Imperial College London
Department of Civil and Environmental Engineering

A Generalised Methodology for the Investigation of Human Behavioural Responses to Hostile Attacks

Alastair Shipman

February 2021

Supervised by Dr Arnab Majumdar

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Civil and Environmental Engineering of Imperial
College London
and the Diploma of Imperial College London

Declaration

I herewith certify that all material in this dissertation which is not my own work has been properly acknowledged.

Alastair Shipman

Acknowledgements

I have greatly enjoyed doing this PhD, and that is probably mainly because of the support I had surrounding me. I would primarily like to thank my lovely wife, Hannah, without whom I wouldn't have started, continued, or finished. You are quite simply my favourite person- I often don't know why you put up with me, but I am very glad that you do!

To Dr Arnab Majumdar, and all the members of my research group, thank you! It has been a hard slog at times, but having a team around you that knows what you are going through, and who can, crucially, answer your questions, is so helpful. I particularly am grateful for Georgia Bateman, as we started this work together, and she has helped me throughout!

A great many more thanks are owed, including to all members of the Sussex School, where I had some of my best ever times. My family have consistently been a source of peace and inspiration throughout. From Sunday evenings at Abdale Road, to being a monster for Lara and Jesse, or sleeping off the thesis stress in Walthamstow, relaxing with you all has been the best thing. Thanks too, to Aidan and Mum, for reading through the awful first drafts-I hope this one is better!

Throughout the this PhD, I've relied on the help and support of many people. To Dr Ruggiero Lovreglio, I will be forever in your debt for your support and guidance. To all members of the Lorentz Center Group, including Dr Steve Gwynne, Dr Aoife Hunt and Dr Guillaume Dezecache, to Dr Angelika Kneidl and the group at Accu:Rate, to Dr Holly Carter and the incredible people at Public Health England - thank you all!

Finally, to my experimental team, including my actors Dr Jack Herlihy and James Phillips, for being amazing, to Ellie Jay, for letting me sleep on her floor, and everyone in Skempton Building who put up with my no doubt irritating questions!

Abstract

Terrorist attacks have become more frequent and more deadly in recent decades. Understanding the movement responses of the individuals involved can save lives.

Pedestrian dynamics is an area of research and industry that investigates the movement of people. Using pedestrian dynamics to predict the responses to terrorist attacks could be hugely beneficial to the survival of those involved.

This PhD developed a generalised methodology for investigating movement responses to terrorist attacks, accommodating the wide range of attack vectors seen in modern terrorism and providing a detailed summary of the data that should be obtained.

Initial work analysed the overlap between pedestrian dynamics and terrorism. This concluded that current interdisciplinary research is not sufficiently advanced to allow for accurate predictions of pedestrian movement responses to terrorist attacks, due to a lack of usable data given the sensitivity of previous incidents and limitations on experimental methodologies.

Two sets of experiments were then performed, examining responses to knife-based attacks by an unexpected and hostile individual. These experiments were split into physical reality (PR) and virtual reality (VR). The PR experiments were performed in December 2018 with 80 participants, and the VR experiments in January-February 2020 with 55 participants.

The experiments obtained positional, physiological, psychological and memory-related measurements. Analysis of these datasets investigated the participant responses to hostile aggressors, including the impact of demographic, emotional state, and the similarity in responses between the VR and PR paradigms.

The analysis on the psychological and physiological datasets proved a stressed response to the hostile individual, resulting in the first ever experimental dataset for these. The subsequent analysis on the positional and memory-related datasets provided key inputs for pedestrian dynamics models and con-

cluded that VR is appropriate as a data-generating tool for human behaviour in emergencies.

Contents

List of Tables

List of Figures

1. Introduction

1.1. Terrorism

At 9.15pm on Friday 13th November 2015, seven terrorists began a coordinated attack in Paris, France. There were incidents in six different locations, including four suicide bombings and four separate hostile shooting events, one of which evolved into a hostage situation. Over a four-hour period 130 civilians lost their lives, and more than 400 were injured.

At 10.06pm on Saturday 3rd June 2017, three terrorists began an attack in London, UK. They drove a car into pedestrians on London Bridge, before running into Borough Market on foot, attacking individuals with knives. Over a ten-minute period eight civilians lost their lives, and 48 more were injured.

Figure 1.1.: Terrorist attacks in London (l) and Paris (r)

Both of the above events shocked the world, showing how quickly it was possible for everyday activities to turn into a life-or-death scenario. The Global Terrorism Database [11] tracks the statistics surrounding terrorist attacks worldwide, showing that the number, lethality and injury rate in attacks has been increasing over the last two decades. Unfortunately, events like these have become far too common.

Traditionally it has been assumed that people will panic during these scenarios, where 'panic' is defined [3] as: irrational, competitive and non-adaptive behaviour. However, it has been mostly observed [12] that individuals and crowds do not panic, but instead act pro-socially. Interestingly, after both events newspapers reported examples of unexpected behaviour, where members of the public acted calmly and altruistically. In the Bataclan theatre in Paris, members of the audience queued patiently as they tried to hide in the attic [13]. In Borough Market, Ignacio Echeverría lost his life attempting to stop the knifemen, saving several members of the public in doing so [14]. Unfortunately, while social norms like queuing seem to be maintained to a certain extent, it remains unknown what an individual will do in a given particular context. Some people will run, some will hide, some fight or freeze, while others might do something completely different. There is little data and still less understanding surrounding these responses, and therefore little ability to predict them. Importantly, while the behaviour of civilians during these events often had a notable effect on their survival rates, little else is known about the typical responses that are seen in terrorist attacks. This lack of understanding is a serious impediment to the mitigation of these events.

1.2. Modern macroscopic engineering and pedestrian dynamics

Modern large-scale engineering projects are increasingly focused on human-centric design principles, with examples ranging from transport provision to large-scale buildings all being designed to provide any occupants with the most intuitive and efficient experience possible. Pedestrian dynamics is a modern "smart-city" example of this approach, where engineers (both practitioners and researchers) seek to quantify human movement through infrastructure within specific contexts (see Chapter 2 for further details). Standard scenarios examined by pedestrian dynamics engineers include:

- Quantifying maximum commuter flow rates through a train station [15]
- Assessing pedestrian exposure to artworks in modern galleries [16]
- Predicting typical flow through large infrastructure [17]
- Understanding the impact of additional seating provision in stadia [18]
- Crowd management for high-density events [19]

Pedestrian dynamics is also used for emergency scenarios, from local-scale modelling used to improve the safe egress of building occupants (e.g.: [20]), to macro-scale modelling, used to understand systematic issues with movement of large numbers of people across multiple transport modes, as for example in response to hurricane evacuations ([21]).

However, it has been noted that there remain many unsolved problems with incorporating human user responses into the design process for large-scale infrastructure [22]. Major issues include the definition and quantification of these interactions, as well as the validation of any models used to predict these. The ability to use computational models to predict likely outcomes in emergency situations will provide a powerful tool that can potentially save lives.

Approaching this problem requires a large amount of raw data on observed interactions, which is expensive to obtain, and an assessment of the predictors for these interactions. Previous research that sought to approach this problem have included physical experiments [23], virtual reality experiments [24], and analysis of real-world data [25]. However, these approaches suffered from many limitations, as any real-world data is uncontrolled and often incomplete, and any data from experimental approaches can lack ecological validity. Furthermore, the data surrounding human responses to emergency situations is rarer still, and often sensitive in nature. Without a sufficient source of valid data, any approach to solving this problem will be unproven at best, and misleading at worst.

1.3. Pedestrian dynamics and emergencies

The nature of emergencies is varied, with examples ranging from natural disasters like hurricanes and wildfires, to man-made disasters like infrastructure collapse, to hostile events such as marauding-terrorist-firearm-attacks (MTFAs). Further to this, behavioural responses to individual emergencies depend on the type of emergency itself, with reactions to a wildfire being fundamentally different to those of a terrorist attack.

However, all of these scenarios combine unique human behavioural responses with a high threat to life and possibility of harm for all involved, and therefore represent a large opportunity for the use of human-centric design principles (such as pedestrian dynamics approaches) that can mitigate their negative effects. In order to fully understand how individuals react, it is crucial to characterise the emergencies themselves and quantify any behavioural responses as well.

Day-to-day pedestrian dynamics models are based on data obtained from experiments, real-world measurements, drills, and many other methods. These models need to be calibrated with appropriate data that is tailored to the specific scenarios in question. For example, when engineers design the staircases for higher tiers in sports stadia, they should consider the mobility of the potential occupants. This mobility might be very different to, say, the occupants of an elderly care home, which would consequently change the staircase design.

It is clear, then, that when considering the types of hostile emergency that terrorist attacks represent, the datasets that are used to inform day-to-day behavioural models are inappropriate, as they do not encapsulate the range of responses that are observed, nor do they include the stressors that induce these responses. This means that the datasets used for any behavioural and simulation models may be contextually unique, thereby increasing the requirements for a fully validated model of human behaviour in emergencies. Therefore, in order to accommodate human responses to emergency incidents in designing infrastructure, it is first necessary to gather data on such scenarios, whether through real-world data or carefully designed experiments.

1.4. This PhD

This PhD research arose out of a need to start addressing the lack of a predictive capability for models of human behaviour in emergencies, with a specific focus on terrorist attacks. There has been a dearth of research investigating human behavioural responses to these type of incidents as they represent a particularly complex and difficult topic to study. However, terrorism is one of the highest priorities for the security and resilience of modern civilisation, having increased in both frequency and lethality over the past decades.

To expand on this point, a hypothetical example of a terrorist attack, shown below in Figure 1.2, will be described. In this example there are four pedestrian members of the public (in blue), and one hostile aggressor (in red) who intends to harm the pedestrians.

The pedestrians in this scenario are able to choose from multiple possible discrete response options. These might be:

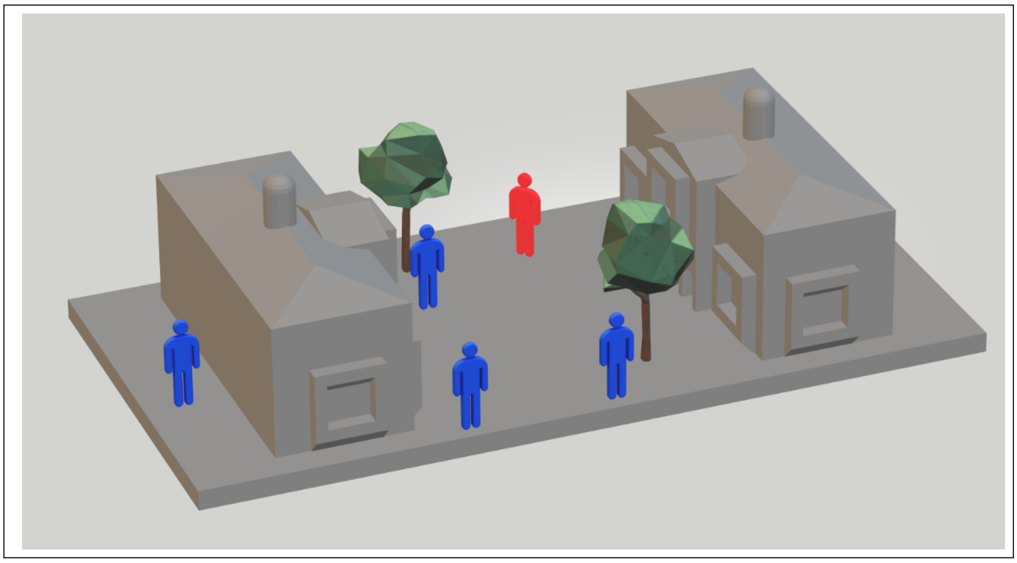


Figure 1.2.: Hypothetical Scenario

- Enter the buildings (hide)
- Evacuate (flight)
- Attempt to stop the aggressor (fight)
- Do nothing (freeze)

Within these options there are also possible variations and combinations of different choices. For example, a pedestrian might hide in the building, before evacuating at the earliest possible opportunity. Then, during the evacuation phase, this pedestrian might choose to hide again in a different location (e.g. behind a tree).

There are also uncertainties regarding the continuous movement responses of the pedestrians, and how they depend on the relative position of the aggressor. One option might be to assume radial movement away from the aggressor at the highest possible speed, or to prioritise movement towards places of safety. The presence of obstacles complicates this further, as these may represent potential areas of temporary safety.

A final consideration is the potential level of social interaction between pedestrians. In the above example there are three pedestrians who can see the aggressor. If one of these three decides to evacuate (flight), then it is likely that the remaining two will also choose to do so; this type of herding behaviour has been seen on multiple occasions during emergency scenarios. However, quantifying this probability is complex, and depends on multiple factors, including the level of trust and social identification among the three pedestrians. Extending this scenario even further, the fourth pedestrian (behind the building) would initially be unable to see the aggressor, nor the actions of the three pedestrians. There is no current quantitative understanding of the impact of any non-visual stimuli on this pedestrian, such as audible gunshots or raised voices. This hypothetical scenario will be revisited in Chapter 10.

One way of modelling this scenario would be to assume certain proportions of pedestrians choosing specific actions, and to predict an outcome from these assumed responses. However, without data to validate these assumptions, there is no guarantee that this modelling would be accurate. Without a developed theoretical model of how the pedestrians in this scenario will respond, it is impossible for engineers, designers, architects, crowd managers, or indeed anybody, to reliably put in place measures to mitigate the impact of the hostile aggressor. Therefore, without this developed theoretical model, based on valid data, the lives of the pedestrians in this scenario remain at risk.

To address this problem, this PhD research needed to understand how to produce valid and quantitative data on pedestrian responses to hostile individuals, before obtaining and analysing this data. Consequently, the research conducted in this PhD has led to the following key results:

- A generalised methodology for performing pedestrian dynamics experiments that can ethically produce the ecologically valid responses expected when investigating hostile emergency scenarios.
- A set of experiments that implement the above methodology, specifically looking at knife-based terrorist attacks.
- A comparison between the use of physical and virtual reality experimental paradigms for these experiments.
- A quantification of the reliability of recall of stressful scenarios for generating data and theory.

These results will allow future engineers to perform experiments to investigate specific types of terrorist attack, while also establishing the validity of the data generated.

1.5. Objectives

The objectives of this thesis are stated below.

Objective 1: Critically analyse the current state-of-the-art relating to pedestrian dynamics in hostile emergencies.

The first objective of this thesis is to explain the current use of, and potential demand for pedestrian dynamics in the case of hostile emergencies. This will inform the remainder of the research performed and will provide crucial insight into the state of the industry and academic research. This will be performed by:

- A critical review of the surrounding literature.
- Interviews with key industry professionals and academics.

The research conducted as part of these indicated both a lack of, and a demand for relevant data in this area. In consequence the remaining objectives focus on tackling this challenge and its implications.

Objective 2: Develop a robust methodology for experimental investigation of behavioural responses to hostile emergencies.

This thesis will describe a methodology that can be used to produce experiments that investigate human behavioural and movement responses to generic hostile emergencies. This methodology will be developed through several stages, namely:

- Ethical considerations
- Experimental design
  - Theoretical constructs
  - Stressor design
  - Measurements
  - Experimental paradigm
- Validation

In order to perform experimental procedures examining stressful events such as terrorist attacks, it is necessary to first understand the ethical considerations of such an undertaking. Secondly, the experimental design itself requires theoretical underpinnings, as well as well-defined stressors and measurements. This thesis will also investigate multiple paradigms for obtaining these datasets, before finally examining the validation of these experiments.

Objective 3: Detail the design and implementation of this methodology.

The next objective within this thesis is to describe the experiments that were as part of this PhD. These experiments were carried out in December 2018 and January-February 2020, producing the first ever set of open-source, valid and controlled results within this field. This included producing data in the following areas:

- Positional
- Psychological
- Physiological
- Recall

These datasets are required for the internal validation of the experimental procedures, as well as the generation of relevant data for pedestrian dynamics simulations. This PhD implemented two separate experimental paradigms, one physical and one virtual, not only generating greater amounts of data, but also allowing a comparison between the two paradigms.

