



Wuthering heights: gauging fear at altitude in virtual reality

Giuseppe Boccignone¹ · Davide Gadia¹ · Dario Maggiorini¹  · Laura A. Ripamonti¹ · Valentina Tosto¹

Received: 15 December 2021 / Revised: 22 February 2022 / Accepted: 3 June 2022
© The Author(s) 2022

Abstract

In this study we propose an approach to assess the fear of heights through a 3D virtual reality environment. We show that an immersive scenario provides a suitable infrastructure to such purpose, when supported by related behavioural and physiological measurements. Our approach is grounded in the principled framework of constructed emotions. This allows to shape fear detection as a case of categorical perception, which is amenable to be formalised as an unsupervised learning problem. Meanwhile, it paves the way for addressing meaningful physiological parameters for the assessment. Gauging fear of heights in individuals, beyond its theoretical relevance, is cogent for the early discernment of workers who are unsuited for operating at altitude and who may require to undergo specific training or, eventually, to be recruited for different positions.

Keywords Virtual reality · Skills assessment · Affective computing · Unsupervised learning

1 Introduction

A considerable number of deaths following work-related accidents is due to fall from heights (around 25% [32]), and the number can be even higher among specific occupations, such as construction workers. Many safety strategies can mitigate the risk of fall from high height accidents. Yet, among several causation factors, human behavior and attitude play a significant role.

Safe working at height training is essential to enhance people's awareness of the issues involved with working at height, as much as gauging the worker's attitude in such conditions. Unfortunately, standard training requires specialized personnel and the setup of a very expensive environment in order to put the trainee under elevation stress. Recently, *Virtual Reality* (VR) technology proved to be a viable alternative to training in a real setup [53]. By using VR, it is possible expose the trainee to a potentially harmful situation by providing

✉ Dario Maggiorini
dario@di.unimi.it

¹ Department of Computer Science, University of Milan, via Celoria 18, Milan, I-20133, Italy

a visual stimulus of the working environment without recreating a real dangerous situation. Moreover, in a virtual environment it is possible to seamlessly monitor physiological parameters to achieve a better understanding of the situation.

In this paper, we consider a VR environment to evaluate a person's suitability to work at height. In the simulation, the subject is supposed to walk on a virtual plank suspended in the void while wearing an Oculus device. A sensor is used to record physiological parameters in real time and a camera records the subject's behavior for later inspection. In such context, we propose an evaluation approach, which, different from other proposals, is grounded in the latest results achieved in emotion theory and emotion neuroscience [10, 11, 14, 15, 58].

An earlier report on the bare experimental setup and preliminary results was presented in [19]. Here, more generally, we spell our work in the principled framework of constructed emotions. To the best of our knowledge this offers a novel perspective to the affective computing field as currently stated [29, 46, 59, 65] and, markedly, to the reasearch problem we are addressing.

We show how methods for solving the problem of fear of height detection can be soundly formalized as a categorization problem, which, in turn, is amenable to be shaped in the form of an unsupervised learning problem. On this basis, the relationship between emotion/affect states and physiological parameters can be investigated in a principled way to assess the attitude of an individual to work at height.

The remainder of this paper is organized as follows. In Section 2 related work is discussed and our contribution outlined. In Section 3 we elucidate and motivate the rationales behind our work. Section 4 describes the virtual plank experiment and the associated protocol. The model-based data analysis is presented and discussed in Section 5. Eventually, Section 6 provides a recap of the main results achieved so far, together with an in-depth discussion of possible limitations and viable developments of this study.

2 Related work and our contribution

A variety of studies have blossomed in recent years concerning the analysis of emotional response (to be intended in a broad sense, in the brief overview that follows) as elicited in virtual environments. By and large, these studies mainly concentrate on how to scrutinize physiological responses and on the way different environments can trigger distinct emotions.

Physiological responses The general approach is to expose the subject to a stressful, demanding context and then correlate physiological parameters, such as heart rate and blood pressure variation, to the proposed virtual situation.

In [49] a subject is placed first inside an elevator and then on an aerial moving platform, from which she is supposed to jump off. Authors claim that, despite an increase of heart rate, blood pressure and hydrocortisone (also known as cortisol) levels remain constant while on the platform. Moreover, the hydrocortisone level decreases when inside the elevator.

In [6], instead, subjects are asked to traverse a grid of ice blocks with the risk of falling down. During the experiment, participant's movements are recorded alongside skin conductance level and facial electromyography. Authors found that a risk-averse behavior was more evident in participants with an high neuroticism personality profile. Moreover, these subjects also made more frequent Risk Assessments than the average.

Authors of [38] measure cardiovascular and cortisol reactivity to the VR equivalent of a *Trier Social Stress Test* (TSST). In a TSST the participant is asked to hold a speech and

to engage in an arithmetic task in front of an audience to reproduce stress in laboratory settings. Virtual reality was used to recreate a virtual audience to the presenter and, for the proposed case, results resembled those obtained in prior studies using a real-life TSST.

Other contributions, such as [25] focus on understanding whether a VR environment is suitable to generate the right psychophysiological condition for an effective exposure treatment. In [25], authors found that VR exposure does evoke psychophysiological arousal, especially in terms of electrodermal activity, making it feasible for cognitive behavioral therapy. With the goal of assessing the effectiveness of VR exposure therapy, in [26] 40 patients with acrophobia and 40 matched healthy controls participated to a VR height challenge and assessed subjective (fear ratings) and physiological (heart rate, skin conductance level, salivary cortisol) fear reactions. Both groups reported statistically significant increase of physiological levels. However in that case, physiological arousal in acrophobic patients, in contrast to subjective fear, was no stronger than that of controls when confronted with height cues in VR. Surprisingly, there was no increase in salivary cortisol levels in either group.

Emotion/affect detection In this case, the paradigm is to devise and perform a reliable classification of the emotional state of the subject and correlate it with her virtual experience.

One such approach, is represented by the *Affective Virtual Reality System* (AVRS) [44]. AVRS detects the level of arousal (i.e., the autonomic nervous system stimulation) through measurement of the heart rate and using the *Self-Assessment Manikin* (SAM) technique. Using SAM, a stylized figure representing the intensity of the affect dimension must be selected on a scale or grid (see Fig. (3)a in Section 4, for an example). In the paper, the levels of arousal solicited by the same video in VR and via standard screen are compared. Authors claim that the average arousal level is higher in VR when the scene is depicting happiness or fear, while there is no significant difference for other emotions. As to fear, in particular, the difference is likely to depend on the negativity of the scene and on the VR visual quality. The only emotion that appears to be stronger on screen is represented by distaste, though changes of the heart rate did not reveal to be significant between the two technologies.

In a similar vein, the EMMA project [7] addresses the development of a *Mood Induction Procedure* based on VR (VR-MIP) with the goal of eliciting sadness, joy, anxiety, and relax in the involved subjects. To this purpose, subjects were exposed to virtual environments populated with objects suitable to elicit emotion states. Along the experiment, the virtual environment proposed to the subject was neutral at first, and subsequently modified so to trigger a specific affective state. In this project the intensity of the experienced emotion has been evaluated on the basis of three feedback forms: *Visual Analogue Scale* (VAS), *ITC-Sense of Presence Inventory* (ITC-SOPI), and *Reality Judgement and Presence Questionnaire* (RJPJ). Results demonstrated that the four environments (one for each emotion) were actually able to induce an affective state change in subjects. Interestingly enough, and differently from other contributions, in [7] anxiety is kept well separated from fear, in accord with Barlow [8] definition of anxiety as “a diffuse, objectless apprehension”, while fear is best characterised as an emotional triggered by a “present” and specific threat.

Somewhat relevant, it has been shown that factors less than obvious can play a role for emotion elicitation and detection. For instance, in [23] authors give evidence that observed geometric shapes and daylight illumination correlate with heart rate and skin conductance. To such end, different building façade patterns have been considered. Each pattern induced light characterized by a specific geometric diffusion (irregular, regular, and “Venetian style”) inside a virtual space. Results suggest that participants find the space to be more interesting when illuminated through an irregular pattern. In that case, the mean heart rate was

lower than in other conditions; meanwhile, subjects reported to be more relaxed. Indeed, the negative correlation between mean heart rate change and interest was statistically significant.

