facilitating communication with the police organization have been vital!

Aniek, you've been more than simply helping out in that stressful data collection endeavor in Sevenum. You've been a very reliable colleague, and your scientific contributions really added something substantial to the first experimental paper I wrote!

Evert, I still can't thank you enough for the role you played in the early years of the project. Your inquisitive mind has been able to spot many of the early imperfections of the project. Even if the final development wasn't your doing, we still owe you for it!

Evan, this section would not have been complete without you. You've been a guide in many ways in the early years of the project, like in your insistence in early stakeholder involvement. You're the one who brought Wendy on board! I still go by all the lessons you taught me. Above all, I remember the motto: "Don't ask for permission, but beg for forgiveness". Rest assured: I went into that head first!

2.2 EPAN

To fit the lab's nomenclature this section should be called Thanks2You, or something of the sort :)

Well **Bob**, I really had to force myself not to make any joke about you boat's unofficial name (Its **Reinoud**'s fault anyway!), on which I had the privilege to place some rivets. I can't wait to sail with you! **Felix** thank you for some mixed models help, fully memorable music and deep conversations. Deep conversations being my favorite, I also have to thank other people that spent some time helping me in my soul-searching moments, namely to **Anna T.** and **Linda**.

I also want to thank all the previous researchers on the Police front of EPAN, **Mahur, Wei, Saskia, Iris** and **Reinoud** for laying the bases of my project! Wouldn't have been here without you!

Anneloes, your help was cardinal to help me navigate through mixed models… and Donders parties!

Moniek, your way of always putting the patients and participants first has always been something I admired in you!

Lars I still feel like you'd have loved to be part of our gaming project, thanks for the inputs you gave after trying it!

Anna D., I still have the gifts you gave me at my housewarming party! **Floor**, I'm still scared to make jokes in front of you, even if **Sjoerd** might empathically laugh.

James and **Agnieszka**, as the other 2 "distant" members of the lab, I was always happy that I could count on you not to be the only one attending meetings from zoom. **Eliana**, grazie per la tua bella energia e le risate in conferenza!

When I look at what is around the corner, I can't wait to see what the continuation of my project will be, also thanks to **Mariana**, **Andy** and more recently **Teun & Jonathan**. I'm so glad you picked up the ball!

2.3 GemHLab

My dear 6th floor crew! It has been such a delight to haunt those corridors! I got a first introduction by our lovely **Anouk T.**, who started introducing me to the lab by telling me who not to hit on (couldn't spare you on this one!). Then only, I had the chance get acquainted to the GemH's game world. It started with **Joanneke** and her project DEEP, which was some of the most inspiring and relatable work I took as a reference for my project. Speaking of games... **Anouk P.**: game-wise, you ARE a silent assassin as **Ken** once said, even though I'm still more scared by **Babette**'s boardgame moves and **Hanneke**'s dissertation book (and the fact that you laughed when I hit your boyfriend while he played my VR game...). **Hanneke, Geert, Aniek and Anouk:** despite all my jokes around the Swangerschapverlof (I purposefully didn't check spelling here), I am still jealous of the marvelous fact that you could start your own family while doing a PhD! **Marlou**, your Christmas cards and waitlist effects always made me wonder! **Nastasia**, or Natasha... anyway, my dear Renaissance Lady, I'm glad we could share a passion for... having passions!

Hiromitsu and **Katy**, we've been part of the same foster-family for a bit, and I will always remember walking up Mount Calvary with the first, whilst having a storytelling-camping-fire-evening-indoors with the latter (where I discovered that you do have a Pony-horse!!!). A big, big hug to you all, my favorite weirdoes!

A

2.4 Police

Wendy, I already thanked you in a previous section, but I need to add that I wouldn't have met any of the people below without you!

René, you've been such an incredible and tireless help in the data collection of my last study, you've been both insightful and funny, a powerful mix!

There are countless **IBT docents and supporting team members** that I would like to thank. But I guess that both for privacy concerns as well as list length it won't be possible to thank you all here. But still, if we happen to meet in Sevenum, den Haag (YP!), Rotterdam, Amsterdam, Alphen aan den Rijn, Leusden, Zaandijk, Sevenum, Breda, Elst, Soesterberg, Huis 't Velde and Apeldoorn, be certain that I would have wanted to mention you and thank you for your time, energy and motivation. I really appreciated every minute that I spent interacting with police officers and trainers. It was always an exciting and thought-provoking journey!!