This data will be placed into an open-source repository, allowing further analysis to be performed on these results. This represents the first set of open-source data on human behaviour in hostile-emergencies.

Objective 4: Detail appropriate datasets and resulting analysis.

This thesis will then describe the analysis performed on the results from these experiments, generating relevant results in many different areas that had previously been theoretical. These will include:

- Empirical analysis techniques
  - Emotional data
  - Psychological approaches

  - Physiological approaches
  - Trajectories
  - Recall quality
- Comparisons of different experimental paradigms

The initial analysis will assess the psychological and physiological data obtained, providing internal validation for the experiments. The next analysis will investigate the resulting positional data, using both discrete and continuous approaches. The final analysis will aim to quantitatively understand the reliability of interviews when gathering data from stressful events.

Throughout this analysis the thesis will compare the different experimental paradigms, providing insight into potential future avenues of research.

1.6. Thesis outline

The outline of the chapters in this thesis is shown below in Figure 1.3.

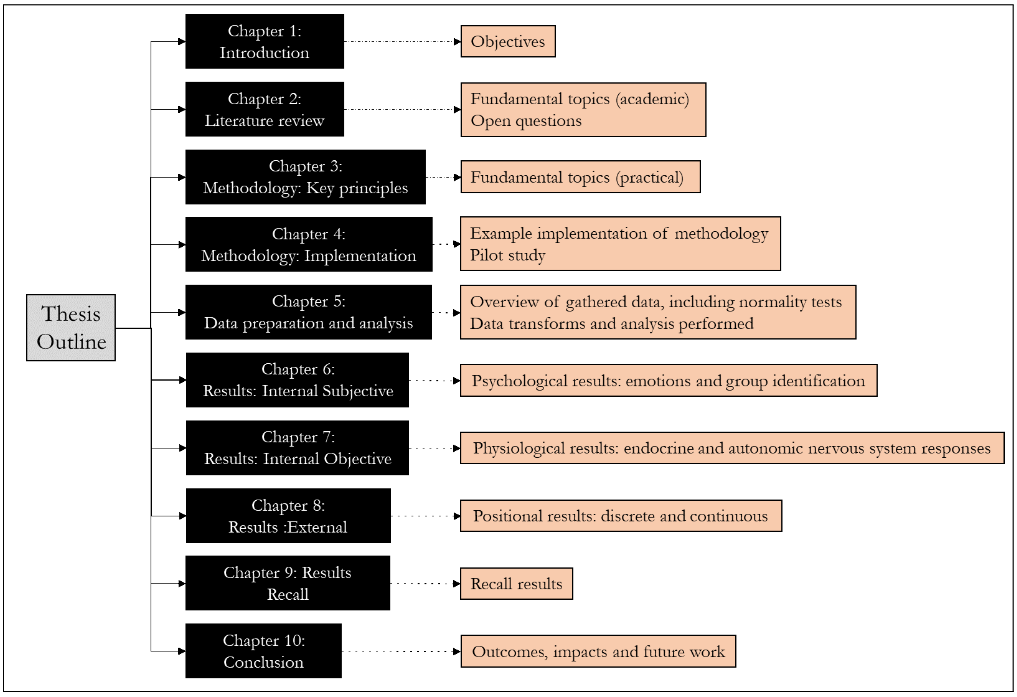


Figure 1.3.: Thesis outline

Chapter 2: Literature Review investigates the different fundamental topics of this type of research, specifically detailing them from an academic perspective. These include pedestrian dynamics, psychology and physiology, experimental design and terrorism. It will conclude by detailing a set of interviews that were performed with key stakeholders in industry, government and academia, investigating the current state of, and demand for, this research. **This chapter concludes Objective 1.**

Chapter 3: Methodology: Key Principles details the different practical aspects of designing and performing experiments that investigate human behavioural responses to terrorist attacks. Specifically, it will focus on the ethical considerations, including using pilot studies and presumptive consent, before detailing the requirements for stressor design. It will then describe the different potential measurements, separating these by experimental paradigm.

Chapter 4: Methodology: Implementation describes an example of the generalised methodology detailed in Chapter 3. It will focus on the implementation of two sets of experiments (one physical, one virtual) looking at behavioural responses to knife-based terrorist attacks. As part of this implementation, it will detail the pilot study and presumptive consenting procedure that was performed in conjunction with Public Health England, before defining the final experimental design. **This chapter concludes Objective 2.**

Chapter 5: Data Procedures, Preparation and Analysis will provide a view of the raw measurements obtained in both sets of experiments, and will perform the initial data preparation. This includes a summary of the obtained datasets, an assessment of the participant demographics, normality tests on the remaining data and a brief summary of the proposed analysis techniques. **This chapter concludes Objective 3.**

Chapter 6: Results: Internal Subjective focuses on analysing the psychological measurements taken as part of the experimental processes. These measurements quantify the subjective emotional state of all participants, as well as the level of group identification between participants, and are used to satisfy the internal validation of the experimental procedures.

This chapter shows that the experimental procedures successfully induced a level of negative emotion in the both the physical and virtual experimental groups that was not induced in the control groups.

Chapter 7: Results: Internal Objective will focus on the physiological measurements taken during the experimental processes. Initially this chapter will quantify a measure of the endocrine system responses of all participants, specifically the level of salivary cortisol and cortisone, both of which are correlated with levels of internal stress. Next, this chapter will use a set of autonomic nervous system measurements obtained from a subset of participants to assess a set of algorithms used to classify stress responses.

This chapter shows that there is a level of salivary cortisol and cortisone produced in the physical experiments that was not seen in the control groups. It also concludes that the autonomic nervous system responses can be used to determine stress states in participants, but also that experiments require further data for specific emotional state quantification.

The analysis shown in Chapters 6 and 7 conclude the internal validation of the experimental paradigms, showing that the experiments produced both a psychological and physiological stress reaction, which was mirrored in both physical and virtual paradigms.

Chapter 8: Results: External will detail and analyse the positional measurements taken from all participants throughout the experiments. As part of this analysis, the discrete choices will be analysed using social, demographic and stressor-related predictors. Further to this, the continuous movement responses of the participants will be analysed. Both sets of analysis are specifically designed to facilitate incorporation into pedestrian dynamics simulation models.

These responses were analysed to show that the most important predictor for an individual's discrete choices is their observation of other participants' social responses. Comparisons between virtual and physical paradigms show an almost identical set of predictors.

The analysis also showed the continuous responses of individuals, based on their relative position to the aggressor. The output from these models showed that participants both moved faster and accelerated more when nearer the aggressor, but that this movement was also asymmetric, with higher velocity and acceleration values being seen in front of the aggressor than behind. A further comparison was made between physical and virtual movement responses, with significant overlap observed.

Chapter 9: Results: Recall quantitatively examines the reliability of testimony when considering data gathering from emergency scenarios by detailing the results from a survey sent to all participants three months after their respective experiments, asking them to recall their experiences.

This chapter shows how participants' recall quality is dependent on the recall domain, as well as on stress level and experimental paradigm. For example, participants remembered information about their neighbours better during stressful conditions than in control conditions, and remembered their own behaviour better in physical experiments than in virtual experiments. **This chapter concludes Objective 4.**

Chapter 10: Conclusion and Further Work will summarise the research done throughout this PhD, detailing how this satisfies the initial objectives, as well as how this research can be extended to different scenarios.

2. Literature Review

The previous chapter introduced the basis of the PhD, providing an overview of the outstanding questions surrounding the analysis of human behaviour in hostile emergencies, before continuing to detail the approaches taken by this PhD and the results obtained. This chapter will review the academic research that underpins these approaches, contextualising the decisions made during the experimental design process. This literature review, outlined in Figure 2.1, was performed by first addressing referential texts (e.g.: [26, 27, 28, 29, 30, 2, 31, 32, 33]), and referential papers (e.g.: [34, 35]) . This was extended to searches of online repositories (e.g., Google Scholar, Web of Science), with key word searches (including ‘human behaviour’, ‘pedestrian dynamics’, ‘terrorism’, and ‘fear’). The literature review was also informed by interviews with experts in industry and academia, which augmented the literature review and established the boundaries of research in this topic. This is discussed in the final section of this chapter.

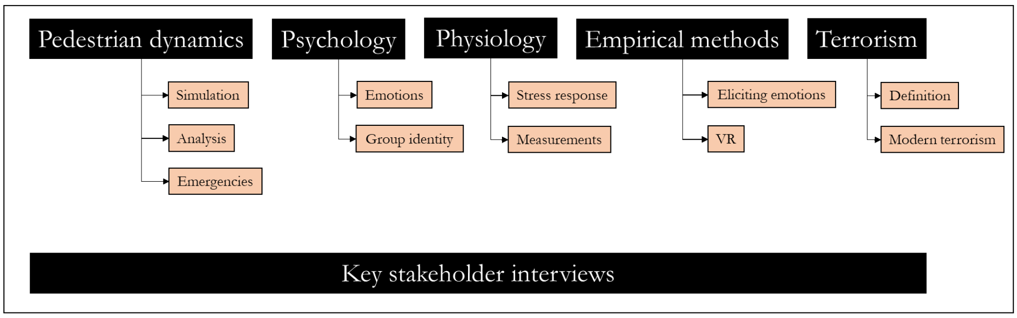


Figure 2.1.: Literature review outline

This chapter will be split into several sections. The first section will review the current state of pedestrian dynamics research, including theoretical, empirical, simulation-based, and experimental approaches that have been performed.

The next section will detail the psychological and physiological elements of responses to emergencies. This will initially focus on previous research surrounding group responses in emergencies, describing the historical context as well as modern research. This section will also detail the use of individual measurements to characterise emotional responses, including objective (e.g. heart rate) and subjective (e.g. surveys) measurements.

The third section will investigate previous experimental approaches, identifying appropriate measures that can be used for pedestrian dynamics experiments. This includes a discussion on the virtual reality (VR) paradigm, as well as other previous experimental approaches.

Fourthly this literature review will detail the type of emergencies that will be investigated within this PhD. Specifically, it will expand on the definition of terrorism, as well as the type of terrorist incident that will be examined within this PhD.

Finally, this chapter will summarise interviews that were conducted in 2016 with key stakeholders, detailing the current usage of these concepts within the resilience, crowd dynamics, local governance and security sectors, as well as investigating areas of desired progress. A thematic analysis is provided in this chapter, and full transcripts are available within Appendix B.

Throughout this chapter there will be reference to several terms, defined below. Where available, these definitions were taken from [36], identified by being placed in quotation marks.

2.1. Pedestrian dynamics

The modern study of pedestrian dynamics has existed since at least the 1950s [37], where the relationships between flow rate, density and individual speed are detailed. This work was then extended through the next few decades to include basic hand calculation methods ([38, 39]), as well as more complex theoretical and computational models (e.g.: [40]). The work performed by both Predtechensky and Milinksy ([41, 42]) is often referenced as some of the first examples of evacuation calculations.

With the advent of available computational power, large strides were made in the theory and simulation of the movement of people. For example, the social force model (which underpins a large amount of contemporary modelling) was established for pedestrian dynamics in the early 1990s [43]. Since then pedestrian dynamics has developed into a wide-reaching area of research, with

Term	Definition
Pedestrian dynamics	A term covering the academic research and industry applications of a range of areas including simulation, theory and experiments, which seek to characterise how people move.
Emergency	A scenario in which the lives, physical and/or mental health of the people involved are at risk.
Pedestrian	"Person moving on foot in a publicly accessible area. Subcategories include: pedestrian-commuter, pedestrian-shopper, pedestrian-traveller, (phases of waiting or pausing included)."
Group	"A group of people or items is a number of people or items which are connected or linked for some time by some measure (e.g. spatially or psychologically)." "In Self-categorization theory (psychological) terms, a group is people who think of themselves as a group (this is a theoretical definition)."
Emotion	A generalised term for behavioural, physiological and psychological responses to external stimuli. See Section 2.3 for further information. A subset of emotion involved in emergencies is stress, fear and anxiety.
Stress, Anxiety, Fear	These are related terms that are often used interchangeably, representing the emotional responses to physical and mental stressors. Stress is an overarching term that can relate to multiple different emotional subsets. Anxiety is an emotion where an individual experiences negative feelings towards a stimulus that is not directly present and possibly ill-defined, while fear is an emotion which is induced by stimuli that are physically present and well known for the individual. This PhD will measure these psychologically (through self-assessed measures) and physiologically (through autonomic nervous system and endocrine system measures). See Section 4.7 for further information.
Behaviour	"Behaviour is the response of individuals or groups to internal and external stimuli. Collective behaviours are responses of groups to internal and external stimuli that arise from interactions between group members." "Collective behaviour in a group/crowd arises when individuals respond not only to stimuli, but also to others (coordination) leading to the spatial or temporal synchronisation of group/crowd dynamics. In contrast, aggregate behaviour lacks synchronisation across individuals even if interactions between individuals may occur (e.g. to avoid collisions)."

Table 2.1.: Definitions used in this PhD

applications in architectural design, crowd management and building safety (e.g.: [15]). Industrial applications of pedestrian dynamics largely revolve around day-to-day interactions, as this area has the most supporting research and is therefore the most validated use case.

Modern research topics in pedestrian dynamics include investigating the links between group psychology and movement behaviour (for example, social and familial groups [44]) and edge cases (for example, inebriated behaviour [45]). Emergency behaviour is also an established area within pedestrian dynamics, but suffers from many limitations including a lack of validation, and difficulties with experimentation. This subset of pedestrian dynamics will be detailed below.

2.1.1. Computational approaches

There are many different methods of simulating pedestrian dynamics, all of which have their relative benefits and drawbacks [46]. It is useful to split these methods up by the scale of their modelling: microscopic, macroscopic and mesoscopic.

On a microscopic scale, simulation models aim to replicate interactions between individuals and their surroundings. This includes reproducing certain phenomena that are seen in real pedestrian flows, such as lane formation, the zipper effect, alternating flows at bottlenecks and arching at exits. There are also phenomena seen in evacuations and higher density flows. These effects (for example, the faster-is-slower effect) are explained in [34]. Examples of microscopic models include social force-based approaches, and agent-based modelling approaches. These are generally more computationally expensive than macroscopic models, with significant research efforts aiming to improve the efficiency and speed of simulations. Recent models have produced performance metrics where several 100,000s of individuals are modelled in real time [47]. A common microscopic method of analysis is agent-based modelling, including social force-based models, which will be described in detail below.

In contrast to these microscopic approaches, macroscopic models aim to replicate large scale movement behaviour, such as city-scale to nation-scale movement characteristics, and have applications in large scale evacuations [21] and pilgrimages [19]. Examples of macroscopic models include fluid dynamics approaches (e.g., [48, 49]), which provide an approximation of pedestrian flow for high density crowds on a macroscopic scale. The applicability of macroscopic approaches breaks down as density decreases, or as the scale of investigation reduces, and as the focus of the model moves towards microscopic (i.e. understanding the effect of individual pedestrians).

Mesoscopic models aim to bridge the gap between macroscopic and microscopic modelling scales, providing some of the benefits of both, while also suffering from some of the drawbacks. As this PhD focuses on microscopic movement, neither macroscopic or mesoscopic models will be considered further.

Two high profile examples of these computational approaches, the social force model and the optimal steps model are detailed below.