Our approach In spite of the growing number of studies in the field and of the apparently intuitive definition of “fear”, there is an overall lack of clarity of what exactly is elicited and measured along experiments [54]. Indeed, the actual experience of fear is a complex phenomenon (cfr. left panel of Fig. 1) binding together the external perception of the world, the internal perception of the body, conceptual knowledge about fear emotion itself, and past experience (e.g., previously experienced situation or episodes of fear). Differently from previous work, we first frame the problem addressed here in a solid framework, namely the theory of constructed emotion (which is summarised at a glance in the right panel of Fig. 1). In such theory, terms like “emotion” and “affect” (and thus possibly related measurements) have a clear meaning at different levels of explanation (conceptual level and core affect/interoceptive level, respectively). The setup of the experiment and measured subjects’ behavior can thus be accordingly interpreted.

Next, based on the definition of emotion as a category, we set up a bridge between a formal Bayesian approach proposed in the field of perceptual categorization and fear perception. This allows to model and to analyze the data collected along the experiment in terms of a straightforward and explainable unsupervised learning technique.

In the following Section we introduce and motivate the theoretical essentials of our approach.

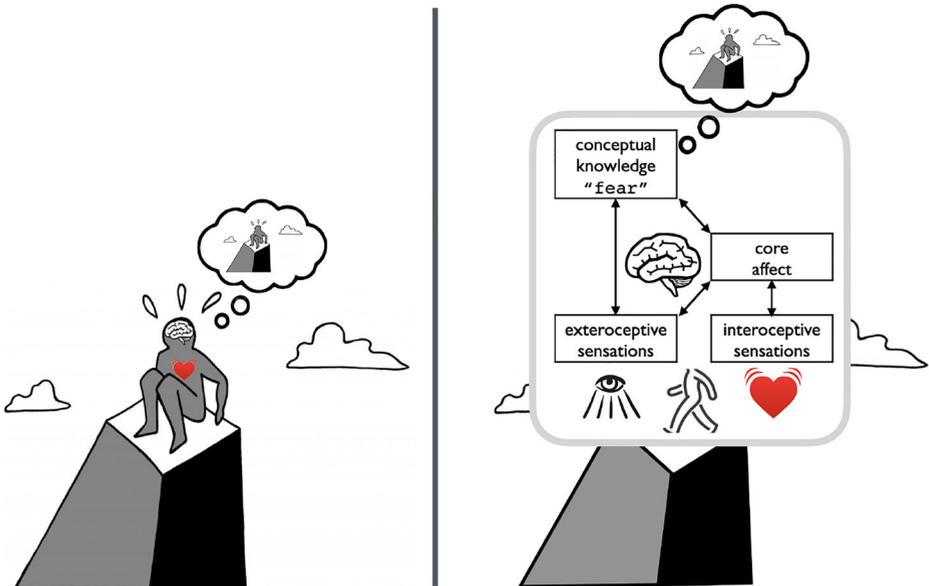


Fig. 1 Fear as situated conceptualization. *Left panel*: the actual experience of fear bounding together the external perception of the world, the internal perception of the body, conceptual knowledge about emotion, and past experience. *Right panel*: the theoretical view of fear as a categorical emotion constructed through a conceptual act, the result of brain’s active, ongoing predictive processing endeavour (see text for details)

3 Background and rationales

In order to 1) elicit (*how* problem) and 2) measure the physical states and subjective feelings that researchers refer to as “affect” or “emotion” (*what* problem), we need to establish a principled theoretical framework that makes clear how these two issues can be addressed. For instance, we preliminary need to exactly define what we intend by spelling “affect” or “emotion” [54]. One such framework is the view of emotion as a categorical construction, which is outlined at a glance in the right panel of Fig. 1 and which is presented to some detail in the remainder of this Section.

It is worth noticing that in the literature concerning the computational modelling of emotions, markedly in the affective computing field, the term “affect” is often used interchangeably with that of “emotion” but they should not be confused. Affective computing is a vibrant area of interdisciplinary research (see, e.g. [29, 46, 59, 65] for up-to-date overviews). Unfortunately, it is often the case that basic concepts and working assumptions are loosely defined; also, the same terms are occasionally adopted to denote different classes of phenomena. For instance, it is common parlance in the field to conflate the categorical description of emotions - fear, anger, joy, etc. - as typically derived from *Basic Emotion Theories* (BET, e.g., Ekman’s Neurocultural Theory [30, 31]) with Russell’s dimensional representation of affect (over the valence/arousal dimensions, [56]). The first one is usually referred to as a discrete representation of emotions; the second, as a continuous representation of emotions [46]. This clearly is, at best, an incorrect statement: Russell’s psychological construction view of emotions posits emotion and affect as different phenomena. Incidentally, BET and the construction view are theories in stark contrast with one another [10].

The theory of constructed emotion: a brief tour In the perspective that motivates the present work, emotions (e.g., fear) are the result of situated conceptualizations (whose overall process is depicted in the right panel of Fig. 1) constructed from affect. In this view, emotional events are specific instances of affect that are linked to the immediate situation and involve intentions to act [12, 14].

Here, differently from what proposed in BET [31, 37, 51, 63], emotions labelled by words as “fear”, “anger”, “surprise” are not natural kinds, Platonic essences wired in the brain, but abstract, *ad hoc* categories [11, 15]. A *category* can be defined as a population of events or objects that are treated as similar because they all serve a particular goal in some context [11]. Perceived fear, for instance, is the result of categorizing the current situation (sensations, context) as fearful when it contains features similar to those of previous situations that have been experienced as fearful [45].

A category has a mental representation, a *concept*, namely the population of representations that correspond to category’s events or objects. Moment by moment, conceptual representations are tested against the incoming sensory evidence - from the external world and from the body - to categorize it according to past experience, in the effort of anticipating body’s needs and preparing to satisfy those needs before they arise (allostasis, [58]). This dynamics is represented in the right panel of Fig. 1 by bidirectional arrows connecting the different components: when the information flows from concept to sensations, the agent is generating predictions; otherwise, the information flows from sensations to concept, representing agent’s inference or categorization

In such endeavour, *interoceptive sensations* from the body play a cogent role because weighing which parts of the world are worth caring about in the moment. Without them, an

actual agent would not appraise relevant features of the physical surroundings. Interoception is fundamental to construct the pivoting psychological primitive [13] named the “core affect”. *Core affect* can be described as a state of pleasure or displeasure, named *valence*, with some degree of *arousal*. Together, valence and arousal form a unified, continuous state-space. It is referred to as “core” because it is grounded in the *internal milieu*, an integrated sensory representation of the physiological state of the body: the somatovisceral, kinaesthetic, proprioceptive, and neurochemical fluctuations that take place within the core of the body.

Core affect is realized by integrating incoming sensory information from the external world, i.e. *exteroceptive sensations*, with internal, interoceptive information from the body (see Fig. 1). The result is a mental state that can be used to safely navigate the world by predicting, for instance, reward and threat. Thus, by no means affect can be equated to emotion.

In this perspective, in every waking moment, brains function as predictive machines that run on concepts - predictive internal models - to give sensations, either exteroceptive or interoceptive, meaning. Note that, under such circumstances, there is no specific difference between emotion, vision or audition: when we focus on some of those sensations that are exquisitely interoceptive, the resulting experience can be an instance of emotion [12, 14].

Having clarified the general framework, we now turn back to the issues of elicitation and measurement raised at the beginning of this Section.

The elicitation problem For what concerns induction (the *how* question) there is a number of methods typically used for evoking affect more generally, and emotion more specifically. These have been in-depth reviewed and evaluated in [54].

Among such variety of methods, VR allows subjects to immerse themselves in a social situation or a scene in a first-person way (as opposed to viewing the scene in a third-person way). Quigley et al. [54] conclude that the VR method is likely to provide a potent way to induce either affect or emotion, and, as other immersive technologies, it has the potential “to radically change affect and emotion research”.

VR main feature is the ability to induce a feeling of “presence” in the computer-generated world experienced by the user [55]. Presence refers to the subjective sense of reality of the world and of the self within the world; more precisely, in VR, presence is used in a subjective–phenomenal sense to refer to the sense of “being now there” in a virtual environment (VE) rather than in the actual physical environment. This sense establishes a behavioral/functional equivalence between virtual and real environments [57, 60]. It has been argued that the sense of “being there” in a VE is grounded on the ability to “do there” [57, 60], the latter being related to the sense of “agency”, namely the sense that a person’s action is the consequence of his or her intention.

The rationale behind this hypothesis has been discussed by Seth et al. [60] and grounds in the situated conceptualization framework that we have previously outlined. The general idea is that presence in a VE is underpinned by good matches between expected and actual sensorimotor signals, which in turn requires interacting interoceptive and exteroceptive processes.