2.5 Other colleagues

Séolane, you've been the colleague I had dreamt to have earlier in my PhD trajectory! Happy my moving back to Switzerland made it happen! **Adrien**, you've been a tether to the Donders in Lausanne! **Tomas**, my former mentor… if I'm here it is all your fault! Thank you so much for it! With you I started my path on complex systems… that I could only fantasize about when looking over **Merlijn**'s shoulder, or **Edmond**'s and **Daan**'s presentations. In another life, I'd do a PhD with **Anna L-A** too. Or maybe **Maaike C.**'s group (Loved being part of your group meetings for a short while)! **Daan** and **Andrea** (hope your brother becomes a scientist too!) you've been wonderful officemates! Speaking of office life, I loved it to be perturbed by **Freek**'s and **Sander**'s "scientific b*ching moments". I also was super happy to drop by other people's desks to bother them with deep ecumenic questions about life, Pony horses and how the heck to survive in the Netherlands (shootout to **Katja**, **Christel** and **Meta**!). **Giel**, merci pour nos échanges pendant ces années! And also, thanks to **Toon** for the most amazing battle dance ever videotaped in the BSI! A big thank you also to all the other colleagues in both the DP and EPT "departments"!

Donders-wise, there is quite a bunch of people that come to mind. Form **Dan**'s knowledge of dive-bars to **Thomas**'s moments of blissful meditation, the conversations with **Mora**, in Italian, rigorously at rush hour! Thank you all!

3. The Dutch crew

3.1 Dollars

There are many names and faces that pop up when I think about my happiest moments in the Netherlands. Many include a beer and a rock band playing endlessly the same 80's and 90's tunes, while I'm having a great time with **Joanne, Marten, Dorien, Sid, Dario, Yves, Hendri, Max, Senne, Bobby, Anne, Collin, Marc, Freek, Kim, Karlijn**... and many more. You all contributed so much at making me feel home in Nijmegen!

3.2 Storyline

Dries, you've been more than a friend in these years of Netherlands (a brother and a dad as someone would say!). I can't count my blessings enough to have you in my life. **Timo & Verena** you too have been part of my extended family for years now, and we all noticed that Timo became better than me at playing Minecraft! **Sergio**, your contribution to the project has been substantial, as most of our scientific image for the outside world is materialized thanks to the amazing videos you made for us. Even my dad who doesn't speak English was impressed by the "behind the scenes" video!

Anita, **Esther**, **Anna**, **Rob**, **Maria**, **Flavia**, **Johanna**, all of you have been playing (literally) a crucial part in my emotional stability in these years of playback theatre rehearsal. Can't wait to play with you more!

3.3 Music friends

Most of you were more than that! Especially my former neighbors **Luuk** and **Dietke**, who still host me from times to times when I come visit in the Netherlands. Jamming in your enchanted flat, then enjoying a deep conversation while drinking tea in a home-made cup were highlights of my time in the Netherlands! Thanks also to the "hardcore team" of jammers that routinely came to the "Jamming Orange" sessions I ended up organizing weekly in my room, namely **Jody**, **Jule**, **Giel**, but also **Kimon**, **Henri**, **Corina**, **Thomas**, and many, many more.

4. The North American crew

4.1 Canada

Marc, I always say that through you I've discovered a different way to experience the relation to my dad. I still haven't finished building my sitar, but once I do, I'll ask for a jam! **Ruben**, Mr Fluff, thank you so much for giving me the alpaca in a moment of need. I thought of you every time I saw it! **Julian**, apart from the fact that I'm jealous

of everything you blissfully undertake, I always need your help to throw your brother in the swimming pool. **Tom**, thank you for all the time you invested explaining grids to me, and where to buy the center of doughnuts, and for your warm presence! **Viktor**, I'd still be marooned on that great lake island if it wasn't for you.

I probably will never visit in winter... but do keep the eastern skies in check in summer time!

4.2 USA

Paul, your work on dispatch priming bias has been quite an important part of my thesis. Thank you for your work and the time you invested in explaining the ins and outs of your task to me!

Skip, I was very happy to have you visiting Nijmegen in the first part of my PhD, your insights gave quite a start to my project. I learned from you at every single moment, up to the time I drove you back to the train station!

5. The Swiss crew

5.1 Family

When Covid hit, most of us went back to our close friends and family. This is also what I did, albeit in gradual steps. Living in Lausanne, right in the geographical middle between my siblings (**Alice**, **Roberta** and **Rocco**) and my dad **Mario**, was a necessity without which this thesis would never have been finished. Their support, along with the one of the extended family (**Pascal**, **Benoît** and **Alice**), could not be understated. A special thank you also to the friends of my mother: **Claire**, **Daniel**, **Beatrice** and **Renato**, as well as all her family in Italy, for helping me feel her presence in these last years. Veglia su di noi **Mutti**!

Lastly, thank you to my girlfriend **Nausicaa** for all the emotional support, as well as her parents **Jim** and **Viviane** for making me feel one of them.

5.2 Friends

Once I decided to settle in Switzerland to finish my PhD, I heavily relied on my friends to feel home again. A special Thank you to **Leo P, Ibasel, Naomie, Marcan, Salomé, Luca, Marion, Nathan, Jeremiah, Leo B, Isabelle, Ana, Barbora, Simon, Mie, Donàl, Marie, Giovanni, Adrien D., Franziska, Xavier, Aline and Adrien K.**

Behavioural
Science
Institute

Radboud University