Social force model

The social force model is an example of an acceleration-based model, which approximates pedestrian movement as a Newtonian process. The interactions between pedestrians, obstacles and other objects are described by forces, which cause the pedestrian to accelerate in a particular direction. A simplified version of this is shown below (adapted from Pedestrian, Crowd and Evacuation dynamics [4]):

$$\frac{dV}{dt} = \frac{F(t) + \gamma(t)}{\tau} \tag{2.1}$$

where:

$$\begin{aligned}
F(t) &= F_{desired} + \Sigma F_{repulsive} + \Sigma F_{attractive} \\
\tau &= \text{relaxation factor} \\
\gamma(t) &= \text{random fluctuations} \\
F_{desired} &= \text{ wayfinding force} \\
\Sigma F_{repulsive} &= \text{all repulsive forces} \\
\Sigma F_{attractive} &= \text{all attractive forces}
\end{aligned}$$

These forces are vectors, and consequently the acceleration resulting from them will be directional. An individual force will have an associated weighting and direction defined by:

$$F_i(\vec{e}_i, r_i) = \omega(\vec{e}_i, f_i) \times f_i(r_i) \tag{2.2}$$

An example of the calibration of these factors is shown below in Figure 2.2

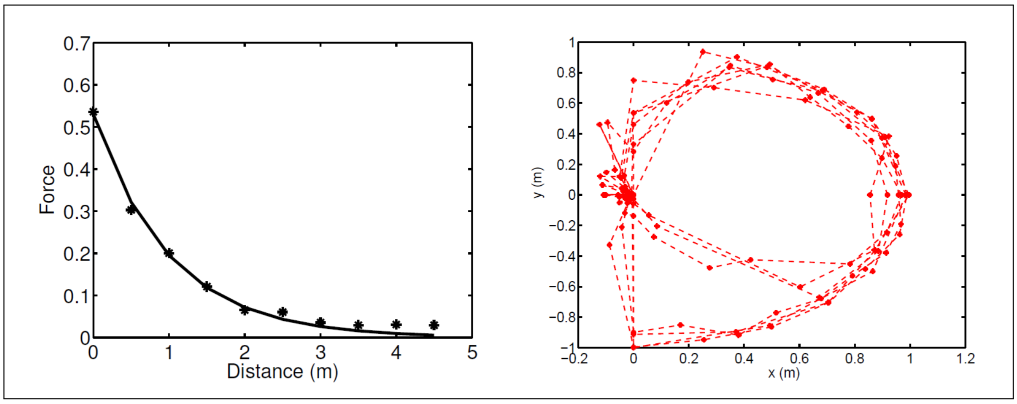


Figure 2.2.: Calibration of social force model, taken from [4], page 10

Optimal steps model

The optimal steps model is an example of a velocity-based model, focused on local optimisation algorithms. These models neglect any inertia of a pedestrian, assuming it to be negligible, and instead focus on selecting the optimal position within the available space. This requires the nearby area of each pedestrian to be interpreted as a scalar utility field, and an optimisation algorithm is used to find the best ("optimal") next step. This utility function can be formulated for a point as below (a simplified version, adapted from [5]).

$$P_i(x) = P_t(x) + \Sigma P_{peds}(x) + \Sigma P_{obs}(x) \tag{2.3}$$

where:

$$
\begin{aligned}
x &= \text{Position} \\
P_t &= Wayfindingutility \\
P_{peds} &= \text{Pedestrian utility} \\
P_{obs} &= \text{Obstacles utility}
\end{aligned}
$$

Individual utility functions can be formulated as:

$$
P_{i,i\neq t}(x) = \begin{cases} \mu & x < g_p \\ \nu e^{-kx} \qquad g_p < & x < g_p + h_p \\ 0 & x > g_p + h_p \end{cases} \tag{2.4}
$$

where

$$
\begin{aligned}
\mu &= \text{maximum avoidance utility value} \\
\nu, k &= \text{utility calibration parameters} \\
g_p, h_p &= \text{distance calibration parameters}
\end{aligned}
$$

These parameters are calibrated to produce a utility map (see Figure 2.3) for each pedestrian at each time point, for the range of movement available within a timestep. The minimum of this utility map is then selected as the next position of the pedestrian.

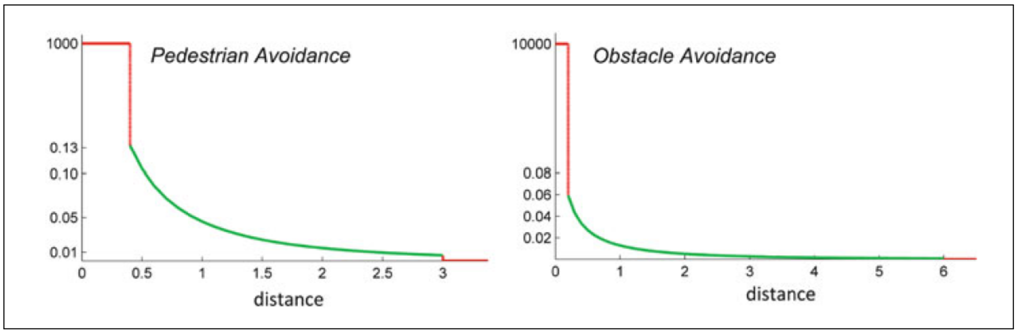


Figure 2.3.: Calibration of utility for optimal steps model, [5], page 173

Analysis techniques

Continuous analysis: Simple flows

Fundamental diagrams are used to calculate the movement of groups within a single direction of motion (the direction of travel), by modelling the relationship between local density and individual speed and group flow rate (see Figure 2.4). These are the basis for the hydraulic model, and are commonly used for hand calculations in estimating evacuation times. However, the model only considers one-dimensional flow, and therefore has limited applicability for complex real-world scenarios.

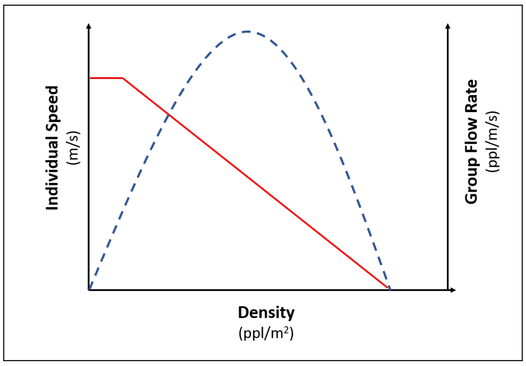


Figure 2.4.: Fundamental diagram

Continuous analysis: Complex flows

To produce models that can simulate more than one-dimensional movement, it is necessary to perform experiments that incorporate another direction of movement. Among several commonly used approaches, one option (example shown below in Figure 2.5) is to perform a T-junction experiment, where participants move in a group, arriving together from two ends of a corridor before turning a corner and combining their groups within a separate corridor. This allows data to be gathered on specific metrics, such as density, velocity and flow rate. These metrics are then used to calibrate movement models.

Discrete analysis

Another type of analysis that is relevant for quantifying human behavioural responses is discrete choice modelling, which is a useful method for understanding the impact of individual factors on discrete outputs (e.g. exit choice or wayfinding). This type of analysis is particularly powerful at reducing the

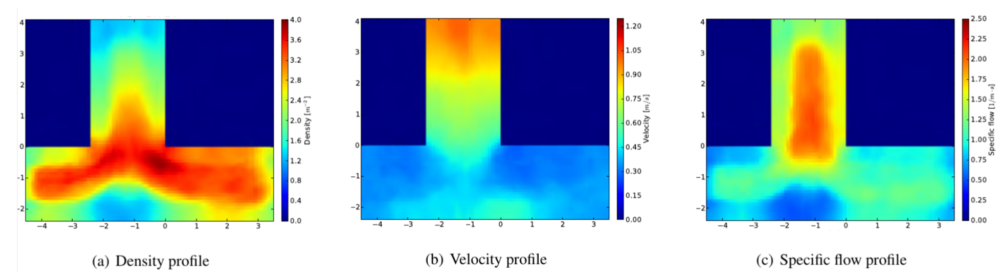


Figure 2.5.: T-junction diagram [6]

level of complexity seen in human behavioural responses while still producing meaningful results. A relevant example of this type of analysis can be found in study by Lovreglio et al. [50]), in which multinomial logistical models were implemented to determine the relevant predictors for the exit choice of individuals. These were shown to be dependent on both continuous and discrete factors, including:

- Number of evacuees at each exit
- Number of evacuees close to 'decision-maker'
- Distance of a 'decision-maker' to the exit
- Exit light signage
- Presence of smoke
- Flow rate through each exit

2.1.2. Pedestrian dynamics and emergencies

Previous research modelling human movement behaviour in emergencies has been wide-ranging, covering a host of different topics, techniques and threat origins. These studies are primarily computational, aimed at informing pedestrian dynamics models and predicting human behavioural responses to real-world scenarios.

Examples: Simulation only

While there have been many studies exploring how to introduce human behavioural and emotional considerations to pedestrian dynamics models, the majority of these have been theoretical simulation models, without any validation of the underlying algorithms. These models are still useful for understanding the different aspects of simulation that should be considered when implementing behavioural responses.

For instance, Hollmann et al. [51] describe a method of understanding changes in behaviour as a result of emotional state in travelling pedestrians. In this model the emotional state is implemented by agents altering their itinerary and route choice depending on the available time until their train departure. This is an initial step towards being able to model a behavioural response to stressful scenarios, as it allows for variation of the agents' responses within the model simulation itself, as opposed to prescribing a behaviour ahead of time.

Another example of this is shown by Helbing et al. [52], where different phenomena purported to occur during 'normal' and 'panic' situations are modelled using a social force based approach. This use of proxy panic variables is common approach. Helbing et al. [53] include herding behaviour in their attempt to explain arching and clogging at exits during an emergency evacuation. They do this by introducing a 'p' factor that dictates the likelihood of an agent following the average direction of their neighbours. Zou et al. [54] describe a model that shows competitive behaviour between agents on a lattice gas grid, examining the various outcomes when agents can either be 'gentle' or 'flustered', with 'flustered' agents being able to overtake the agent in front of them. Tsai et al. [55] incorporate a 'fear' characteristic that takes an integer value between 0 (no fear) and 2 (full fear). This characteristic affects the 'regular traveller' agent's actions, causing them to move radially away from a 'threat' and towards an exit.

Within the literature there are numerous different attack vectors considered. For example, Ma et al. [56] model the responses to a sarin gas attack within a railway station, using an agent-based approach for pathfinding and behaviour selection. Conversely, Fa et al. [57] focus on bladed attacks, investigating the impact of movement speeds and exit choice on the level of casualties during an incident. There are also many studies that look at responses to firearms attacks [58, 59, 60]. Finally, Lu et al. [61] combine terrorist attacks and stampedes to provide an initial look at multiple-vector threats, which is particularly relevant for modern terrorist attacks (see Section 2.5 for more information).

There are a variety of different underlying algorithms for these studies. For instance, many studies use social forces approaches [57, 56, 62, 63], while others use a cellular automata model [64], a Netlogo framework [65], or an adapted velocity-based model [66].

Modelling approaches often also include the concept of emotional contagion [66, 67], where the emotional state of an individual agent can affect that of another. This has a biological basis [68], and has been examined on a theoretical level on multiple occasions. For instance, Elzie et al. [69] studied group dynamics under various emotional and cognitive conditions, allowing the spread of information to vary between simulations.

The recurring feature within these approaches is the reliance on assumptions about how agents move in response to these emotional states, and in response to hostile threats. While it may seem an obvious choice to implement movement behaviour away from a threat, this does not necessarily agree with the data (see [70]). Without underpinning data or other validation, these simulation approaches remain completely theoretical.

Examples: Simulation and sociological theory

There have been examples of pedestrian dynamics research which attempt to implement the sociological theoretical constructs that have been established for crowd behaviour during emergencies (see Section 2.2 below for further information on these constructs). For example, Aguirre et al. [71] write an account of how agent-based modelling (ABM) can be used to replicate concepts from social sciences surrounding responses to emergencies. They conclude with seven different recommendations to computer modellers, finally arguing for the use of video footage to design and calibrate these approaches.

Pan et al. [72] provide an algorithm-based approach for modelling the behavioural responses of people in a state of stress. This study includes theoretical sociological constructs, focusing on modelling emotional and sociological concepts, while being flexible regarding the specific theories that are implemented.

Examples: Empirical analysis

A common feature in each of the papers detailed above is the difficulty in comparing the results with real-world data, owing to the lack of readily available data (either through sensitivity concerns, or rarity) and the difficulty of obtaining experimental evidence. This is the case both for the underlying theory of pedestrian dynamics during different emotional states (e.g. where do people move when they're scared?), and for the abstract theory of emotional contagion (e.g. how does fear spread?).

Tsai et al. [25] examine calibration by comparing a pedestrian dynamics model, combined with an emotional transmission model, with CCTV footage of two separate incidents. By analysing the difference in pixel location of several agents within the footage and within the model, they evaluated the performance of both the behavioural and the emotional transmission models. Similar approaches have been performed examining movement during earthquakes [73, 74, 75].

Some of the only physical experimental studies that have been performed was reported by Li et al. [23], investigating the initial movement responses to a hostile, knife-wielding actor from a range of different starting positions. This study provided a useful insight into the potential data collection methods, and implemented a simulation based on driving forces.

In a differently targeted approach, Li et al. [76] analysed the November 2015 terrorist attack in Paris, France, generating a behavioural and decision-making model for individuals. They use video footage to extract data, and analyse various concepts, including how the level of perceived threat changes with attack vector and distance from source.

Other approaches include analysis of the 2014 Kunming, China terrorist attacks via video footage to assess how crowds move in response to knife-based terrorist attacks [77]. This study notes the need for further research incorporating the psychological and agent-based methodologies.

These are some of the few examples of where research has used data (either experimental or real-world) to inform metrics and simulation models. However, serious questions remain about the level of generalisability and applicability of these studies. Ensuring accurate representations of real-world movement is an ongoing area of progress within pedestrian dynamics models.

Pedestrian dynamics Review papers

There are several review papers that provide an overview of various software packages, and a comparison of them in terms of certain metrics. For example, Duives et al. [34] provided a review of the different types of models available, comparing them with respect to computational efficiency and phenomena reproduction. This is a useful introduction to the separate modelling techniques (from macroscopic to microscopic) and their relative capabilities. Further to this, Kuligowski [40] provides a review of the available egress models, for use with fire planning and safety engineering. This review built on the research done by Gwynne and Galea [46].

Kinateder et al. [78] produced a thorough overview of the concepts (including risk perception and theoretical constructs), as well as a list of empirical data sources. While this focused on fire evacuation behaviour, it is still a useful resource for understanding potential avenues for the investigation of behavioural responses to terrorist attacks.

Finally, there are reviews that incorporate social science concepts with computer modelling. Templeton et al. [35] performed an analysis from a sociology viewpoint, classifying computer models from 1977 – 2014 in terms of their ability to represent human behaviour (from ‘homogeneous mass’ to ‘cognitive groups’). They argued that there is a need to include behavioural and group effects to pedestrian dynamics models to achieve a satisfactorily realistic representation.

2.1.3. Summary

This section has discussed the significant progress in the area of pedestrian dynamics over the past decades. However, this section has also established that the state-of-the-art in this area still does not accommodate emergencies well. Previous approaches have been based on outdated theoretical constructs such as panic theory (see below for more information), or not based on theoretical constructs at all. Such simulation approaches have limited applicability, as they represent the modellers’ opinions, rather than established phenomena.

These previous approaches are also similarly limited by the lack of data. Research that uses any data, whether obtained through experiments or real-world observations, is an improvement over research that is entirely theoretical. However, the real-world data sources discussed above are uncontrolled and incomplete, and the experimental approaches suffer from a significant lack of ecological validity.

Without the required datasets, and the theoretical backing, any approach will be hypothetical at best, and misleading at worst. The overlap between pedestrian dynamics and emergencies (such as terrorist attacks) has a huge potential for mitigating the negative outcomes of such incidents, including saving lives. However, this potential has yet to be realised.

2.2. Behaviour in emergencies

2.2.1. Theoretical constructs

This section will detail the theories that have been developed on how people react during emergencies. These are largely group-based (sociological) approaches, detailing the types of response, and the key scenario factors in eliciting this response.