Functionally, this is achieved by two primary components, an “agency component” and a “presence component,” mutually interacting and connected, respectively, with the sensorimotor system and the autonomic/motivational system. Briefly, their model associates presence with successful suppression by top-down predictions of informative interoceptive signals evoked by autonomic control signals and, indirectly, by visceral responses to afferent sensory signals. The predicted interoceptive signals will depend on whether afferent

sensory signals are determined, by the parallel predictive-coding mechanism controlled by the agency component, to be self-generated or externally caused. Thus, the role of the agency component with respect to presence is critical: it provides predictions about future interoceptive states on the basis of a parallel predictive model of sensorimotor interactions (see [60], for model details and possible neural correlates of functional components). It is however clear, that presence, agency and actual emotional experience are closely intertwined, by sharing a common functional architecture.

The measurement problem Since we are interested in the interoceptive pathway (see Fig. 1, right panel), in the present study we will use physiological measures of the Autonomic Nervous System (ANS) activity. It is clear from the model that, affective states may be primed by top-down processes (along the predictive/generative step) or bottom-up, when making sense of ANS activity, and that the top-down and bottom-up mechanisms may mutually reinforce one another (e.g., as in panic disorder).

The ANS innervates smooth muscles (e.g., the heart) and glands, and is divided into the sympathetic and parasympathetic branches (SNS and PNS, respectively). Indeed, autonomic changes are integral to affect and emotion [40, 50, 54]. In particular, we will adopt the most frequently used physiological measures of heart and electrodermal activity (EDA) [26, 27].

Heartbeat is primarily produced by the sinoatrial node, which generates action potentials that course throughout the cardiac tissue, causing regions of the heart muscles to contract in the orchestrated fashion that characterizes a heartbeat. Meanwhile, the heart is innervated by the sympathetic and parasympathetic branches, which regulate the heart rate by influencing the activity of the sinoatrial node. Activation of sympathetic fibers has an excitatory influence on the firing rate of the sinoatrial node, resulting in increased heart rate. Alternatively, the parasympathetic activation has an inhibitory influence on the pace-making activity of the sinoatrial node and produces decreased heart rate. As to cardiac activity, recorded via ECG, parameters of interest for this type of study are the Heart Rate (HR, number of beats per unit of time) and Heart Rate Variability (HRV, variation in heart period, or rate, as a function of central respiratory drive or peripheral respiratory afferent input). In particular, HRV has been widely used in affective science.

EDA reflects changes in the electrical conductivity of the skin that results from both internal and external stimuli. Changes in EDA reflect changes in the activity of the eccrine sweat glands, which are exclusively innervated by the SNS. Because of this relationship between EDA activity and SNS activity, EDA has been widely used as an indirect measure of peripheral physiological arousal. The EDA signal is composed of two components, both a background, or tonic, level and phasic changes that are of shorter duration (on the order several seconds). The tonic level is referred to as the skin conductance level (SCL). The phasic changes are referred to as skin conductance responses (SCRs). SCRs can arise from stimuli external to the person (often called event-related skin conductance responses) or the responses can be non-specific meaning that there is no apparent external stimulus. Negatively valenced stimuli, such as anger and fear, are likely to elicit greater skin conductance (or emotional sweating) than positive stimuli.

However, the measurement of autonomic changes in relation to emotion and affect is not as simple. Although for several decades BET has led to a search for autonomic states specific to different emotions, current research has highlighted that varying degrees of overlap in autonomic response patterns do likely exist across discrete categories of emotions, accompanied by individual variation in the bodily expression of emotion [50, 54]. Affective states and responses are neurobiologically mediated by a network of distributed interacting

neural circuits ranging from the spinal cord to the limbic system and neocortex, and the link between psychological events and individual physiological responses is a complex one [22].

Even limiting to the fear emotion, it is well known that fear can come in many flavors (i.e., fear of predation, fear of starvation, fear of pain, but also social fear such as in public speaking, etc.), each associated with the integration of information at different levels that are subsequently associated with distinct neurobehavioral and physiological responses extending across different temporal scales [24, 43, 50]. Interestingly enough, studies of fear learning in rodents have found evidence of variability in ANS responses and neural circuitry [36, 43].

By and large, studies point to broad sympathetic activation, including cardiac acceleration, increased myocardial contractility, vasoconstriction, and increased electrodermal activity. However, a number of studies reported decreased HR and increased SCR in response to threatening material (e.g., spiders) [40]. It has been surmised that when stimuli elicited a stronger degree of self-involvement, leading to higher imminence of threat, participants were characterized by immobilization (freezing effect) rather than an active coping response that leads to sympathetic inhibition [40]. As to cardiac activity, it has been further noticed that aversive conditioned stimuli can produce coactivation of the sympathetic and parasympathetic branches, yielding accelerated, decelerated, or even unchanged heart rate, depending on the relative strength of sympathetic versus parasympathetic activation [9, 50].

A limited sample of studies have investigated physiological arousal during exposure in VR with respect to other elicitation techniques [26, 27]. A number of studies have shown significant HR reactions of phobic and fearful participants in fear-related VR environments. Overall, however, the evidence for an impact of VR challenges on HR is mixed, but there is convincing evidence that VR exposure leads to significant variations in EDA [26, 27].

4 Experimental setting and procedure

Materials We exploit *Richie's Plank Experience* as a virtual environment to expose the participants to a simulated height. To perform the test in the virtual space, the subject is first required to take a virtual elevator riding upward for a long time. This is to induce the feeling that an extremely high floor has been reached and to increase the sense of presence (see Section 3). When the elevator doors open, an altitude view of a modern cityscape is presented. The doors open on nothing but for a small plank protruding in the void for around one meter (see Fig. 2). The participant is supposed to walk in a real (and safe) environment to reach the end of the plank. In order to add realism, we also place a piece of wood on the floor for the participant to walk on. The experiment ends as soon as the subject reaches the end of the plank. The experimental setup can be appreciated from Fig. 3, presenting snapshots of the video recorded along the experiment.

Participants The experiment has been performed on 33 volunteer students; 28 males and 5 females in the age range between 22 and 30 years. Volunteers did not receive any payment or credit for their collaboration. All of them reported to be in good health: no cardiovascular pathologies, no anxiety disorders, and no neurological alterations. Moreover, in order to be accepted as volunteer each student must not had been exposed to the *Richie's Plank Experience* before.



Fig. 2 The plank protruding out from the elevator door

Procedure Each experiment articulates in three steps:

1. **Pre-treatment step.** Participants provided their informed consent and received a set of instructions about the experiment.
Each subject is required to sit and see a soothing video in order to stabilize her physiological data. While watching the video, the subject is also wearing the physiological sensor and a calibration is performed.
2. **Treatment step: VR exposure.** The subject is equipped with the Oculus Rift visor and instructed about the use of the equipment; then the simulation virtual environments is started. This step is taking one to four minutes to complete, depending on participant's immersion.
3. **Subjective affect rating step.** The participant using an affect annotation tool [20], is requested to evaluate her experience as described below.

Withdrawal from the study At the end of the experiment, 26 subjects provided valid data among the 33 participants; the other 7 were not able to complete the simulation due to being too scared from the virtual environment, and thus discarded.

4.1 Measurements

Physiological Parameters To collect physiological data we use the E_4 wristband from *Empatica* (Empatica Srl, Milan, Italy). The wristband contains four sensors: (1) an electrode for electrodermal activity (EDA), (2) a photoplethysmogram sensor to measure blood volume pulse (BVP) from which it derives heart rate (HR) and the inter beat interval (IBI), (3) 3-axis accelerometer and (4) a temperature sensor.

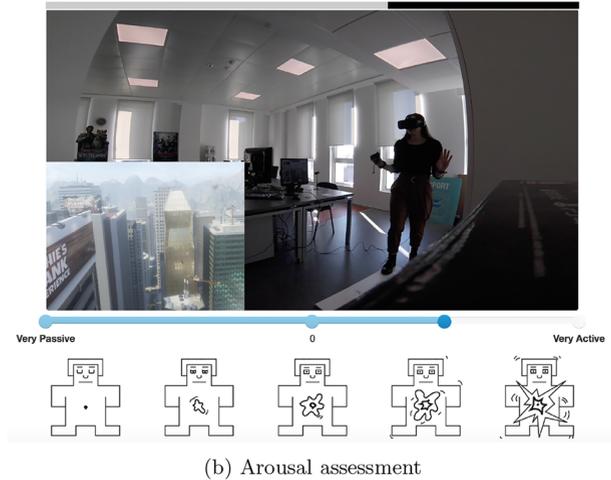
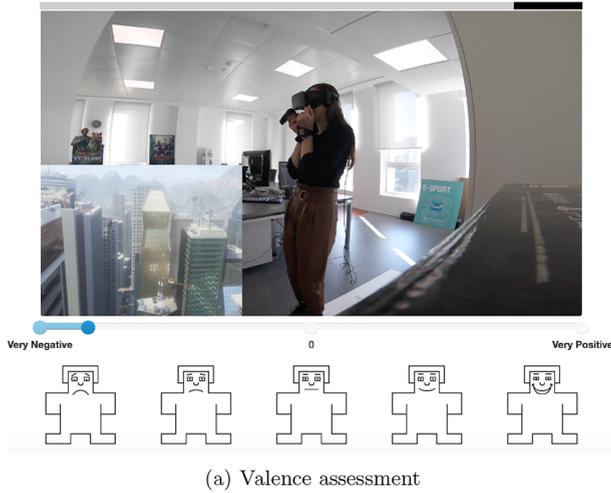


Fig. 3 Valence and arousal self-evaluation via DANTE. The interface provides a sliding bar for the continuous labelling of the video and the corresponding SAM (Self Assessment Manikin to guide participant's labelling choice)

In particular, the EDA sensor measures the variations in skin conductance. For this purpose, a small amount of alternating current is passed through the skin between the two silver-plated electrodes placed at the strap of the wristband. The skin conductance is recorded at a sampling frequency of 4 Hz.

Photoplethysmography (PPG) is a non-invasive, optical technique used to measure the volumetric changes in arterial blood resulting from the heart cycles. A photoplethysmogram sensor measures blood volume pulse (BVP) from which other cardiovascular features like HR may be derived. The E4 PPG sensor outputs two main time-series for each individual, which give us insight into HR function: HR and IBI. HR is sampled at 64 Hz. As to IBI, it represents the time in milliseconds between two successive heart-beats (R–R), detected after removing wrong beats. The sensor also uses an in-built motion artefact removal algorithm to remove unwanted signals from the data.

When the simulation is over, all data is downloaded from the wristband via bluetooth link for later processing.

Affect state assessment Collecting physiological parameters is not enough to reconstruct the overall status of the subject during the virtual experience. Thus, affective states are also taken into account. These are measured through subject's self-evaluation.

After the virtual plank walk, the subject is required to provide a self assessment about the levels of arousal and valence felt during the experiment.

To serve the purpose, each participant was recorded on camera during the experiment. This video, merged with another video recorded in first person from inside the virtual environment, is used to collect the affect measurements.

We ask the subject to annotate her own recorded video with an online freely available tool¹ named *Dimensional ANnotation Tool for Emotions* (DANTE) [20] (see Fig. 3).

While replaying the video, the participant can report in real time the levels of arousal and valence, using a sliding bar (with values ranging from -1 to $+1$ and a step of 0.001). DANTE interface includes a SAM visualization specific for the selected affective dimension (arousal or valence), to help annotators. The video timestamps allow to map each affect state to the right position in the data sequence collected by the sensor.

5 Data analysis

In this Section we first set up a bridge between a rational analysis model of categorical perception and the constructionist view of emotion categories previously introduced. Then we introduce the features most apt to characterize physiological responses. Eventually, we analyze and assess results obtained along the experiment at the light of the introduced framework.

5.1 Data modelling

Recall that the goal of the present work is to gauge the attitude of an individual to work at height as related to a possible fear experience. Methodologically, on the one hand, we need to set up a data-driven procedure for such assessment, exploiting data collected after an experiment/evaluation session; on the other hand, such procedure, should be grounded in, or, at least, informed from the framework we have in Section 3.

A viable solution is to take into account the close theoretical link between the constructionist view of emotion categories and rational analysis models of categorical perception [3, 33].

Considers again the model outlined in Fig. 1, in particular the interoceptive pathway. The generative process undertaken by a subject relying on a set of categories C to produce actual multi-modal physiological responses, say R , given a context model \mathcal{M} (situation, subject's history, traits, and so on) can be simply summarised by the following sampling steps: given context \mathcal{M} , sample the most plausible category c ; sample an affect state a , given c ; sample a response r , given a .

¹<https://github.com/phuselab/DANTE>

Clearly, subject's affect states are not directly available, except for participants' self-evaluations collected at the end of the procedure (see Section 4). For our purposes, thus, it will suffice to boil down the full sampling procedure to the following.

Denote: $\mathbf{R} = \{\mathbf{r}_n\}_{n=1}^N$ the set of observable (or computable features of) multi-modal physiological responses (vectors); $c_n \in \mathbf{C}$ the random realization of a hidden category from the set \mathbf{C} of K categories, which generates the n -th multi-modal response $\mathbf{r}_n \in \mathbf{R}$. Category c_n can be shaped as a 1-of- K binary random vector of components $\{c_{nk}\}_{k=1}^K$, in which a particular element c_{nk} is equal to 1 and all other elements are equal to 0, that is $c_{nk} \in \{0, 1\}$ and $\sum_k c_{nk} = 1$.

The probabilistic generative model is represented by the joint distribution $P(\mathbf{R}, \mathbf{C} \mid \Theta, \mathcal{M}) = P(\mathbf{R} \mid \mathbf{C}, \Theta_R)P(\mathbf{C} \mid \Theta_C, \mathcal{M})$, $\Theta = \{\Theta_R, \Theta_C\}$ being a suitable set of parameters. Then, the generative process from categories to observed responses, for $n = 1 \dots N$ observations, can be written as:

1. sample category

$$c_n \sim P(\mathbf{C} \mid \Theta_C, \mathcal{M});$$

2. sample response, given the category

$$\mathbf{r}_n \sim P(\mathbf{R} \mid c_n, \Theta_R).$$

In the scope of this work, we instantiate the procedure in the following simple model. We let the likelihood term to account for Gaussian observations $P(\mathbf{R} \mid \mathbf{C}, \Theta_R) = \prod_{n=1}^N P(\mathbf{r}_n \mid c_n, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{n=1}^N \prod_{k=1}^K \mathcal{N}(\mathbf{r}_n, \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})^{c_{nk}}$, each observation \mathbf{r}_n being generated by the k -th Gaussian, indexed by category k , with mean and precision (inverse covariance) $\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k$, respectively; thus, $\Theta_R = \{\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k\}_{k=1}^K$. The prior term is the categorical distribution $P(\mathbf{C} \mid \Theta_C, \mathcal{M}) = P(\mathbf{C} \mid \boldsymbol{\pi}) = \prod_{n=1}^N P(c_n \mid \boldsymbol{\pi}) = \prod_{n=1}^N \prod_{k=1}^K \pi_k^{c_{nk}}$, with $\boldsymbol{\pi} = \{\pi_k\}_{k=1}^K$ the set of prior probabilities for each category (\mathcal{M} , has been omitted for notational simplicity).

Note that in such form the model corresponds to the Gaussian Mixture Model (GMM), $P(\mathbf{r}_n \mid \Theta_R) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{r}_n; \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})$, with π_k playing the role of the mixing coefficients of the K Gaussians. In this case, parameter learning can be easily accomplished through approximate maximum likelihood estimation via the iterative Expectation-Maximization (EM) algorithm [47]. The EM algorithm is well-known in the machine learning literature as an unsupervised clustering algorithm (which, in the case of isotropic Gaussian components and hard-assignment of points to Gaussian clusters, "degenerates" to the classic k-means clustering algorithm, but see [47], for details)

By and large, either in studies of emotion, and markedly in affective computing, researchers use supervised learning (i.e., classification), guided by emotion labels, to attempt to discover "fingerprints" in the brain or body for the corresponding emotion categories. Unsupervised clustering methods consistently can treat the number of clusters as a statistical parameter to be learned which, in principle, allows researchers to discover a more flexible and variable category structure should it exist, while also discovering similarities across subjects, should those exist [4].

In what follows we describe the set of descriptors of EDA and cardiac activities selected to form the response vector $\mathbf{R} = \{\mathbf{r}_n\}_{n=1}^N$ from the collected time series and related analyses.