There have been numerous theories surrounding behavioural responses to emergencies. This section will analyse the most immediately relevant and well-known ones, namely: Panic Theory, Emergent Norm Theory, Social Attachment Theory and Self Categorisation Theory. This section will also outline an individualistic theory, the Protective Action Decision Model.

Gavin Le Bon's Panic Theory ([27, 79]), published in 1895, was the primary forerunner theory in understanding how human beings react to physical threats, arguing that within a group acting in the presence of a threat, human beings lose their individuality and become an anonymous part of a whole. Therefore any action within the group is immediately copied unquestioningly by every other individual, causing a 'mob mentality'. Le Bon's theory fits only loosely with this PhD's definition of 'panic' as behaviour which comprises the following features:

- Competition: the actions of individuals seek to prioritise themselves over their immediate neighbours (e.g. pushing people out of the way to get to a door).
- Irrationality: the actions of individuals are not optimal (e.g. ignoring an immediate threat and focusing on prior actions, as in [80].
- Non-adaptivity: the actions of individuals do not change based on new information (e.g. continuing to push on a door that has 'pull' written on it).

Quarantelli [81] described the required features for such a panic to develop, before concluding that 'compared with other reactions panic is a relatively uncommon phenomenon' (page 275). This has also been argued by others ([82, 83]), and seen empirically ([84]). Despite this academic research, Panic Theory and the conceptions of panic as a response to emergencies are still prevalent both in popular culture and in official guidance ([85, 86]), leading to significant misconceptions in the public mind.

There have been review papers on sociological and individual responses to emergencies. In 1965, Schultz [87] discussed various theories of panic, analysing the required elements for this response to occur, either not aware of or dismissing Quarantelli's 1954 arguments. Instead Schultz focuses on Le Bon's crowd theories, which provides an insight into the historical considerations given to the topic, even if panic is much rarer than this paper implies. Chertkoff and Kushigian [26] also provide a review of the theories, again focusing on panic theories, before evaluating their applicability to several historic incidents.

This is not to say that panic does not occur, but rather that it occurs in rarely and in unique situations. Schultz [88] has offered compelling evidence of panic in military settings, and Johnson [89] summarises several different incidents involving some form of emergency crowd movement. The existence of panicking individuals might be explained by Entrapment Theory, which outlines the scenarios in which people begin to ignore social rules, and to act in irrational, non-adaptive and competitive ways. These are when there is [90]:

1) An immediate, pressing threat to the individual's life.
2) An opportunity to escape, which is diminishing quickly.
3) A feeling of helplessness in otherwise dealing with the threat.

This definition is mirrored by Chertkoff and Kushigian [26], who go on to state that large-scale emergencies and fatalities occur when several preconditions are met in combination, namely:

1) Severe limitations on space.
2) Large numbers of people present.
3) Lack of knowledge about the local layout.
4) Lack of emergency management and guidance.

A successor to Panic Theory was the Emergent Norm Theory (ENT) [27], which suggests that, instead of panicking, people group together during an emergency and, out of an ambiguous scenario, define what is occurring and the best course of action is developed from within the group. This is achieved by the generation (the 'emergence') of new social norms. This theory permits a heterogeneous group to start acting in a homogeneous way, and has been used to inform conceptual models [91].

The Social Attachment model [92] attempts to explain the empirically observed actions of individuals who seek out 'familiar' others, such as friends or family, in an evacuating crowd ("affiliative behaviour"), and has been examined against various scenarios, including combat and fires. This theory provides a potential avenue into modelling the likely choices of people moving within previously defined groups, especially in an evacuation scenario. This is strongly linked to Affiliation theory [93], which examines how individuals who are unfamiliar with a built environment will seek out previously-known people and locations in the event of an emergency.

Self-Categorisation Theory (SCT) allows people to define themselves dynamically within different groups, an effect which is exaggerated during an emergency. This self-categorisation details how 'emergent properties of group processes can be explained in terms of a shift in self-perception, from personal to social identity' ([94], page 2). Crucially, SCT permits an explanation of why strangers tend to help each other in emergencies [95]. SCT has been examined against a real-world backdrop on several occasions, as for example during the 1990 poll tax riots in the UK [96].

The Protective Action Decision Model (PADM) [97] focuses on individual behaviour, rather than grouped, social behaviour. PADM describes the internal processes of an agent through an emergency scenario, and can be readily used to explain the 'milling' (people waiting until a course of action is determined) and 'keynoting' (individuals taking charge of a group) processes that occur in ENT. However, it is not easily applied to the social aspects of emergency evacuation, detailed by the SCT and the Social Attachment model.

2.2.2. Data

To inform these theoretical constructs, data on behavioural responses to emergencies has been assessed throughout the past century. As detailed above, some of the first analysis was performed by Quarantelli [81], in refuting the validity of Panic Theory. Another key example includes the investigation by Feinberg and Johnson [98] examining the behaviour of occupants of the Beverly Hills Supper Club fire, aiming to correlate their ultimate fate with one of several models of the breakdown of social order.

Similarly, Baddeley's investigation [99] examined various papers on human performance in a range of extreme scenarios, including combat, parachuting and scuba diving. During this research he observed several effects, including 'attention-narrowing', where peripheral activities are given less attention and consequently are performed worse. Schultz's [88] research on examples of panic within military contexts was referenced by Mawson [92] to help form the basis of the Social Attachment model.

Further to this, Kuligowski [100] analysed 245 separate interviews with survivors of the attacks on the World Trade Centre, 11th September 2001, and described two key thresholds in behavioural responses, where occupants recognised either a threat to personal safety, or the presence of social cues to evacuate. This work built on several theoretical bases, including the PADM and the ENT, and developed a further model of occupant behaviour.

However, as mentioned above in Subsection 2.1.2, there exists little quantitative data surrounding movement behaviour in emergencies that can be used to calibrate pedestrian dynamics simulation models.

2.2.3. Summary

The theories that have been discussed aim to explain exactly the behaviour that this PhD seeks to understand. However, there remains significant work to be done before these theories can be satisfactorily used in computational models. Primarily, these theories require large amounts data to fully validate their key principles. This data is necessarily difficult to obtain, owing to the sensitivity of real-world observations and the lack of appropriate experimental approaches that include pedestrian dynamics concepts.

Additionally, these theories need to be predictive and quantitative before their use in simulation models can be assessed. Without predictive capability, the models cannot be compared with real-world outcomes, and therefore validated for future hypothetical scenarios.

2.3. Psychology of emergencies

2.3.1. Emotions: Conceptual elements

In order to model behavioural responses to emergencies, it is necessary to first understand the psychological modelling that has been done previously. This field has produced several major theories covering the concept of emotions, which will be summarised here.

James [101] is often credited as the first person to study the concept of emotions within the psychophysiological field, arguing in 1884 that emotions are the by-products of physical effects that occur in response to external stimuli. While this James-Lange theory of emotions has been discredited (see [102] for further detail), it did bring into question the concept of emotions and their relationship with physical reactions.

Study in this area has been ongoing since James [101] proposed the theory of psychophysiological responses to emotions. More recent theories instead focus on two major conceptual models: those that favour individual basic emotion sets (e.g., anger, disgust, fear, joy, sadness, surprise, as defined by Ekman [103], in 1982), and those that favour dimensional emotion models. Dimensional models often characterise emotions by valence (positive or negative, happy or sad) and by arousal (high or low, agitated or drowsy), as shown below in Figure 2.6.

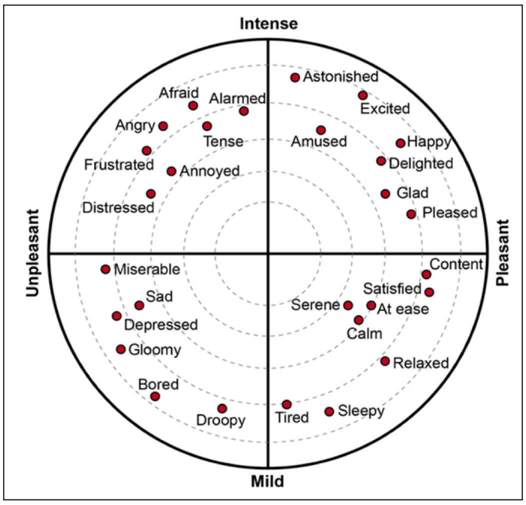


Figure 2.6.: Valence-Arousal affective space, taken from [7]

One obstacle is the fact that consistently eliciting a particular emotion is an inherently difficult task, as it relies on individual responses, which vary according to said individual. Therefore it is possible that an experiment that elicits a particular emotion in one subject could elicit a completely separate emotion in another. Indeed, this effect is often an area of research, focusing on phobias and trauma [104]. One solution to this obstacle is the International Affective Picture System (IAPS) [105], which provides pictures that have consistently produced particular emotional responses.

2.3.2. Fight-or-Flight responses

The 'Fight-or-Flight' response was first proposed by Cannon [70] in the early 20th century. Here, Cannon sought to explain the emotional response seen in times of extreme stress, and also provide a counterpoint to the James-Lange theory of emotions.

In 2016 McCarty [102] produced an overview of the history and context of the 'Fight-or-Flight' response theory, including identifying many examples of seminal studies in this area of research. Further to this, in 2010 Romero [106] provided an in-depth review of the physiological effects of such a response, concluding that it is a complex interaction, and that some physiological reactions are counterintuitive. For instance, in some animals, the presence of a predator induces a reduction in heart rate and the temporary decision to remain immobile. This is sometimes known as freezing, and in extremes is known as 'playing dead'. If the threat passes, there is a subsequent increase in heart rate, typical of a standard 'Fight-or-Flight' response. Romero does note, however, that the production of catecholamines (e.g., adrenaline and noradrenaline) is a common feature of this response.

2.3.3. Quantitative measures of emotion

A significant obstacle in emotional modelling is the availability of a single established, reliable and quantitative measure of emotions. According to Bradley and Lang [105] and Mortillaro and Mehu [107], emotional responses fall into three categories: behavioural, physical and affective. This approach groups multiple reactions under one umbrella term for emotion. So, for example, an anger emotion might be characterised by a particular change in physiology (heart rate, skin conductivity), a change in behaviour (walking cadence, muscle activation, volume and pitch of voice) and a subject-specific internal feeling (often measured on a self-assessment scale, e.g., the Self Assessment Manikin, SAM, [108]).

For this PhD to provide a validated experimental method of measuring behavioural responses to hostile emergency scenarios, it will be necessary to both produce and measure the emotional responses that are predicted in such a scenario. To measure these emotional responses, it is therefore necessary to first understand the different methods of measuring these states, and the relative benefits and drawbacks.

Starcke and Brand [109] presented a review of papers exploring decision making under stress, examining a range of concepts such as the neuro-biological and psycho-physiological concepts of stress, before going on to critically examine papers investigating decision-making while exposed to stressful scenarios. This research both reviews and informs several of the quantitative measures of stress within the studies assessed here and will help develop the experimental procedures detailed below in Chapter 3.

Autonomic Nervous System

Psychophysiological emotional modelling has provided several ways of quantitatively measuring emotion, a prominent example being the Autonomic Nervous System (ANS) responses. In measuring ANS responses, it may be possible to infer an emotional state. These ANS responses include Heart Rate (HR), Heart Rate Variability (HRV, usually measured by an electrocardiogram, ECG), Skin Conductance (SC), Skin Temperature (ST), Respiration Rate (RR), Muscle activation (as measured by an electromyogram, EMG), and many others. The major benefit of using ANS responses is the low level of conscious control exerted by the subject, therefore resulting in unbiased, quantitative responses.

Using ANS responses as a measure of emotion is vulnerable to artefact generation, specifically by the movement of the volunteer undergoing measurement. Healey [110] found that physiological measures were a powerful method of differentiating between emotions (up to 100% for the generated dataset), but that in unconstrained ambulatory settings, the signals produced by emotional responses were overwhelmed by those produced by physical activity.

Furthermore, it is often noted that these ANS responses are not unique to emotional states (e.g., ‘fear’ and ‘anger’ both produce high positive changes in heart rate, despite being distinct emotion sets [111]), and instead there is a many-to-many relationship between ANS responses and emotions [112]. There has been some success in using multivariate analysis to produce dimensions along which individual emotions can be discerned. One of the most powerful of these approaches is the classification procedure proposed by Kreibig [113] in 2007, which can predict the type of emotional response with up to 85% reliability. These dimensions take many ANS, central nervous system (CNS) and behavioural measurements and weight them depending on the level of their discriminating power. This approach improves as the number of different responses is measured.

As Mauss and Robinson [114] comment, there is evidence for both dimensional and specific basic emotions. They also comment that the sole use of ANS responses is insufficient to distinguish between emotional subsets; however it is a powerful method of quantifying valence and arousal levels.

Endocrine system

Given the restrictions on ANS response measurement, namely that the movement of subjects can obscure emotional identification, further methods of measuring the biological responses to stressful scenarios have been developed. One such method is to use the responses of the human endocrine system.

It has been observed that during fear responses the body produces several different hormones, including Adrenocorticotropic hormone (ACTH), cortisol and adrenaline [115]. This was examined thoroughly by Dickerson and Kemeny [8], who examined 208 previous laboratory studies involving psychological stressors and cortisol responses. They note that this cortisol response also occurs in other types of stress, with uncontrollable and social-evaluative elements being associated with the highest response and longest recovery duration. Figure 2.7 shows a typical ACTH and cortisol response to a social-evaluative stressor in ten-minute intervals.

Furthermore, cortisol measurements have been assessed as a response biomarker by Hellhammer et al. [116], including analysis of the different sources of cortisol measurement, including serum, blood, urine and saliva. Hellhammer et al. also commented that the use of saliva samples is an acceptable form of measurement, and that it has the added benefit of being non-invasive and easy to perform.

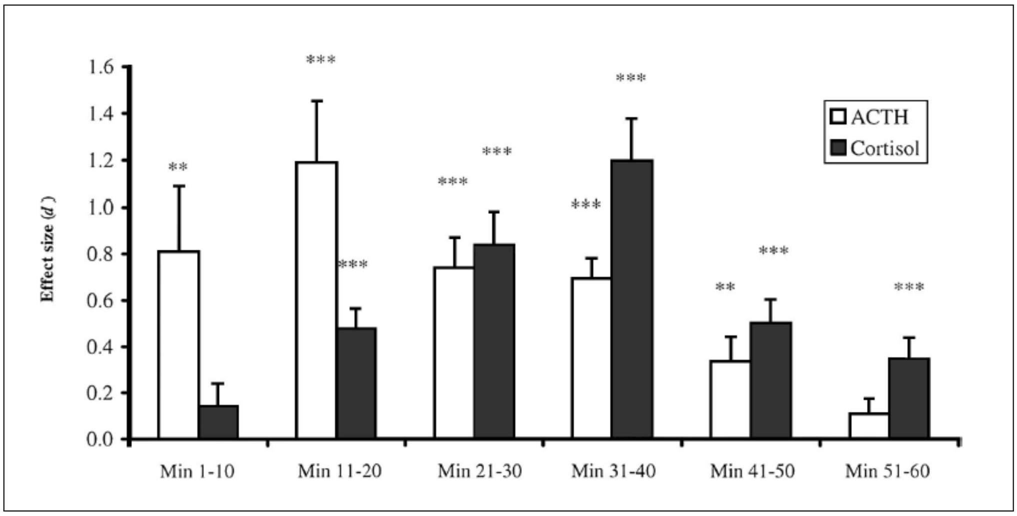


Figure 2.7.: Adrenocorticotropic hormone and cortisol effect sizes (d). [8], page 375.

Self-assessed measures

Self-assessed (survey) approaches have been in use since the establishment of the field of quantifying emotional states. They are the most commonly used method and there are significant amounts of validation work underpinning the most widely used measures. Examples of measures that are used to quantify generalised emotions include the Self-Assessment Manikin (SAM) [108] and the Discrete Emotional Questionnaire (DEQ) [117]. However, there are also specific questionnaires that assess negative emotions, such as the State Trait Anxiety Inventory (STAI, [118]) and the Positive Affect, Negative Affect Schedule (PANAS, [119]).