5.2 EDA analysis

Skin conductance (SC) changes are mainly dependent on the activity of sweat glands innervated by the sympathetic branch of the ANS. The time series of SC can be characterized

by a slowly varying tonic activity (the skin conductance level, SCL) and a fast varying phasic activity (skin conductance response, SCRs). A sudomotor nerve burst corresponds to an observable SCR. In mathematical terms, sudomotor nerve activity can be considered as a driver, consisting of a sequence of mostly distinct impulses (i.e., sudomotor nerve bursts), which trigger a specific impulse response (IRF, i.e., SCRs). In mathematical terms, SC activity thus can be assumed to be composed as follows:

$$SC = SC_{tonic} + SC_{phasic} = SCL + Driver_{phasic} \otimes IRF, \tag{1}$$

where \otimes denotes convolution, and $SC_{phasic} = Driver_{phasic} \otimes IRF$. Currently, there are many software libraries that allow automatic detection of SCRs. Here, we have adopted the Python NeuroKit toolbox² to recover the SCL and the SCRs.

Based on the the current literature, we focused on 15 descriptors for EDA and 15 for heart rate. The detailed discussion of each descriptor is out of the scope of current paper (see [1, 35, 52, 61, 64]). A summary in the next two subsections will suffice.

EDA Descriptors Leveraging on [1, 35, 52, 61, 64], we considered six *event-related descriptors* [61]. They are: SCR peaks number, SCR total and mean amplitude, SCR total and mean signal raise time, area under the curve (AUC). The only non-trivial descriptor is the AUC. This is related to the number and amplitude of spontaneous oscillations [5] and based on (1) can be computed as:

$$r_{AUC} = \int SC(t)dt - SCL = cn\bar{a} + \epsilon, \tag{2}$$

by defining $Driver_{phasic}(t) = \sum_{i=1}^n a_i \delta(t - T_i)$, δ being the Dirac delta function, $c = \int IRS(t)dt$, and where \bar{a} is the mean amplitude of spontaneous fluctuations occurring at times $T_i : i1, \dots, n$, ϵ denotes some error that absorbs random fluctuations and any violations of time invariance and linearity assumptions.

We then selected the standard *statistical descriptors* of the SC signal: mean, standard deviation, kurtosis, and asymmetry. In addition we computed Shannon’s Entropy

$$r_H = \sum_x p_x \log_b p_x, \tag{3}$$

as a measure of the uncertainty for a random variable.

Eventually, we measured the three Hjorth parameters, originally conceived for electroencephalogram (EEG) signals over time [61]: activity

$$r_{Ax} = var(x(l)) = \frac{1}{L-1} \sum_{l=1}^L (x(l) - \mu)^2, \tag{4}$$

mobility

$$r_{Mx} = \sqrt{\frac{var(\frac{dx(l)}{dl})}{varx(l)}}, \tag{5}$$

and complexity

$$r_{Cx} = \frac{r_M(\frac{dx(l)}{dl})}{r_M(x(l))}, \tag{6}$$

²<https://github.com/neuropsychology/NeuroKit>

where $x(l)$, $l = 1 \cdots L$ denotes the discrete time series of length L . Activity, mobility, and complexity are related in some respect to signal power, mean frequency, and the change in frequency, in the frequency respectively.

The last descriptor is the classic power spectral density (PSD) defined using Fourier transform. The bulk of the energy of the phasic component is considered to be in the band from 0.05 to 1-2 Hz and the total power [μS^2] within that frequency band was computed [35, 64]. Spectra of EDA signals were calculated using Welch's periodogram method with 50% data overlap [52].

Eventually, all features were z-normalized to have mean 0 and to have a standard deviation of 1.

5.3 Heart rate analysis

As to cardiac activity, we based our analysis on heart rate variability (HRV) or R-R interval variability, which quantifies the successive variations in the interval from the peak of one QRS complex to the peak of the next. HRV is recognized in literature as an indicator of physiological stress and arousal. An increase of the arousal level is associated to low HRV values while a decrease in arousal is associated to high levels.

To measure HRV, time intervals between heart beats (R-R), are considered, or more precisely, Normal to Normal (N-N) intervals where outliers or ectopic beats that originate outside the rights atrium's sinoatrial node have been removed. The HRV values have been investigated in both time and frequency domains.

HR descriptors The first four descriptors are based on standard statistics in the *time domain*: mean, standard deviation, minimum and maximum values for HRV. Seven descriptors are based on the IBI: NN mean, median and range (difference between minimum and maximum values), SDNN, SDSN, RMSSD, and NN50.

SDNN is the standard deviation of NN intervals and represents the variability over the entire recording period, giving the overall autonomic modulation regardless of sympathetic or parasympathetic branch. SDNN tends to be higher when the LF band has more power compared to the HF band. SDSN is the standard deviation of the differences between subsequent NN intervals. RMSSD is defined as the root mean squared differences of successive difference of intervals, also based on normal sinus beats,

$$\text{RMSSD} = \sqrt{\frac{\sum_{i=1}^N ((R - R)_{i+1} - (R - R)_i)^2}{N - 1}} \quad (7)$$

RMSSD is the main estimation for PNS-mediated changes in HRV.

Last, NN50 is the number of successive differences of intervals which differ by more than 50 ms (pNN50 percent expressed as a percentage of the total number of heartbeats analyzed) and similar to RMSSD it has been developed to quantify high frequency variations arising from parasympathetic activity [48].

The next three indicators are in the *frequency domain*. The default bandwidths used to calculate the total power are, ultra-low frequencies (ULF, 0 – 0.003 Hz), very low frequencies (VLF, 0.003 – 0.04 Hz), low frequencies (LF, 0.04 – 0.15), and high frequencies (HF, 0.15 – 0.40). Total spectral power indicates overall HRV, and is used to assess overall autonomic cardiac modulation. LF and HF together are generally exploited to estimate sympathetic modulation.

Selected features are spectral density, HF and LF variance for HRV. For the spectral density we use the same PSD estimate already presented for the EDA value. HF and LF variance are measurements of the sympathetic nervous system activity and together outline a balanced and controlled behaviour of the two branches of the autonomic nervous system.

The last indicator is a *geometric* one, the HRV triangular index. The major advantage of geometric methods lies in their relative insensitivity to the analytical quality of the series of NN intervals. Geometrical indices are calculated on the sample density distribution of the NN intervals, which corresponds to assignment of the number of equally long RR intervals to each value of their length. The sample density distribution (histogram) presents discontinuities, and is then reconstructed by smoothing the curve using a moving average function. The HRV triangular index is calculated as the integral of the density distribution divided by the maximum of the density distribution. The histogram can be interpolated as a triangle, using the minimum square difference. The triangular interpolation of the RR interval histogram is the baseline width of this triangle [1, 48].

5.4 Participant cluster analysis

Given the response vector $R = \{r_n\}_{n=1}^N$, the standard EM algorithm was used to learn the set of response parameters $\Theta_R = \{\mu_k, \Lambda_k\}_{k=1}^K$. Since here we are committed to a data-driven analysis, prior probability for sampling from each category $\pi = \{\pi_k\}_{k=1}^K$ was initialised to a uniform distribution (which amounts to sampling $\pi \sim P(\pi)$, $P(\pi)$ being the conjugate Dirichlet prior with equal hyperparameters). The number of iterations of the EM steps was not fixed *a priori*; iterations were performed until convergence.

Here the most interesting aspect for the purposes of this work is the model selection problem, namely, determining the number of Gaussian components (i.e., participant clusters) K .

To such end, rather than resorting to any of the automatic model selection criteria adopted in the literature [16, 47] that are agnostic with respect to cluster semantics, we opted to perform a joint analysis considering both a measure of cluster quality and participants' self-evaluation reports.

A popular measure of cluster quality is the average silhouette [39], which is computed as follows: define the silhouette of element j as $(b_j - a_j) / \max(a_j, b_j)$, where a_j is the average distance of element j from other elements of its cluster, b_{jk} is the average distance of element j from the members of cluster C_k , and $b_j = \min_{k: j \notin C_k} b_{jk}$. The average silhouette is the mean of this ratio over all elements (quality ranging between -1 and $+1$).

The index is computed on clustering results considering all descriptors (EDA+HRV), EDA-only descriptors and HRV-only descriptors, for $K = 2, 3, 4$ components/clusters. Results (see Fig. 4) show a moderate quality with the EDA+HRV and EDA descriptors favouring the $K = 2$ model, whilst the HRV only setting favours the $K = 3$ one. For assessing these proposals and to give the obtained clusters a semantics, participants belonging to each cluster were compared on the basis of their self-evaluation and the behaviour exhibited in the recorded video. Results can be summarised as follows.

- EDA-only setting. Subjects in one cluster reported a medium-low level of arousal, a wide range of valence from neutral to positive and a variety of movement behaviours; subjects assigned to the other cluster reported a high level of arousal (> 0.8 in the range $[0,1]$), neutral to negative valence, similarities in moving and gazing behavior in the virtual environment. Thus, the latter cluster can be conceived as the “fearful” cluster, the first one can be labelled as a generic “non-fearful” subject group.

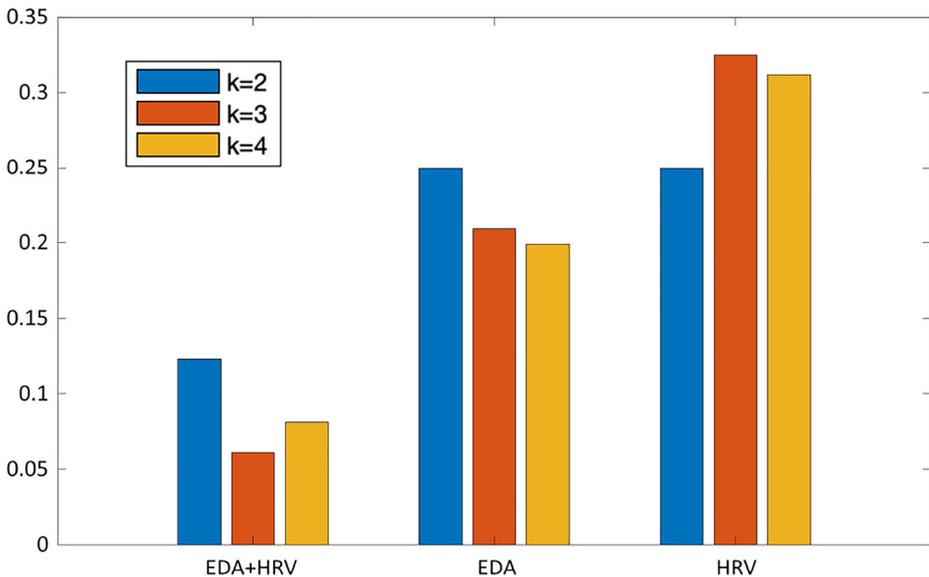


Fig. 4 Average silhouette index [39] computed on results of clustering by using all features (EDA+HRV), EDA-only features and HRV-only features only for $K = 2, 3, 4$ components/clusters

- HRV-only setting. In this case, the mapping between the $K = 3$ clustering results and self-reported affect state, and thus emotion category, is not readily obvious as in the previous case.
- EDA+HRV setting. This case is similar to the EDA-only case, but with a higher variability within clusters and resemblance between clusters as witnessed by the quality index, which is likely to be introduced by the HRV descriptors.

We then considered whether a feature selection step could further improve on the above results. A feature selection technique was thus applied to identify the best features to be used in the clustering step. This is a standard approach in machine learning [62].

We adopted the Spectral Feature Selection (SPEC) algorithm. SPEC estimates a descriptor relevance based on its consistency with the spectrum of a similarity matrix calculated using Radial Basis Function (RBF) [21]. In simpler words, the more relevant descriptors are those exhibiting a uniform behavior between clusters. The number of descriptors was reduced from 30 to 10.

The descriptors for the EDA-only setting were selected as the SC mean, SC kurtosis, SCR peaks number, SCR mean amplitude, SCR total and mean raise time, r_{AUC} , r_{Ax} , r_{Cx} .

The descriptors for the HRV-only setting were determined as the HRV standard deviation, SDNN, NN range, SDDSD, RMSSD, NN50, PSD total power, LF and HF variance and the HRV triangular index.

As to the EDA+HRV setting, the ten identified descriptors were: SC kurtosis, SCR mean signal raise time, standard deviation of HRV, NN range, SDNN, RMSSD, NN50, PSD total power, LF variance and the HRV triangular index.

Clustering learning and evaluation was repeated as in the previous analysis. The overall pattern did not change, except for a slight quality increase in clusters and computational efficiency.

As a final, complementary check, a one-way MANOVA test (category/cluster as the independent variable) was applied [34] on results obtained from all the configurations described so far: complete descriptor set and feature selected descriptors; HRV+EDA, EDA, HRV combinations; $K = 2, 3, 4$ models. MANOVA test rejected the null hypothesis of the clusters being similar in the sole case of $K = 2$ obtained with EDA parameters ($F(9, 17) = 2.6384, p < 0.05, \text{Wilk's lambda} = 0.524$).

To provide the reader an intuitive picture of results achieved, we display in Fig. 5 the scatter plot matrix to visualize bivariate relationships between pairs of features. The figure reports scatter plots of clusters distribution for the EDA indicators, best indicators being those separating the clusters in an optimal way i.e., with no overlapping in the scatter plot.

To summarise, in this final setting, with respect to the 26 valid tests, 8 subjects have been assigned to cluster 1 (unsuited for working at height) and 18 to cluster 2 (suited for working at height).

By and large, all participants assigned to cluster 1 reported a very high level of arousal (> 0.8 in the range $[0, 1]$). When considering, collected raw data (sensor), the EDA values have been steadily increasing during the simulation for all subjects within this cluster.

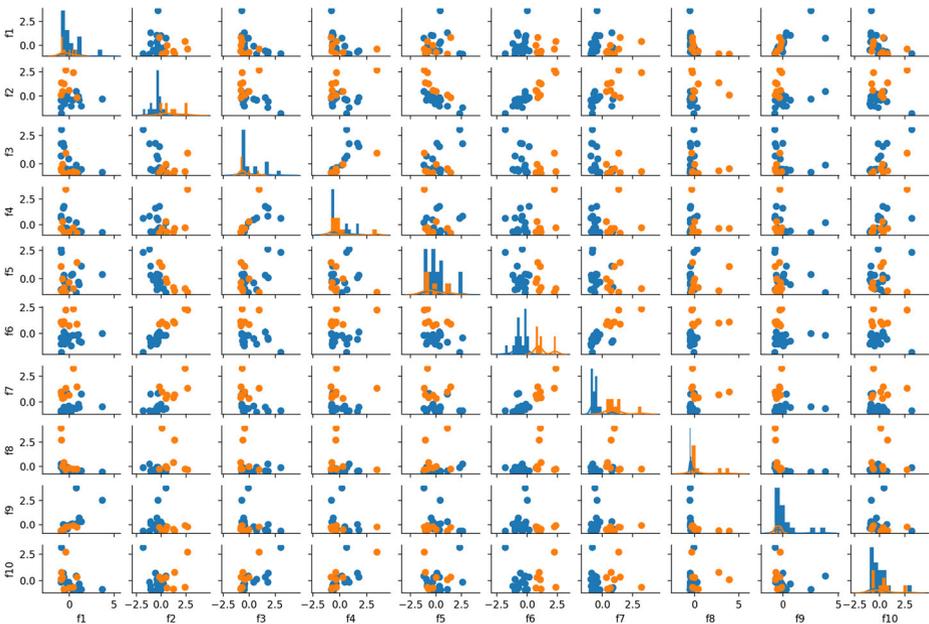


Fig. 5 Scatter plot matrix for EDA indicators. Cluster 1 is in blue while cluster 2 is in orange. The diagonal, where only one descriptor is available, shows the univariate distribution of the clusters for that descriptor. Best EDA indicators are the number 2, 6, 7, and 8 which are SCR peaks number, SCR total raise time, and the Hjorth parameters r_{A_x} and r_{C_x} , respectively

By examining the videos, similarities could be observed concerning the way participants moved on the plank and how they looked around in the virtual environment. In particular, two subjects exposed visible traits of fear and confusion during the simulation.

To complete the picture, when considering cardiac activity of individuals within this group, we can observe two distinct behaviors: for some subjects the heart rate raised sharply and stood there for the whole time, for the others it decreased smoothly during the simulation. Qualitatively, to an external observer, subjects reporting a sharp increase in the heart rate presented also a bit more clearly their fear in the recording.

This nuanced pattern concerning heart rate, is consistent with results reported in the psychological literature [2, 9, 17, 26, 27, 40, 50], while it is likely to be the reason for HRV indicators characterizing as less effective for straightforward clustering aims. We surmise that, grounded in the general framework we have outlined, this behaviour might be the result of how each individual is mentally representing and coping with the category of fear(s).