An issue with these measurements is the reliance on the subjectivity of an individual's responses, and the consequent difficulty in establishing an objective comparison between two individuals. This issue has been assessed throughout the research history of these methods, and generating these surveys requires a lengthy amount of validation work before they are considered reliable. [120, 121]

Alternative measures

There are several other methods for measuring a fear response in a quantitative manner. These include the use of functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and electrocorticograms (ECoG) [122, 123, 68]. These tools measure activity within specific regions within the brain, and therefore can detect when the amygdala (which reacts to fearful stimuli) is undergoing a particular response. These methods are particularly well suited to quantitatively and objectively measuring the response of an individual; however they are also limited to certain settings (i.e. within a hospital), and have a variety of further limitations (e.g. cannot be ambulatory, must not have any metal on the person).

Possible ambulatory systems that mirror these approaches include near infra-red spectroscopy (NIRS) and electroencephalography (EEG). These methods are becoming increasingly established for the purposes of quantifying and spatially locating metabolic activity within the brain; however, they currently lack the level of validation, spatial or temporal resolution, and the reliability to provide accurate measures [124, 125]. Consequently, these methods will not be pursued within this PhD.

Review papers

Balters and Steinert [126] performed an exceptional review of emotional measurement techniques, with detailed reference to the physiological reactions (ANS) detailed above. They go on to detail the overlap between human factors engineering, design engineering and emotional responses. Mauss and Robinson [114] also looked at both ANS and alternate methods of measuring emotional responses. Further to this, Bradley and Lang [105] and Mortillaro and Mehu [107] both produced authoritative summaries looking specifically at the various measures of emotion.

Camerer and Hogarth [127] provide a detailed review of the use of financial means as a method of incentivising participants, concluding that there is a complex relationship, with some studies showing improved performance, some showing degraded performance, and some showing no relationship.

2.3.4. Summary

The field of experimental psychology that quantifies and measures emotional responses has extensive background literature. However, many questions remain surrounding the appropriate models and measures for specific emotions, with evidence for multiple approaches. This lack of clarity implies that a wide-ranging approach is required, measuring as many physiological and psychological elements as possible to ensure any experimental outcomes are observed.

2.4. Empirical measurement of behavioural responses

To model emergencies, the types of behaviour and cognition that are observed in these scenarios must be used to calibrate the models. This section will examine some of the previous research that measured these behavioural responses, including virtual reality, laboratory, animal and field experiments, and finally review several alternative measurement methods, including real-world observations.

2.4.1. Experimental design

Experiments that allow measurement of human reactions to a stressful situation are inherently difficult to perform, given both practical and ethical considerations. Recently Haghani and Sarvi [128] studied the methods that have previously been used to obtain data for use within this area of research, finding seven major techniques:

1) Virtual reality (VR) and hypothetical choice (HC) experiments

2) Animal experiments

3) Controlled laboratory experiments with human subjects

4) Evacuation drill experiments

5) Analysis of natural disasters

6) Video analysis of pedestrian crowd movement in natural environments

7) Post-disaster interview with survivors

Further to this, Lawson [129] compares several different methods, including virtual reality, incident reports, eyewitness statements, footage of the event, field experiments, drills (announced and unannounced), participant and expert predictions, and finally survey methods for hypothetical events.

Haghani produced a two-part paper in 2020 examining the state-of-the-art for experimental methods and highlighting controversial topics for which there is mixed or inconclusive evidence. Part 1 [130] detailed contemporary experimental methods, identifying the following key areas:

- Evacuation of vulnerable people (mobility impairment and disability, elderly and children)
- Evacuation training, education and guidance
- Vertical evacuation (movement on stairs, evacuation of high-rise buildings)
- Group behaviour
- Evacuation under limited visibility

Part two [131] identified the following controversial topics :

- Faster is slower
- Effect of partial obstruction
- Relationship between exit width and capacity
- Additional exits and blind evacuation
- Corner versus middle exit flow rates
- Symmetry breaking and social influence on exit choice
- Stress and navigation
- Stress and decision changing
- Social groups and evacuation efficiency

Methods of eliciting emotion

Experimental approaches that investigate behavioural responses to emergencies also need to consider how to elicit the appropriate emotional subsets. Kučera and Haviger [132] provide insight into the various methods of eliciting emotional response, examining:

- Displaying material (e.g., photographs)
- Imagination techniques (including recall)
- Pre-set interactions
- Facial feedback paradigm
- Administration of drugs
- Use of a real interaction

They use self-assessment forms to compare three methods of eliciting emotional response and find that anger is most effectively induced via pre-set interactions and fear is best induced via real interactions.

There have been numerous methods used to elicit states of stress, including the use of financial incentives, time pressure and distractions. The Trier Social Stress Test (TSST) is a type of non-physically threatening test that is regularly used to produce this type of response. Allen et al. [133] have examined the varying responses using the TSST, and have included that any methodology should consider age and gender differences, as well as control for caffeine, smoking and contraceptive use.

Further to this, Buckert et al. [134] use cortisol and heart rate measures to examine the level of stress, and compare this with the performance within a financial competition. This approach was then extended in [135], where a cognitive (mental arithmetic) competition with financial incentives is used to induce a state of stress. In this example it is suggested that a 'winner-takes-all' tournament induces a higher stress level than a format where participants can reliably store their winnings.

Marín-Morales et al. [136] produce a comprehensive review of previous studies that have used virtual reality techniques to emotional responses, before analysing the potential for machine learning algorithms to distinguish between these emotional states.

2.4.2. Virtual reality

Virtual reality (VR) has been used for decades as a data gathering and generation tool in multiple formats. It is a particularly powerful tool when complete repeatability is required, and when field experiments are not possible. VR can come in many forms, varying from basic computer screen simulations (for example, Second Life [137] and computer games) to highly immersive virtual environments (IVEs).

VR is distinct from Augmented Reality (AR) and Mixed Reality (MR). In both AR and MR systems, the participant is aware of their physical surroundings, able to see people and objects. In MR digital objects can react to real-world objects, whereas AR typically only provides a graphical overlay. In contrast to both of these, VR provides a complete digital environment that often bears no resemblance to the participant's physical surroundings. This is shown below in Figure 2.8.

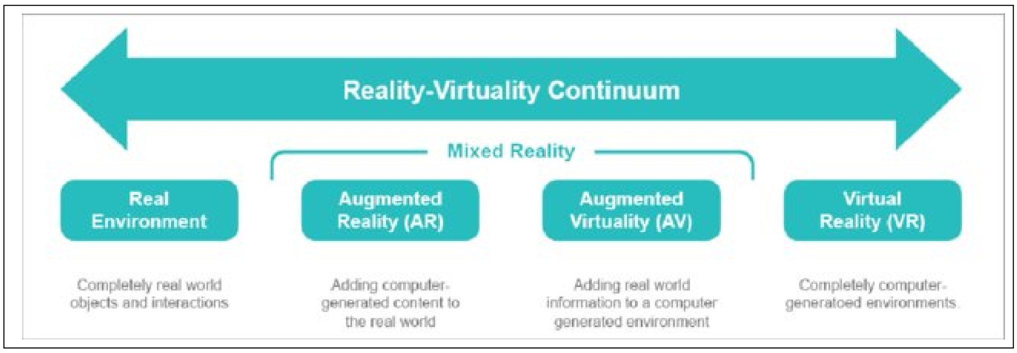


Figure 2.8.: Reality- virtuality spectrum [9]

VR types

There are numerous different VR systems available, including Cave Automatic Virtual Environments (CAVEs) and Head Mounted Displays (HMDs), as well as non-immersive virtual reality environments, such as powerwalls and desktop computers.

Cave Automatic Virtual Environment (CAVE)

Cave Automatic Virtual Environments (CAVEs) are a set of four (or six) walls where three-dimensional images are projected to present the IVE to the subject [138, 139]. In a CAVE system, a participant stands in a large container, and the virtual environment is projected onto the walls, as shown in Figure 2.9. The participant is tracked while moving, and the wall projections subsequently updated to represent movement. Further movement and interaction are provided by an interface, for example a controller or haptic gloves.

This method is immersive, requiring polarised glassed for three-dimensional experiences.

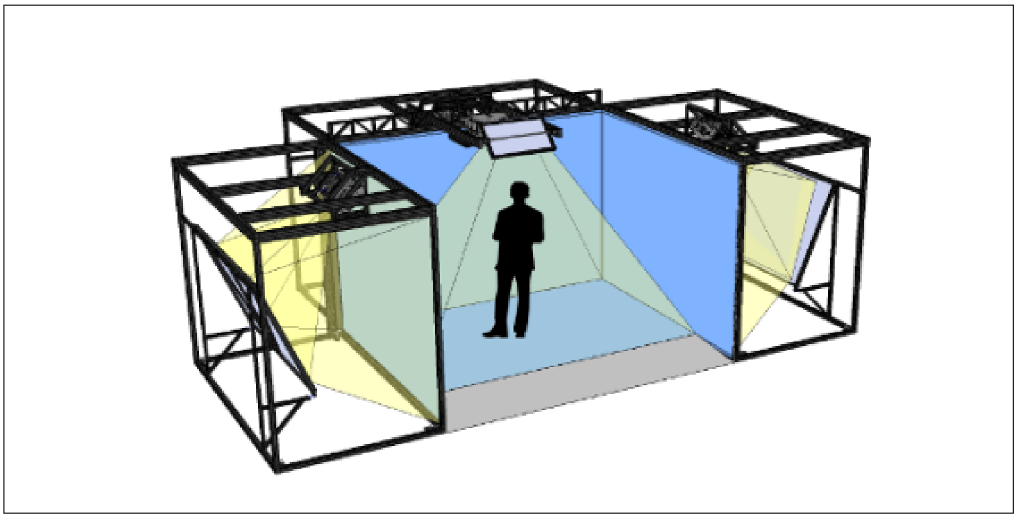


Figure 2.9.: A CAVE sysem

The CAVE system provides several benefits, including a wider field-of-view than typical HMD systems while having no requirement for a headset. However these systems result in increased cost and space requirements over HMD systems.

Head Mounted Display (HMD)

Head Mounted Displays (HMDs) provide another way of providing an immersive environment. However, there are specific issues with the feeling of motion-sickness resulting from low latency systems (usually characterised by a motion-to-photon time, or how quickly the virtual representation updates in reaction to the motion taken by the occupant) and different locomotion techniques [140, 141].

In HMD systems, the participant wears a headset that completely covers their eyes, while the headset's internal screens present views of the virtual environment. This allows three-dimensional viewing as standard, but resolution is limited. Furthermore, the headsets can become uncomfortable during long sessions. Within HMDs, there are further subdivisions:

Outside-in tracking

HMDs require high quality positioning systems to ensure that the pictures shown to the participants are indicative of their position in the virtual environment. This requires a six DoF system, providing millimetre positional and sub-degree rotational accuracy, and an update rate of 50Hz or higher. To perform this, some HMDs use base stations that scan the area to locate the participant and determine their position and orientation. This data is then sent to a central computer, which in turn provides the HMD with the required display. These systems generally have a higher graphical power than their inside-out counterparts, but are more expensive when including the cost of a sufficiently powerful computer tower. An example of this form of HMD is the HTC Vive Pro, which has an optional eye-tracking and wireless movement format.

Inside-out tracking

HMDs using inside-out tracking can be used as standalone systems, without being tethered to a PC tower, or using base stations to aid in positioning. These systems can perform this tracking using highly developed simultaneous localisation and mapping (SLAM) systems, including computer vision and onboard IMU chipsets. They typically have lower graphical power than their outside-in counterparts, but also tend to be cheaper, and are more flexible when considering experimental set-up and mobility. An example of an HMD using inside-out tracking is the Oculus Quest.

The difference between Outside-in and Inside-out tracking is shown graphically in Figure 2.10.

Powerwalls and Desktops

Previously the definition of virtual environments has included non-immersive, two-dimensional screens, such as powerwalls and desktop-based experimentation [142, 143]. Both of these environments provide the ability for large-scale experimentation, as well as relatively easy shared-space design (a 'multiplayer' approach). However, they also suffer from a lack of immersion and the consequent lack of ecological validity. These environments will not be discussed here.

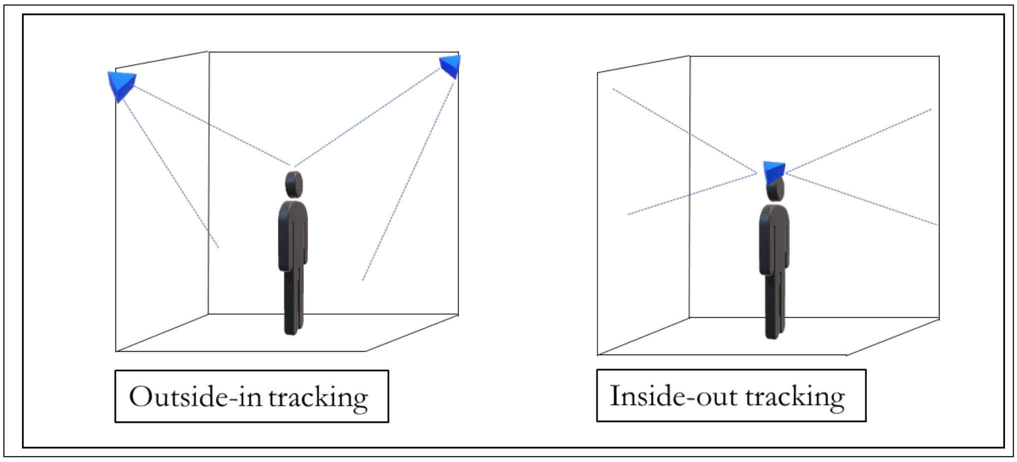


Figure 2.10.: Outside-in and Inside-out tracking

Ongoing research

There are several unanswered research questions surrounding the use of VR as a data-gathering tool, especially when considering behaviour in emergencies. In 2014 a research study by Kinateder et al. [1] performed a Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis examining these, concluding that there are large potential benefits to using VR as an experimental paradigm, but that research on validation should be performed before any results can be incorporated into the existing body of research. This SWOT analysis is summarised in Table 2.2.

Feng et al. [144] performed a systematic review examining how VR has been used, focusing on emergency evacuations and pedagogical research, and concluded that there has been a prior focus on using VR for fire-related emergency evacuations. Therefore, there is an opportunity for the use of VR to investigate further emergencies, such as terrorist attacks.

The concept commonly used to discuss whether VR can be used as a scientific research tool is the extent of 'presence' experienced by volunteers within the environment. 'Presence' is defined as the extent to which the volunteer imagines they are actually within the virtual environment, and therefore how closely behaviour and responses mimic those that would occur in reality. However, there are also limitations in using the concept of 'presence' as, much like the concept of emotions, it is difficult to measure quantitatively, and furthermore it is difficult to establish an objective comparison between participants.

Strengths	Weaknesses	Opportunities	Threats
Internal validity - specific measurements External validity- more generalisable than individual physical drills Replication and repeatability Ecological validity - visual stimuli Participant safety Real-time feedback Multi-modal simulations Precise measurement Repeated measures Psychophysiological measurement Low running cost Flexibility Control of confounding variables Experimental logistics	Validation Interaction-UX/UI Technical limitations Side-effects Time-intensive development	Intuitive navigation Graphical developments Multi-modal simulation and feedback Usability for researchers Exchange of environments	Ethical challenges Ecological validity Side-effects Misleading expectations Technical faults

Table 2.2.: VR SWOT, adapted from [1]

There have been several different approaches in measuring presence, including self-assessed measures ([145, 146]) and ANS responses. For example, in 2001 Meehan [147] proposed using physiological responses to objectively quantify the level of presence in virtual environments, using heart rate, skin conductance and skin temperature as a way of measuring the response to a virtual environment.

In 2005, Meehan et al. [148] reviewed four studies that investigated the use of physiological effects as a measure of presence during a stressful event within IVEs, finding that HRV and SC provide strong, objective measures for presence, whereas skin temperature did not. Further to this, in 2006, Brogni et al. [149] examined this effect further, attempting to understand whether the physiological responses are an artefact of the IVE. Vinayagamoorthy [150] examined the effect that virtual avatars had on volunteers within an IVE, using physiological measures (ECG and SCR), self-assessment, and behavioural responses (for example, moving away when avatars encroached on their 'personal space'). Zou et al. [54] produced an IVE experiment in which subjects were asked to evacuate from a burning building. By comparing the self-assessment and typical ANS responses, they showed that subjects experience a significant level of presence within such a scenario.

Although these are clear examples of a validation of the effect VR has on the ANS responses, there has been comparatively little work done to validate the behavioural responses during VR scenarios (such as movement patterns and reaction times). One example of this type of work was performed in 2013 by Slater et al. [151], who examined the use of IVE to examine bystander responses to violent emergencies. A further example is Moussaid et al. [152], who provide varying financial incentives to move through a bottleneck within a corridor, measuring herding and other social effects while varying levels of 'stress'. Another serious games approach was performed by van den Berg [153], allowing participants to choose the type of transport after hearing an in-game alarm, as well as react to what other participants are doing, this research measures the effect of herding during an emergency. Lin et al. [154] also produced a study looking at how participants in an IVE responded to avatar movement during an emergency in a railway station. Crucially, these examples all suffer from a lack of quantitative real-world data to compare against, which presents a problem for the validation of VR as a data-generating paradigm.

A relevant recent study was performed by Awada et al. [24], focusing on using heart rate and self-assessed questionnaires to assess emotional responses in VR, while participants moved in response to an active shooter. This study investigated different locomotion methods, identifying that walking in place was the method that allowed the highest sense of presence and produced the largest emotional responses.

Further issues regarding interactions between participants within VR environments include the fact that individual reactions are limited to how the VR environment is implemented- since there is not as much flexibility for the participant's action choice in VR as in physical experiments. For example, crypsis, or 'playing dead', is a possible approach for participants in a physical experiment, but becomes a particularly complex problem for implementation within a VR environment.

2.4.3. Laboratory experiments

Laboratory experiments offer an opportunity to measure specific reactions in a controlled context, and therefore minimise the potential errors in determining the size of reaction to a given stimulus. This is particularly important with respect to human behavioural measurements, as any exogenous variable can affect individuals in a number of ways, skewing the reactions and therefore the measurements.

Mintz [155] formulates a thought experiment into the conditions necessary for panic to occur and provides some of the first experimental evidence of how actions change in the presence of a threat or fear stimulus. He shows that people can engage in non-adaptive group behaviour when their needs conflict with others. Kelley et al. [156] comment on Mintz's research, decrying the lack of a sufficient stressor to properly investigate the demonstrated effect. They further Mintz's research by threatening participants with an electric shock to investigate how social cues affect their actions while under threat. Keinan et al. [157] subsequently timed several different effects seen when subjects were provided with cognitive tests while exposed to the threat of electric shocks. These effects include: 'temporal narrowing' (assessing each option for less time than previously), 'premature closure' (choosing an answer before having assessed all options), and 'non-systematic scanning' (not checking the answers in order). These effects provide a potential basis for modelling individual responses to threatening stimuli in a computer simulation, for a microscopic-level agent based, cognitive model.

More recently, experimental research has been performed by Drury et al. [158], who use a desktop based VR scenario to measure how likely participants were to push past or help each other during the evacuation of a train station, depending on the level of agreement between the perceived category of the simulated agent and the self-categorisation of the volunteer. Bar-Haim et al. [159] investigated how people might experience time during an emergency scenario by asking volunteers to estimate how long a picture showing an actor producing a 'fearful or calm expression' was shown. They further examined the effect of anxiety on this measurement as potentially informing various elements of a cognitive model. This approach would also likely serve to support Baddeley's [99] (see Section 2.2) proposition of attention narrowing.

Laboratory experiments have been used extensively to produce quantitative inputs into computer modelling software. However, these generally focus on day-to-day events, including the fundamental diagrams defined in Section 2.1. There exist several datasets (e.g., [160, 161, 162, 163, 164]) of human movement in 'stressful' scenarios, where people are asked or incentivised to move through a scene as though they were in an emergency evacuation. While this is a more appropriate source of data for emergency modelling calibration, there are often insufficient stressors to provide ecological validity.

2.4.4. Animal experiments

Animal experiments represent a paradigm for creating scenarios with real risk to experimental subjects while remaining within ethical boundaries. These have been commonly used in the past, and often inform theoretical constructs that are used for simulation models. For example, Shiwakoti el al. [165] achieved this by placing ants within an area and introducing ant repellent (citronella), causing the ants to move towards the exit. In the same study they then go on to develop an emergency evacuation simulation model ('Em-Sim'), based on the Social Force Model. Altshuler et al. [166] also attempted to understand herding behaviour in humans by observing ants in an 'acrylic drum' with diametrically opposed exits, after also introducing ant repellent to the system. This study observed a statistically significant effect of symmetry braking (i.e. herding behaviour) in ants, as they followed other ants towards a presumed exit rather than choosing another, potentially more beneficial exit.

Calibration of these models by approximating ant movement data with human responses has been used to analyse design of turning on emergency evacuations ([167, 168, 169] and multiple stream merging behaviour ([170, 171]).

Woodlice have also been used as a source of data. Sobhani et al. [172] placed 120 woodlice within a box with a single exit, and use a lamp to approximate the threat of sunlight. They measure the density and flow rate through the exit as the woodlice evacuate the box, finding that exit flow rates fluctuate, and that exit capacity increases with exit density. However, it must be noted that they comment that the woodlice were climbing over each other, which could be a source of the increased exit capacity, rather than an increase in individual movement speed.

There have been experiments using sheep as a source of experimental data. Zuriguel et al. [173] recorded sheep as they moved through a door into a barn. They installed an obstacle at varying distances, measuring the flow rate in each instance. They conclude that providing an obstacle a short distance from the opening improves the flow rate through the opening, mirroring what has been seen elsewhere (e.g.,[174]).

However, with all of these approaches there are significant concerns regarding the extrapolation of these methodologies and subsequent datasets to inform models that seek to predict human behavioural response. This is partially because of the difficulty in extrapolating from animal to human behaviour, and partially owing to the limitations in the results reported and their lack of applicability to real-world scenarios. As this PhD focuses on producing valid results that are applicable for pedestrian movement, it will not pursue animal experimentation.

2.4.5. Field experiments

Field experiments are a commonly utilised element of research into human behaviour in emergencies. These are experiments that take place in public settings, often with members of the public, while other day-to-day activities are taking place nearby. This setting gives a higher level of ecological validity, but suffers from potential confounding variables that are often both unrepeatable and impossible to discern from the data gathered.

Examples of these field experiments include a set of unannounced fire drills in hotels, performed by Kobes et al. [175], where naïve occupants have their responses (reaction times, wayfinding, exit choice, movement speed, etc.) recorded after a fire alarm goes off in the middle of the night. Another example is Proulx and Sime's [176] examination of how people react in an underground train station to different types of fire alarm, including different Tannoy announcements and verbal cues from the emergency services. Zinke et al. [177] performed a similar study, in which commuters in underground stations were asked to imagine an emergency scenario, and then evacuate the underground station. Their results indicated strong group effects, where people maintained cohesion within certain groups, and then showed altruistic behaviour in all groups, providing solid supporting evidence for the SCT.

However, the results of these experiments often have limited validity, and consequently reduced applicability. This is because the ecological validity of such an experiment runs counter to the controlled nature of the variables examined, and the extraneous effects that are often present due to the experiments taking place outside of a laboratory and/or with members of the public. These extraneous effects are not repeatable, representing significant confounding factors for the subsequent analysis. There are also further issues that can provide barriers to successfully performing a field experiment, including logistical complexity and financial constraints.

2.4.6. Video footage data extraction

Video data has a huge potential to obtain vast amounts of reliable, quantitative and ecologically valid data. This datasource is especially important, as the subjects within the dataset are often not conscious of the fact they are being observed, and consequently their responses are less 'staged'. Automated tracking allows researchers to access this data in a repeatable and accurate way, while simultaneously reducing the workload on researchers that would otherwise be required for manual tracking. Boltes and Seyfried [178] present a seminal paper on tracking pedestrians from video footage, identifying ellipses within the footage. This provides a solid basis for automated tracking, and therefore accurate data extraction.

The main source of real-world video data is CCTV footage. However, while CCTV cameras exist in large numbers across the majority of the world's major cities, there is a lack of available data for researchers, due to complex issues of ownership and data protection, and in particular the sensitivity of video data of pedestrian responses in hostile emergencies such as terrorist attacks. Moreover, CCTV data is often sub-optimally placed for data extraction, while any non-positional data (e.g. demographic) is usually based on a qualitative assessment of the researchers.

An example of the type of CCTV footage analysis that has been performed for day-to-day behaviour include the classification of dyad and triad social groups [179], and the corresponding movement types. For emergency scenarios, Tsai et al. [25] use video footage to calibrate a pedestrian dynamics model (ESCAPES), combined with an emotional transmission model. By analysing the difference in pixel location of several agents within CCTV footage of two separate events and in the model, they evaluated the performance of the behavioural model, and the behavioural transmission model. Similar approaches have been used examining video footage of real and simulated movement during earthquakes [180, 181, 75, 182, 74, 73].

2.4.7. Survey Methods and Hypothetical Choice experiments

Survey methods are regularly used as an inexpensive and logistically simple way of obtaining data on responses to emergencies. These are usually phrased as hypothetical choices, where participants are provided with written or graphical information and asked what their responses might be in these situations. The information provided might relate to the density and flow direction of nearby pedestrians, and the questions might ask about the specific exit choice of the participant.

Haghani and Sarvi [183] and Shiwakoti et al. [184] have each utilised this approach on train station occupants, asking hypothetical questions about a variety of scenarios, and a varying level of knowledge about the local area, including information such as direction of pedestrian flows and the density of the flows.

An extension to this methodology utilises narrative structures to provide participants with new data at multiple points during the survey, depending on the participants previous actions. Lawson [129] developed a new test using this extension, providing a low cost, repeatable test which can be used to examine how subjects imagine they will react in an emergency.

However, this research method does not produce the required level of ecological validity for generating reliable data. Establishing this degree of reality presents an ethical obstacle, as well as potentially financial and logistical issues. Research has been done to establish the accuracy of the results obtained [185], but large questions still remain surrounding the applicability of this methodology.

2.4.8. Summary

Previous experimental efforts to measure behavioural responses to emergencies have suffered from many limitations, especially when the results from these efforts have been used in pedestrian dynamics simulations. Technological advances in the past decades have led to improvements in video capture, positional and physiological monitoring, and computational analysis. Prior to these, any research was severely limited in the applicability of its results.

More recently, a large amount of the experimental work performed in this area has focused on animal experimentation, which suffers from a lack of ecological validity, psychological and emotional quantification. Furthermore, these approaches make assumptions about the similarity between animal and human behaviour which have not been proven valid. Another commonly used experimental approach is the hypothetical choice experiment, which also suffers from ecological validity, along with a number of cognitive biases that need to be measured.

Previous research that has used real-world observations also suffers from limitations, including incompleteness. CCTV footage is rarely optimally placed for data extraction, and only shows part of an unfolding incident. Further to this, the data gathered in these instances is sensitive, leading to a lack of demographic information on any individuals within the incident.

There are several possible experimental paradigms that could overcome the limitations detailed above. Field experiments provide the opportunity to investigate behavioural responses in emergencies, combining controlled environments and ecological validity. However, these are logistically complex, and can be prohibitively expensive. Another paradigm that could overcome these limitations is virtual reality, which can accurately represent scenarios with complete repeatability and optimal data collection. However, this paradigm has not yet been validated for data collection.

2.5. Terrorism

Terrorism is a multi-disciplinary research field that has typically been dominated by qualitative research, with few quantitative approaches available. This PhD aims to quantitatively investigate microscopic-scale behavioural responses to hostile emergencies.

2.5.1. Background

There is a significant amount of literature on terrorism, focusing on its history and current manifestations. Reference texts (e.g., [2, 30, 186]) often cover its origins within the Zealots (∼10-80 AD) and Assassins (∼1100-1250 AD), before describing state terrorism in the French revolution (notably the Reign of Terror, 1793-1794), and then moving on to modern examples.

Historically terrorism has been characterised by small, targeted attacks, using fear to achieve a secondary aim of publicising a specific cause or achieving a political goal. This has included the use of assassinations and kidnappings used against high-profile individuals. Terrorism across the twentieth century largely falls within this definition, including the acts preceding the Russian revolution, the First and Second World Wars, the Troubles conflict in Northern Ireland, and state terror within the USSR.

However, modern-day terrorists have diversified their methods in many ways. Neumann [30] argues that the key characteristic of contemporary forms of terrorism is the aim to injure and kill as many civilians as possible, in as publicly visible a way as possible. This differentiates contemporary terrorism from previous historical forms, where the targets would typically have been public officials or military targets, and the aim would not have necessarily been the physical harm of bystanders. Another differentiating feature in modern terrorism is the use of multiple attack vectors, including explosives, firearms, bladed weapons and vehicles. For example, the incident in London's Borough Market, 2017, which included two separate attack vectors (vehicles and bladed weapons).

Commonly reported recent terrorist incidents fall into religion-based terrorism (e.g., al-Qaeda, Islamic State of Iraq on the Levant, also known as ISIL). However numerous examples exist of non-religion-based terrorist groups and terrorist acts. Previous reports by the Center for Strategic and International Studies [187] and a joint report by the Nation Institute and the Center for

Investigative Reporting [188] in the USA found that the majority of terrorist attacks in the USA over the past two decades were right-wing-inspired terror incidents. A further European example of a non-religious-based terrorist group is the Euskadi Ta Askatasuna (ETA), a Basque separatist movement active from 1961 to 2011 [189].

2.5.2. Definition

It is necessary to acknowledge the great variety in the approaches taken when studying terrorism, complicated by the vast number of definitions used (e.g., [186]). These generally define terrorism as the use or threat of violence in order to achieve political aims. However, there are disparities in definitions regarding the size of the forces involved (for example, insurgent cells with relatively few members, to full nation state forces), the targets of this violence (for example, a distinction is often drawn between civilian, political and military targets) and the methods used. Easson and Schmid [190] compiled an extensive list of more than 250 different definitions from academic, governmental and intergovernmental organisations, organised chronologically from 1794 (Robespierre, the French revolution and the 'Reign of Terror') to 2010 (Tinnes), (Appendix 2.1, [190]).

For the purposes of this research, this PhD will take direction from the taxonomy and definition developed by Chaliand and Blin [2], detailed below in Table 2.3, who describe terrorism as a method of violence distinct from conventional and guerrilla wars. This definition identifies terrorist attacks as specialised and violent attacks carried out by small, non-state actors on political and public targets. These attacks are carried out worldwide, without defined borders and with covert operatives.