Eventually, all subjects within cluster 2 (suited for working at height) reported a medium-low level of arousal, varying valence state dynamics and quite steady physiological data.

6 Discussion and conclusive remarks

The goal of the work presented here was to evaluate VR as an avenue for assessing a person's suitability to work at height. Our main contributions to such research problem can be summarised as follows

- The approach is set out in the principled framework of the theory of constructed emotions. This allows a clear identification of the different levels of analyses (emotion, affect, physiology) for either elicitation and measurement. To the best of our knowledge this offers a novel perspective to the affective computing in general, and for the specific problem we addressed.
- Based on such framework and by taking stock of previous work in the field of perceptual categorization, we have formalized, from first principles, a suitable probabilistic generative model, which under adequate assumptions can provide a straightforward model for analysing data via unsupervised learning technique
- Under the above circumstances and by drawing on a preliminary analysis reported in [19], we have shown i) how VR is a viable technology to elicit and contextualize fear of heights at the emotion level, by exposing the trainee to a potential harmful situation without recreating a real dangerous situation, and that ii) physiological measurements and core affect evaluation can be soundly exploited to early characterize a worker's suitability to operate at altitudes

More specifically, at the experimental level, data were collected in a simple setup, then automatically clustered. Clusters were assessed based on cluster quality and affective semantics derived from subjects' self-evaluation reports. The partition in two clusters seems to be confirming the goal of this research. It is actually possible to divide the subjects at test into two categories: suited and unsuited to work at height.

Second, we found that for this specific purpose, the electrodermal activity is likely to be a better indicator compared to heart rate. As a side result, our analysis is also hinting to the fact that four features of the EDA (peaks number, total signal raise time, and the H_jorth parameters for activity and complexity) could be enough to identify the two groups.

It is worth remarking that the two group outcome is indeed a significant result, since in principle, for the specific purpose of this work, one is expected to deal with subjects

enacting two categories (fearful, not fearful). This is not a foregone conclusion at the light of the theoretical construct we have discussed and, clearly must be carefully handled. Fear is a response to threatening situations and it is typically considered aversive, which suggests that we should avoid fear. However, as previously mentioned, fear can partition in different types of fear, with either different or overlapping autonomic responses. For instance, in a VR context, pleasurable fear might be “sampled” by a subject. Pleasurable fear refers to cases where people seek out frightening experiences and take pleasure in them: participating in extreme sports, watching horror movies, and going to haunted houses [2, 17]. In such cases, subjects are motivated to experience fear, in a safe context. It has been surmised that these situations are likely to be felt by subjects as a form of “threat simulation” [2], that they might have to deal with in the future. Though, not everyone enjoys being afraid by engaging in this kind of aversive play.

Notwithstanding the evidence so far achieved, the study presented here has some limitations. First of all, in our experiment we are not dealing with a large sample size. This could be expanded in future work. Secondly, it is assumed that in the very specific context we are operating and the kind of constrained environment the participants are experiencing, a link can be established between their affect dynamics and an emotional conceptualization enacting their interoceptive and exteroceptive sensations, at least for the fearful condition. However, as we previously discussed, even an apparently “basic” emotion as fear can be subtle to interpret (e.g., [2, 17]). This aspect has been highlighted when also HRV parameters have been considered in the analyses. Although VR is very useful in realistic inducement of emotional states, a generalization of the results of the present work should be undertaken with caution.

For instance, in this study the interplay between anxiety (occurring when approaching the threat) and fear (occurring in presence of threat) is not taken into account. It is often assumed in the literature a sharp differentiation between anxiety and fear, though it has been shown in VR conditions that anxiety can be a factor of threat magnification in the fear of heights [41]. Also in a more general setting or research perspective, relations between perceived stress and primary stress responses, among which fear (intended as an experienced threat to self) has been surmised to have a part. Indeed, during a stressful experience, negative emotion could take the form of anxiety, displeasure, fear, anger, sadness, or a combination thereof [42]. These analyses would entail a finer assessment of subjects’ traits and involvement beyond the self-assessment evaluation (e.g., using questionnaires of Approach and Avoidance Motivation, Pain Catastrophizing Scale, and so on), which is admittedly out of the straight scope of this work, but however it might be adopted in the form of a follow-up, second step screening.

The unsupervised learning method we have exploited relies on the assumption of independent data observations. For the specific purpose of the present work, this “batch processing” is reasonable to produce a final judgement. Yet, the actual raw data collected (and the valence/arousal self evaluations) are time series and time information could be exploited. The GMM model can be easily generalized to this case (most straightforward and simple example being the Hidden Markov Model with Gaussian observations [16]). However, this would entail a more subtle interpretation, at the category latent variable level, of mental state dynamics.

Here, the affect self-evaluation is principally exploited for assessing the clusters and their most plausible number K . Alternatively, evaluations themselves could be embedded in model construction. It has been shown that it is possible to build a latent affect space on this basis, whose dynamics well correlates with valence/arousal trajectories [18]. This way

the data could be effectively handled as time series (but see the remarks above). However, that method was based on probabilistic deep generative modelling, which in our case would require a much larger subject sample than that we have currently employed.

Eventually, on a more practical stance, the experimental setup could be extended by considering additional physiological modalities, as body temperature and respiratory rhythm, and/or automated analysis of subjects' movements and action behavior from the recorded videos, which is affordable with current computer vision techniques [28, 65].

Funding Open access funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement. This project has not been funded by any institution, public or private.

Declarations With respect to Compliance with Ethical Standards, we declare the following.

Ethics approval and consent to participate An informed consent has been signed by all participants at the time of the experiments (Summer 2019). In particular, with respect to Informed consent, following the rules of our institution related to the experiments with humans, due to the fact that the research did not include any implications in public health or general welfare, no reference number is assigned, and only the described informed consent procedures were required. The experiment participants were all volunteers, adults, consenting, without any physical or cognitive disease, and they are not identifiable with sensitive ages (the participants' age is between 18 to 40 years old). Lastly, their data are anonymized and the experimental videos are not publicly available to the scientific community.

Competing interests The submitted work is the first in this research line for our lab. As such, there are no interests that are directly or indirectly related to the work submitted for publication within the last 3 years.

Conflict of Interests There are no conflicts of interests in the present work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Acharya UR, Joseph KP, Kannathal N, Lim CM, Suri JS (2006) Heart rate variability: a review. *Med Biol Eng Comput* 44(12):1031–1051
2. Andersen MM, Schjoedt U, Price H, Rosas FE, Scrivner C, Clasen M (2020) Playing with fear: a field study in recreational horror. *Psychol Sci* 31(12):1497–1510
3. Anderson JR (1991) The adaptive nature of human categorization. *Psychol Rev* 98(3):409–429
4. Azari B, Westlin C, Satpute AB, Hutchinson JB, Kragel PA, Hoemann K, Khan Z, Wormwood JB, Quigley KS, Erdogmus D (2020) Comparing supervised and unsupervised approaches to emotion categorization in the human brain, body, and subjective experience. *Sci Rep* 10(1):1–17
5. Bach DR, Friston KJ, Dolan RJ (2010) Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *Int J Psychophysiol* 76(1):52–55
6. Baker C, Pawling R, Fairclough S (2020) Assessment of threat and negativity bias in virtual reality. *Sci Rep* 10:17338
7. Baños RM, Botella C, Liaño V, Guerrero B, Rey B, Alcañiz M (2004) Sense of presence in emotional virtual environments. In: 7th annual international workshop on presence, pp 156–159
8. Barlow DH (2004) *Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic*. Guilford Press, New York