In addition to this definition, this PhD will only consider terrorist attacks that specifically aim to immediately harm members of the public, thus excluding possible specialised incidents such as kidnapping and hostage taking. This definition is adopted for several reasons. Firstly it covers the range of terrorist attacks that have been seen in the recent years, and is therefore the most pertinent for contemporary applications. Secondly, since this PhD will focus on the behaviours of people affected by these types of attacks rather than would analysing the hostile stressing agents, it is out of this research's scope to consider the wider debates about terrorism, such as the ethical implications

Metric	Conventional war	Guerrilla war	Terrorism
Unit size in battle	Large (armies, corps, divisions)	Medium (platoons, companies, battalions)	Small (usually fewer than ten persons)
Weapons	Full range of military hardware (air force, armour, artillery, etc.)	Mostly infantry-type light weapons but sometimes artillery pieces as well	Handguns, hand grenades, assault rifles, and specialised weapons (e.g., car bombs, remote control bomb, barometric pressure bombs)
Tactics	Usually joint operations involving several military branches	Commando-type	Specialized [sic]: kidnapping, assassinations, car bombing, hijacking, barricade-hostage, etc.
Targets	Mostly military units, industrial and transportation infrastructure	Mostly military, police and administration staff, as well as political opponents	State symbols, political opponents, and the public at large
Intended impact	Physical destruction	Mainly physical attrition of the enemy	Psychological coercion
Control of territory	Yes	Yes	No
Uniform	Yes	Often	No
Recognition of war zones	War limited to recognised geographical area	War limited to the country in strife	No recognised war zones; operations carried out worldwide
International legality	Yes, if conducted by rules	Yes, if conducted by rules	No
Domestic legality	Yes	No	No

Table 2.3.: Characteristics of Terrorism, Guerrilla War, and Conventional War as Modes of Violent Struggle. Adapted from [2], page 26.

and the differentiation between State and non-state sponsored terrorism.

2.5.3. Current research and approaches

Current research within terrorism studies often focuses on characterising the trends of modern attacks, and the risk profiles represented by certain factors. The overlap between engineering and terrorism studies has produced several different avenues of research, including improving the defensive capabilities of structures (e.g. against explosive blasts [191]), and detection of hostile individuals [192].

As part of the official response to the threat of terrorism, The UK Security Services have distributed official guidance under the "Run, Hide, Tell" campaign, informing members of the public how they should act if they are involved within a terrorist incident [193]. This campaign represents a crucial example of a pedestrian dynamics intervention, as it specifically informs people of the desired movement characteristics. However, there is a notable absence of quantitative pedestrian movement data when considering terrorist incidents. This is partially owing to the sensitive nature of the data required for model calibration, as well as the fact that the type of analysis (i.e. pedestrian dynamics) that would use this data has only become widespread relatively recently, and has so far largely focussed on everyday behaviour. This lack of quantitative data has meant that the available interventions are simultaneously broad and non-predictive.

There have been many reviews of terrorism literature, including that of Schmid and Jongman [194], who classified the research according to several themes and found that there was an over-reliance on secondary material and a lack of primary sources of data. In 2001 Silke [195] repeated this research, finding similar limitations and decrying the lack of scientific and specifically statistical approaches, with researchers instead focusing on reframing freely available literature.

More recently, Schmid [190] analysed potential future avenues of research. In a stakeholder survey engagement to understand the current shortcomings in the literature on terrorism, he noted that '[modern research] has little or no practical application' (page 468), 'few studies combine theoretical analysis with empirical research' (page 469), and finally 'there is very little empirical work of true relevance.' (page 469).

Santifort et al. [10] established the trends of contemporary attacks, finding that the distribution of attack vector has changed over the last few decades (shown in Figure 2.11), with individual assassination less prevalent now, while the number of bombings has increased. Pool Reinsurance [196] regularly assesses the risk profile to the UK and within Europe, recently classifying the terrorist threat level as ‘severe’, with the most likely type of incident being a bladed or vehicle based attack, aimed at crowded places, symbolic or iconic buildings, and police, military, or government buildings.

Human factors engineering approaches within terrorism research are relatively rare, with a key example of the potential benefit is shown by Eachus et al. [192], who use hormonal, physiological and pheromone measures to quantify the effect of hostile reconnaissance. They propose this as a method of remote detection of hostile actors, which would act as an early warning system.

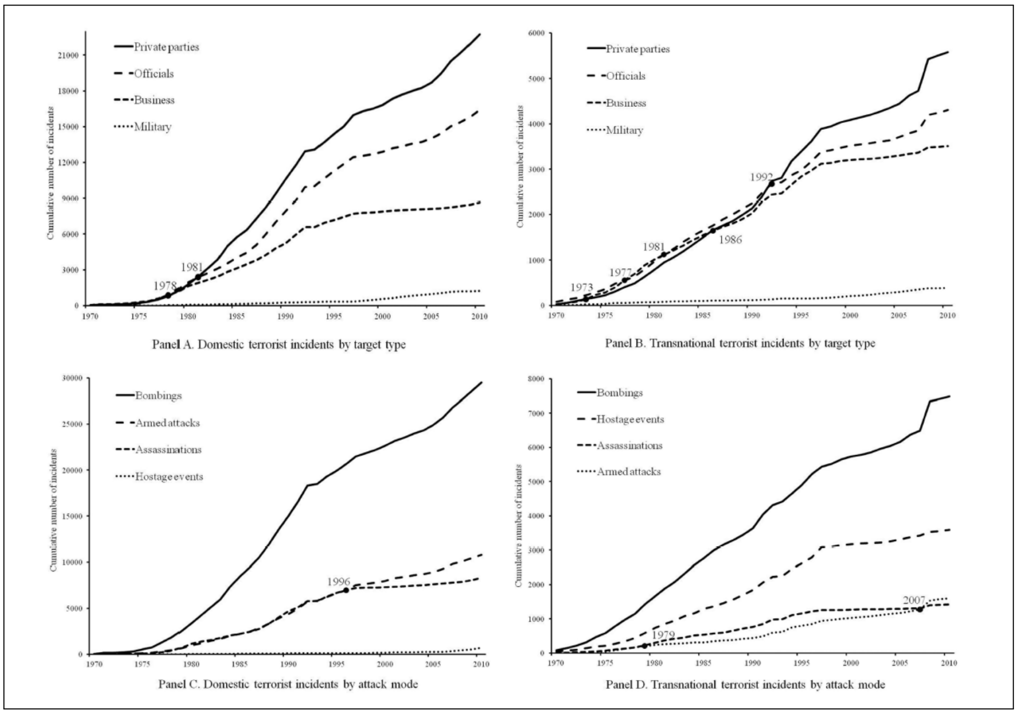


Figure 2.11.: Cumulative incident breakdown by attack mode and target type (from [10])

Further work indicates the growing number of terrorist incidents around the world [197], based on data provided by the Global Terrorism Database

[11]: there were fewer than 1,000 incidents a year before 1977, increasing to 11,999 in 2013. This study also notes the increase in lethality of the attacks, coinciding with the rise of the ISIL.

Finally, there has been recent overlap between the fields of pedestrian dynamics, virtual reality and terrorism. Examples of these were discussed above, in Section 2.1.

2.5.4. Summary

Modern research on terrorism has largely focused on qualitative work, aiming to understand the trends and causes of terrorist incidents. The few studies that broach the overlap between terrorism research and engineering concepts have mainly focused on structural elements, with very little human factors work being performed. There is clearly, therefore, a large opportunity for research that encompasses human factors engineering approaches and terrorism.

There will be references to specific terrorist attacks throughout this thesis, used as examples to clarify any points. These are detailed below in Table 2.4.

Date	Location	Details
20th March 1995	Tokyo, Japan	Five members of the Aum Shinrikyo group released sarin gas on the Tokyo Metro, killing thirteen people.
11th September 2001	New York, USA Virginia, USA	Nineteen members of the Al-Qaeda group hijacked planes and crashed into multiple targets, including the World Trade Centre (New York) and the Pentagon (Virginia), killing over 2900 people.
26-29th November 2008	Mumbai, India	Ten members of the Lashkar-e-Taiba group attacked twelve separate locations around Mumbai, using guns and bombings, killing 174 people (including nine of the terrorists).
22nd July 2011	Oslo and Utøya, Norway	Fjotolf Hansen (previously Anders Breivik) detonated a vehicular bomb in Oslo, before attacking a youth camp on the island of Utøya, using guns while posing as a police officer, killing 77 people.

Date	Location	Details
22nd May 2013	Woolwich, London, UK	Michael Adebolajo and Michael Adebowale, acting alone, killed Lee Rigby on the street in London, using knives.
1st March 2014	Kunming, China	Eight assailants attacked a crowd at Kunming railway station with knives, killing 31 people and wounding over 140.
13th November 2015	Paris, France	Seven members of ISIL began targeted attacks at five locations across Paris, using guns and bombings, killing 130 people.
14th July 2016	Nice, France	Mohamed Lahouaiej-Bouhlel, a member of ISIL, drove a truck into a crowd in Nice, France, killing 86 people, before being shot and killed by police.
17th June 2015	South Carolina, USA	Dylann Roof, acting alone, attacked members of the Emanuel African Methodist Episcopal Church with a gun, killing nine people.
22nd March 2017	Westminster Bridge, London, UK	Khalid Masood, acting alone, drove a car into pedestrians on Westminster Bridge, before moving on foot to attack people with knives. He killed five people overall before being shot and killed by police.
22nd May 2017	Manchester, UK	Salman Ramadan Abedi, acting largely alone, left a bomb in Manchester Arena, detonating as people left the venue, killing 23 people.
3rd June 2017	Borough Market, London, UK	Three members of ISIL drove a van into pedestrians on London Bridge, before moving on foot to Borough Market and attacking people using knives. They killed eight people overall before being shot and killed by police.
12th August 2017	Virginia, USA	James Alex Fields Jr., acting alone, attacked a crowd in Charlottesville using a car, killing one person.

Date	Location	Details
1st October 2017	Nevada, USA	Stephen Paddock, acting alone, attacked a crowd at the Las Vegas Route 91 Harvest music festival from the 32nd floor of a hotel using a gun, killing 60 people before committing suicide.
4th March 2018	Salisbury, UK	Sergei and Yulia Skripal were unsuccessfully targeted for assassination by the A-234 Novichok nerve agent. The British government accused Russian state actors of the attack.
15th March 2019	Christchurch, New Zealand	Brenton Tarrant, acting alone, attacked the Al-Noor Mosque in Christchurch using guns, killing 51 people.
24th February 2020	Toronto, Canada	An unnamed 17-year old, acting alone, attacked a massage parlour in Toronto with a knife, killing one person.

Table 2.4.: Example incidents

2.6. Interviews with key stakeholders

As part of the initial objective of this PhD, several interviews were conducted in 2017 with key stakeholders in the area of resilience, terrorism and research in human factors and pedestrian dynamics. A summary of these interviews was published in [3].

These interviews aimed to fully understand the current use of human behavioural models and pedestrian dynamics models within the context of emergencies. They also investigated open areas of research within this area, as well as any industrial demand for specific pieces of work. Nine semi-structured interviews were conducted, taking a mean of 33.4 minutes (range: 23.7-48.9 minutes). They were transcribed and then grouped according to common themes.

The interviewees were chosen according to their work in the area of human reactions during emergencies, and/or their position as stakeholders in the area of resilience and pedestrian dynamics. Their job titles and descriptions are detailed in Table 2.5.

Interviewee	Job title and description
1	Resilience Practice Leader, Engineering Consultancy
2	Resilience Planning Manager for UK City Council.
3	Social psychologist: "my work primarily focuses on risk perceptions and behavioural responses to extreme events (such as terrorism)".
4	Researcher in "psycho-social aspects of the management of emergencies and disasters, in particular looking at human behaviour in mass emergencies".
5	Industry pedestrian dynamics practitioner and part time researcher.
6	Reader in social psychology: "Behaviour in emergencies and disasters is one of my research areas and so is behaviour in pedestrian flow. I specialise in crowd behaviour".
7	Pedestrian dynamics modeller and consultant on industrial projects, and "also involved in research related to crowd modelling".
8	Associate Lecturer in evacuation simulation and human behaviour in fire.
9	Senior Adviser, UK Security Services.

Table 2.5.: List of interviewees and job roles

2.6.1. Interview questions

The interviews aimed to establish the extent to which the interviewees used human behavioural considerations in their day-to-day work, and specifically whether their work involved potentially fearful scenarios. The questions are detailed below.

Section A: Job description

- What is your current main sector of research/industry?
- How does your work relate to crowd/pedestrian movement and emergency evacuations?

Section B: Human behaviours involvement

- How do you incorporate human behaviour into your work?
- More specifically, pedestrian and pedestrian group dynamics?

Section C: Use of fear and desired progress

- How do you incorporate fear in your work?
  - If your work is related to emergencies, how do you incorporate behavioural responses with respect to this?
  - If you think that you could make use of an improved model of behavioural response to fear, what would be the most useful tool you would use?

2.6.2. Responses

After establishing the job description of each interviewee in Section A, the interviews investigated the extent to which each interviewee used the concept of human behaviours in the roles, and specifically in relation to fearful stimuli. Their responses are detailed below.

Section B: Human behaviours involvement

Section B sought to understand each interviewee's area of work, and the extent to which it included individual and group behaviour during emergencies, alongside pedestrian dynamics concepts, models and approaches.

There was near-unanimous agreement between all interviewees that they needed to consider human behaviour when performing their job, the only exception being that Int. 1 did not routinely use human behavioural concerns in their counter-terrorism work, focusing instead on structural design. However, Int. 1 then went on to say that behavioural considerations were very important when working in resilience in general.

In their roles as researchers, Int. 5, 7, 8 all used pedestrian dynamics models daily for a variety of purposes, including validation and crowd modelling, noting that each type of model has its benefits and drawbacks. Int. 3, 4, 6, 8 all regularly performed research into different human behaviours, either by investigating previous emergency scenarios, or by developing experiments to test and develop models of behaviour. Int. 3 and 4 commented that their work is aimed at providing insight into the behavioural responses seen during extreme emergencies, and consequently at informing policy decisions and guidance regarding these scenarios. Int. 2 stated that they used behavioural concepts in their work in resilience alongside the operational and structural responses, as this is a required element of the planning (such as communication methods, cordon zones, etc.), and as such they are also required to train in behavioural science as part of their Continuous Professional Development (CPD). Int. 9 routinely considers human behaviour in their work in counter-terrorism, attempting to improve the resilience and preparedness of the UK via multiple approaches such as event management and building and infrastructure design.

Group dynamics are implicit in the work of Int. 3, 4 while Int. 5, 6, 7, 8, 9 noted that their work can specifically include the concept of groups (or crowds). Int. 5 notes the difference in capability between macroscopic and microscopic models in replicating group behaviour, with microscopic (Agent Based) models able to incorporate the most intricate concepts, and therefore being the most computationally expensive. They then comment on the limitations with these models: "You don't have models that have a default behavioural system that predicts what people will do. You as the user have [to decide] what proportion of things happen". Int. 5 also comments on the current limitations of modelling group behavioural theories in emergencies with models such as Emergent Norm Theory (ENT) and Protective Action Decision Model (PADM), saying that they could be replicated in computer models, but "you'd have to make big assumptions about what people would do". Int. 6 comments on how the different theories have treated group behaviour in emergencies, citing ENT, the Social Attachment (SA) model and Self Categorisation Theory (SCT). They note that ENT and SA "seem less good [than SCT] at explaining collective co-ordinated behaviour amongst strangers [...]. There are studies showing [...] that even when there aren't people that we know, people still behave in [a] socially structured way".

Section C: Use of fear and desired progress

This section sought to understand the interviewees' work on behavioural reactions in states of stress and negative emotion, and to determine where the focus should be for improvements in this area.

The primary theme was the interviewees' unanimous agreement that a better understanding of human behaviour during fear-inducing scenarios would be beneficial, either for mitigating the effects of an emergency (Int. 1, 3, 5, 7, 8, 9), or for informing guidance for the emergency services (Int. 2, 4, 9). From a computer modelling viewpoint, Int. 5, 7, 8 all confirmed the lack of an accurate emotional modelling (including fear) in various software models, but also mentioned a number of implementations that could reasonably be used for this purpose. Int. 5 referenced Hollmann et al. [51] on utility optimisation and itinerary adjustment "which presumably could then be used in terms of modelling other emotions such as fear".