9. Berntson GG, Thomas Bigger Jr J, Eckberg DL, Grossman P, Kaufmann PG, Malik M, Nagaraja HN, Porges SW, Saul JP, Stone PH (1997) Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology* 34(6):623–648
10. Barrett LF (2016) Navigating the science of emotion. In: Meiselman HL (ed) *Measurement emotion*. Woodhead Publishing, Sawston, UK, pp 31–63
11. Barrett LF (2017) Categories and their role in the science of emotion. *Psychol Inq* 28(1):20–26
12. Barrett LF (2017) The theory of constructed emotion: An active inference account of interoception and categorization. *Soc Cogn Affect Neurosci* 12(1):1–23
13. Barrett LF, Bliss-Moreau E (2009) Affect as a psychological primitive. *Adv Exp Soc Psychol* 41:167–218
14. Barrett LF, Wilson-Mendenhall CD, Barsalou LW (2015) The conceptual act theory: a roadmap. In: Barrett LF, Russell JA (eds) *The psychological construction of emotion*. The Guilford Press, New York, pp 83–110
15. Barrett LF, Satpute AB (2019) Historical pitfalls and new directions in the neuroscience of emotion. *Neurosci Lett* 693:9–18
16. Bishop CM (2006) *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, Secaucus
17. Bloom P (2021) The paradox of pleasurable fear. *Trends Cogn Sci* 25(2):93–94
18. Boccignone G, Conte D, Cuculo V, D'Amelio A, Grossi G, Lanzarotti R (2018) Deep construction of an affective latent space via multimodal enactment. *IEEE Trans Cogn Develop Syst* 10(4):865–880
19. Boccignone G, Gadia D, Maggiorini D, Ripamonti LA, Tosto V (2021) Predicting real fear of heights using virtual reality. In: *Proceedings of the conference on information technology for social good. GoodIT '21*, Association for Computing Machinery, New York, pp 103–108,
20. Boccignone G, Conte D, Cuculo V, Lanzarotti R (2017) Amhuse: a multimodal dataset for humour sensing. In: *Proceedings of the 19th ACM international conference on multimodal interaction*, pp 438–445
21. Buhmann MD (2000) Radial basis functions. *Acta numerica* 9:1–38
22. Cacioppo JT, Tassinary LG (1990) Inferring psychological significance from physiological signals. *Am Psychol* 45(1):16
23. Chamilothori K, Chinazzo G, de Matos R, João P, Dan-Glauser E, Wienold J, Andersen M (2018) Perceived interest and heart rate response to façade and daylight patterns in virtual reality. Technical report. Academy of Neuroscience for Architecture
24. Critchley HD, Harrison NA (2013) Visceral influences on brain and behavior. *Neuron* 77(4):624–638
25. Diemer J, Mühlberger A, Pauli P, Zwanzger P (2014) Virtual reality exposure in anxiety disorders: Impact on psychophysiological reactivity. *World J Biol Psychiatry* 15(6):427–442
26. Diemer J, Lohkamp N, Mühlberger A, Zwanzger P (2016) Fear and physiological arousal during a virtual height challenge—effects in patients with acrophobia and healthy controls. *J Anxiety Disord* 37:30–39
27. Diemer J, Alpers GW, Peperkor HM, Shibani Y, Mühlberger A (2015) The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Front Psychol* 6:26
28. D'Mello SK, Kory J (2015) A review and meta-analysis of multimodal affect detection systems. *ACM Comput Surv* 47(3):43
29. D'Mello S, Kappas A, Gratch J (2018) The affective computing approach to affect measurement. *Emot Rev* 10(2):174–183
30. Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. *J Personal Soc Psychol* 17(2):124
31. Ekman P, Sorenson ER, Friesen WV (1969) Pan-cultural elements in facial displays of emotion. *Science* 164(3875):86–88
32. EuroSafe (2016) *Injuries in the european union, summary on injury statistics 2012–2014*, 6th edn. EuroSafe, Amsterdam
33. Feldman NH, Griffiths TL, Morgan JL (2009) The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychol Rev* 116(4):752
34. Gat-Viks I, Sharan R, Shamir R (2003) Scoring clustering solutions by their biological relevance. *Bioinformatics* 19(18):2381–2389
35. Greco A, Valenzao G, Scilingo EP (2016) Modeling for the analysis of the eda. In: *Advances in electrodermal activity processing with applications for mental health*. Springer, New York, pp 19–33
36. Gross CT, Canteras NS (2012) The many paths to fear. *Nat Rev Neurosci* 13(9):651–658
37. Izard CE (1993) Four systems for emotion activation: cognitive and noncognitive processes. *Psychol Rev* 100(1):68
38. Jönsson P, Wallergård M, Österberg K, Hansen ÅM, Johansson G, Karlson B (2010) Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the trier social stress test: a pilot study. *Psychoneuroendocrinology* 35(9):1397–1403

39. Kaufman L, Rousseeuw PJ (2009) *Finding Groups in Data: An Introduction to Cluster Analysis*, vol 334. Wiley, Hoboken
40. Kreibig SD (2010) Autonomic nervous system activity in emotion: a review. *Biol Psychol* 84(3):394–421
41. Krupić D, Žuro B, Corr PJ (2021) Anxiety and threat magnification in subjective and physiological responses of fear of heights induced by virtual reality. *Personal Individ Differ* 169:109720
42. Lebois LA, Hertzog C, Slavich GM, Barrett LF, Barsalou LW (2016) Establishing the situated features associated with perceived stress. *Acta Psychol* 169:119–132
43. LeDoux JE (2014) Coming to terms with fear. *Proc Natl Acad Sci* 111(8):2871–2878
44. Liao D, Zhang W, Liang G, Li Y, Xie J, Zhu L, Xu X, Shu L (2019) Arousal evaluation of vr affective scenes based on hr and sam. In: *MTT-S International microwave biomedical conference (IMBioc)*, vol 1. IEEE, Piscataway, pp 1–4
45. Lindquist KA, Barrett LF (2008) Constructing emotion: The experience of fear as a conceptual act. *Psychol Sci* 19(9):898–903
46. Ma H, Yarosh S (2021) A review of affective computing research based on function-component-representation framework. In: *IEEE Trans Affect Comput*, p 1
47. MacKay DJC (2004) *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge, UK
48. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, Schwartz PJ (1996) Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Eur Heart J* 17(3):354–381
49. Martens M, Antley A, Freeman D, Slater M, Harrison PJ, Tunbridge E (2019) It feels real: Physiological responses to a stressful virtual reality environment and its impact on working memory. *J Psychopharmacol* 33(10):1264–1273
50. Norman GJ, Berntson GG, Cacioppo JT (2014) Emotion, somatovisceral afference, and autonomic regulation. *Emot Rev* 6(2):113–123
51. Panksepp J (2004) *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press, New York
52. Posada-Quintero HF, Florian JP, Orjuela-Cañón AD, Aljama-Corrales T, Charleston-Villalobos S, Chon KH (2016) Power spectral density analysis of electrodermal activity for sympathetic function assessment. *Ann Biomed Eng* 44(10):3124–3135
53. Quadrio G, Bujari A, Gaggi O, Palazzi CE (2019) New interactive vr/ar media: A challenge worth to address. In: *Proceedings of the 5th EAI International conference on smart objects and technologies for social good*, pp 128–129
54. Quigley KS, Lindquist KA, Barrett LF (2013) Inducing and measuring emotion and affect: tips, tricks, and secrets. In: Reis HT, Judd CM (eds) *Handbook of research methods in social and personality psychology*. Cambridge University Press, New York, pp 220–252.
55. Riva G, Mantovani F, Capideville CS, Preziosa A, Morganti F, Villani D, Gaggioli A, Botella C, Alcañiz M (2007) Affective interactions using virtual reality: the link between presence and emotions. *Cyberpsychol Behav* 10(1):45–56
56. Russell JA (2003) Core affect and the psychological construction of emotion. *Psychol Rev* 110(1):145
57. Sanchez-Vives MV, Slater M (2005) From presence to consciousness through virtual reality. *Nat Rev Neurosci* 6(4):332–339
58. Schulkin J, Sterling P (2019) Allostasis: A brain-centered, predictive mode of physiological regulation. *Trends Neurosci* 42(10):740–752
59. Schuller BW, Picard R, André E, Gratch J, Tao J (2021) Intelligent signal processing for affective computing [from the guest editors]. *IEEE Signal Proc Mag* 38(6):9–11
60. Seth AK, Suzuki K, Critchley HD (2012) An interoceptive predictive coding model of conscious presence. *Front Psychol* 2:395
61. Shukla J, Barreda-Angeles M, Oliver J, Nandi G, Puig D (2019) Feature extraction and selection for emotion recognition from electrodermal activity. *IEEE Trans Affect Comput* 12(4):857–869
62. Tang J, Alelyani S, Liu H (2014) Feature selection for classification: a review. *Data classification: Algorithms and Applications*:37
63. Tomkins SS (1962) *Affect, Imagery, Consciousness*. Springer, New York
64. Valenza G, Lanata A, Scilingo EP (2011) The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE Trans Affect Comput* 3(2):237–249
65. Wang Z, Ho S-B, Cambria E (2020) A review of emotion sensing: Categorization models and algorithms. *Multimed Tools Appl* 79(47):35553–35582