A further unanimously agreed theme was the requirement for high-quality data for future work. This varied between the interviewees: e.g., Int. 2 specifically requested more case studies, for a lessons-learned style of analysis. By contrast, Int. 3, 5, 7, 8 wanted more basic quantitative data in order to improve modelling capabilities, by producing theoretical approaches and improving software models. Int. 1, 9 stated that a better understanding of how people react would be beneficial in providing guidance on design processes and test cases within the context of terrorist attacks. Finally, Int. 4, 6 prioritised an improvement to the methods of gathering data, as a way of improving the ecological validity of the data and the opportunity to examine a wider range of scenarios.

Finally, a common theme in several interviews (Int. 3, 4, 6, 8) was that an appropriate level of fear was sometimes a prerequisite for adaptive and preservative behaviour. For example, it is commonly observed that, during floods and fires, people will move slowly, if at all, until the threat is confirmed (e.g. visually). However, in the context of mass decontamination, there is often a level of reaction that is unhelpful (e.g. running through the decontamination unit without being properly processed). These are two examples of behavioural reactions that lead to suboptimal outcomes. Int. 8 went further by stating "The problem is to get people out of the [burning] building" and observing that the popular concept of panic is a "false myth", since "in the majority of cases we have the opposite problem". This was neatly summarised by Int. 3: "Fear isn't necessarily always something that is negative but in the context of emergency planning, we have to think of fear as something that needs to be managed".

These themes are key findings for this PhD, which aims to provide a quantitative method of analysing behavioural responses to terrorist attacks. For a computational model to be reliable and validated, it needs to be calibrated against relevant data, which can take numerous different forms, including experimental work, CCTV data and interviews.

2.6.3. Interview summaries

Int	**Section A key points**	**Section B key points**	**Section C key points**	**Key area of desired progress**

Int	Section A key points	Section B key points	Section C key points	Key area of desired progress
1	Works on resilience, including on counter-terrorism "bomb proofing" and "in-vacuation"*. Other areas of the company work on people flow modelling, including in hospitals and security queues. "Consulting, specifically in the area of resilience for businesses and cities."	Routinely uses human behavioural considerations in resilience work. Does not use human behaviour in counter-terrorism work.	Suggests that providing guidance will allow for an improvement in the design of buildings. Also, the development of a predictive model will allow for more in-depth analysis of critical infrastructure. An example would be design case guidance to protect against a Marauding Terrorist Firearms Attack (MTFA).	Predictive model and guidance for architects.

Int	Section A key points	Section B key points	Section C key points	Key area of desired progress
2	Planning emergency procedures and preparedness for UK city, including cross-organisational responses. “I am responsible for planning major emergencies that can impact the city, and ensuring the council is able to respond to them.”	Does not use pedestrian dynamics models, but needs to include likely human behavioural responses in plans. Training involves continuous professional development (CPD) in latest behavioural considerations.	Further to answers in Section B, an improvement would be further case studies of emergency planning and reactions- more data would help predict key issues.	More data (case studies)
3	Studies human behaviour in emergency scenarios as a research area.	Aims to produce studies on human behaviour that can inform policy and guidance.	Fear is something that needs to be managed, as it is often needed for a behavioural response, but too much can produce unexpected results.	Improved modelling capability. More data.

Int	Section A key points	Section B key points	Section C key points	Key area of desired progress
4	Studies human behaviour and management of emergencies as a research area.	Has worked on operational considerations in mass decontamination events, including behavioural considerations. Aims to produce studies on management of emergencies and human behaviour that can inform policy and guidance.	The difficulty is that when doing experiments, you cannot be sure that the subjects are showing the same behaviour as in a real scenario, and you can only go so far to increase the reality of the experiment. Qualitative research is carried out to enable data to be triangulated across different research methods. However there is still a need to develop experimental studies with high ecological validity.	More data and improved research capabilities

Int	Section A key points	Section B key points	Section C key points	Key area of desired progress
5	Works as a senior consultant in modelling pedestrian dynamics and flows. "Senior consultant, in pedestrian and evacuation dynamics [...]. As an academic role I research with various international institutions in this same area."	All their models (macroscopic and microscopic) have some form of behavioural assumptions, so all work involves human behaviour. Also works as a researcher, aiming to improve the quality of models.	There is a way of including emotional modelling in the software, by way of variable itineraries. However, this is not a true fear response. By collating responses, and then statistically representing them as likelihoods, you can edit models to mimic fear responses. Improving the theory of how people respond will also be helpful.	Theory development and improved data.

Int	Section A key points	Section B key points	Section C key points	Key area of desired progress
6	Works as a researcher in social psychology, looking at crowd behaviour in emergencies.	Works at verifying and modifying crowd psychology theories of crowd behaviour during emergencies. This includes laboratory and field experiments and interviews.	Any computer models need to take into account the sociological and crowd psychological theories of behavioural responses, or risk being inaccurate. Fear is a huge variable in response to these scenarios, so any test needs to include it as much as practicable.	Improved testing capabilities

Int	Section A key points	Section B key points	Section C key points	Key area of desired progress
7	Works with commercial and in-house pedestrian dynamics software, analysing large scale events, including for real time prediction of safety concerns and evacuations. Also works with researchers developing models.	The behavioural inputs are user specific. The models themselves can be altered to replicate a certain type of behaviour ("any behavioural element is done by the modeller").	The use of fear is unquantified in current models- and it can have a large effect on what people do. There is speculation on how fear transmits from one person to another, but without validation, the work is still based on modeller experience. If this information were available, a large impact would be the ability to improve crowd management ("the sort of information you need to provide the crowd [...] how you need to guide them").	"What effect does fear actually have", from a quantitative point of view, including movement speeds, directions and emotional interdependence.

Int	Section A key points	Section B key points	Section C key points	Key area of desired progress
8	Works in academia as a researcher in human behaviour during emergencies, and simulating this behaviour. Works with international agencies to validate current models.	Has conducted data collection for human behaviour during emergencies, both in fire and other emergencies. Understanding human behaviour is integral to approaching this field, and to modelling evacuations. There are several considerations when gathering data on human behaviour in emergencies- key being ecological validity.	Many models do not explicitly incorporate fear- "not because it can't be done, but because there is a limited understanding of the psychological basis of the consequences of fear". The specifics of modelling fear might include spatial navigation algorithms, for example, avoidance behaviour	"My answer is very simple: data". Basic research will provide calibration for models.

Int	Section A key points	Section B key points	Section C key points	Key area of desired progress
9	Works as part of a counter-terrorism unit within the UK Security Services.	A huge part of their work revolves around classifying the types of behaviour to expect within emergency scenarios, specifically terrorism. This is the behaviour of both an aggressor and a civilian.	Use fear in their work, but "not in a scientific way"- the aim is to improve the design of scenarios (e.g., buildings, events) such that their use is intuitive when people are scared. This understanding of how people behave is integral to the planning for emergencies.	"A better understanding of fear". This will enable testing of threat scenarios, and assess different designs.
* The term "invacuation" refers to the use of buildings and structures, such as a stadium, as temporary accommodation within extreme scenarios (for instance a terrorist attack). This poses a complex issue, as there may be a requirement for quick ingress or egress in buildings that are often designed to improve flow in only one direction				

Table 2.6.: Summary of stakeholder interviews, from [3]

2.7. Chapter summary

This literature review examined several areas of research that were relevant to human behavioural responses in emergencies, and found significant interactions and relationships between these areas. These are shown graphically below in Figure 2.12.

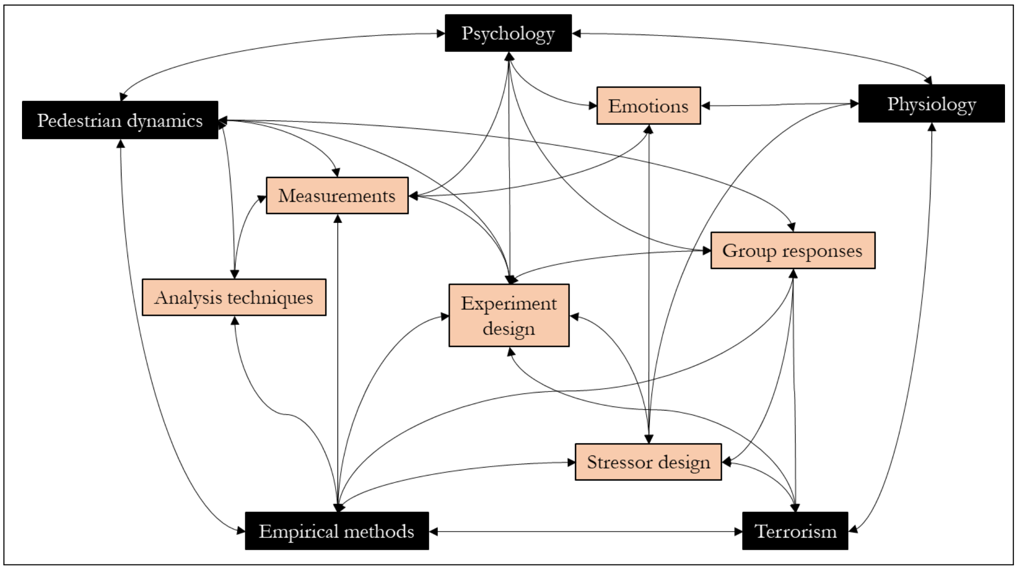


Figure 2.12.: Literature review topic relationships

This literature review has described the concepts and papers that are pertinent to investigating how people react during emergencies. These areas are summarised below in Table 2.7.

Area	**Previous research summary and limitations**	**Opportunities**
Pedestrian dynamics	Modern simulation techniques are extensive, but have focused on day-to-day modelling, and are not able to predict these responses to emergencies. This predictive capability requires advances in theory and large amounts of data.	The simulation of human movement within emergencies can save the lives of those involved.
Psychology	These can be contentious and are not quantitative models. For a valid, predictive theory, this area needs more data and an experimental approach.	A valid, quantitative theoretical model will allow predictions of the effects of different interventions, such as crowd management techniques and infrastructure design.

Area	Previous research summary and limitations	Opportunities
Physiology	This area is required for a valid investigation into human behavioural responses to emergencies but suffers from the lack of defined measurements. Any experimental approach will need to be gather as much data as possible to compare different measures and see which is most appropriate.	By taking as many simultaneous measures as possible, the relative benefits of each can be quantified as methods of studying emotional responses.
Empirical methods	Previous research that experimentally investigated behaviour in emergencies suffered from severe limitations, including a lack of ecological validity, relevant measurements and appropriate stressors.	The development of an appropriate experimental technique would allow progress in all areas.
Terrorism	Previous work has focused on qualitative questions, and therefore suffers from a lack of quantitative understanding.	A quantitative understanding of behavioural responses to terrorist attacks will allow the development of the overlap between engineering and terrorism research.

Table 2.7.: Chapter summary

This review has revealed that while there has been considerable research on both theoretical and empirical fronts, there is a lack of a predictive model of human behaviour within a stressed cognitive (or fearful) state. It has also highlighted that significant fundamental research is required before individual and group behaviour in emergencies is sufficiently well modelled in computer software.

Finally, this review detailed several interviews with key stakeholders in relevant industries: resilience, pedestrian dynamics, counter-terrorism and emergency planning.

The outcome of this chapter is the completion of Objective 1, which aimed to critically analyse the state of the art. This chapter has performed this analysis, as well as investigating the demand for further work, from both an academic and an industrial perspective. The interviews with stakeholders verified many of the outputs from the literature review, and also revealed that there is a significant demand for an improved understanding of behaviour within emergency contexts, with a particular need for more high-quality data.

Of significant note is the need for more overlap between human factors, pedestrian dynamics and terrorism research. The previous lack of this may be due to an inability to produce reliable results in ethically acceptable experimental scenarios, or the supposed intractability of the challenge of quantifying the behaviour of people involved in emergencies.

As a result of the work detailed in this chapter, this PhD determined that the further objectives would be to develop, implement and analyse experimental procedures that mimic hostile emergencies. These experimental procedures would be as generalisable as possible, and the analysis would be tailored to form useful results for pedestrian dynamics simulation models. These would include:

- Trajectories
- Emotional responses
- Sociological effects

In the next chapter, this PhD will detail the key principles to follow when designing an experimental procedure that investigates movement behaviour in emergencies, including the state-of-the-art technology that can be used to take measurements.

3. Methodology: Key Principles

The previous chapter discussed the academic background of the required topics for investigating movement behaviour in hostile emergencies. Several definitions and a theoretical basis were adopted, which will be used for the remainder of this PhD. Using these, this chapter will detail the specific areas to be considered when developing experimental techniques for generating experiments for the investigation of behavioural responses to hostile emergencies. This will include any extension to the academic principles discussed in the previous chapter where necessary, as well as practical examples of how to obtain various datasets. These practical examples will include a description of the ideal dataset, a best-case or 'gold-standard' practical option, as well as further sufficient options. References will be provided to further literature where relevant.

The different areas considered in this chapter are shown in Figure 3.1, split into ethical considerations, stressor design, experimental paradigm, measurements and validation techniques.

While some of these areas are fundamental to the design of an experiment, initially this chapter will detail how these elements are related to each other. The next section will discuss the most important factors when designing an experiment: the ethical considerations. This discussion will include consideration of the various methods of ensuring experiments are both ethically viable and ecologically valid. There will also be a particular focus on informed and presumptive consent.

The next section will investigate stressor design, as the approximation of the 'hostile' element of the experiment is key to obtaining ecologically valid data, i.e. the extent to which the experiment produces data that is representative of effects seen in real-world events. Previous methods of producing this element, and their findings are discussed below, including an analysis of the different desirable effects and an outline of the stressor design process. This chapter will also investigate the potential benefit of pilot studies to improve both stressor design and ethical viability.

The next section will discuss the measurements that must be considered, alongside various methods of obtaining these datasets. The measurements in this PhD have been categorised into three sections: Internal Subjective, Internal Objective, and External. Internal measurements will focus on individual responses to the experimental procedure; the Internal-Subjective measurements will focus on those elements which are inherently difficult to compare between participants, such as emotional state. Conversely, the Internal-Objective measurements will focus on the measurements that are easier to produce comparisons, such as heart rate. External measurements will focus on those elements which can be obtained from sources that are separate to the participant, for example video footage. Specifically, the External measurements will investigate the position and orientation of the participants. These measurements will also be discussed in the context of specific experimental paradigms (namely, VR and physical experiments).

Finally, this chapter will discuss ways to provide validation for experimental approaches.

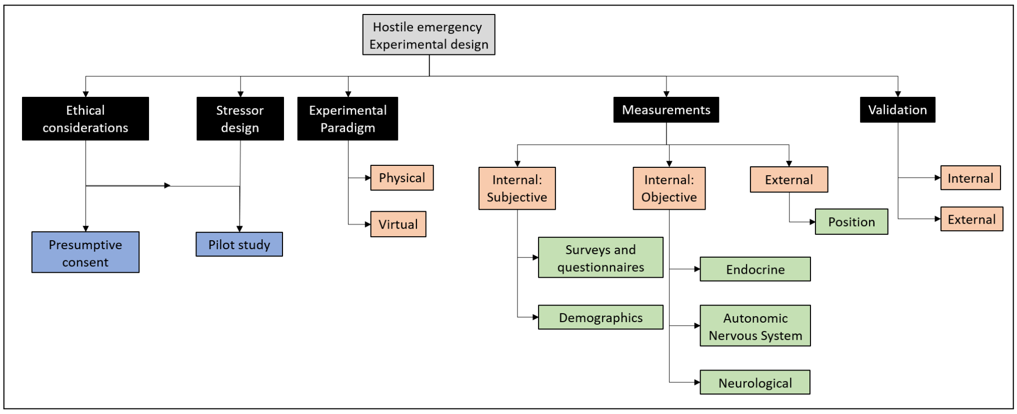


Figure 3.1.: Individual considerations for experimental design

This chapter aims to complement general experimental design principles (such as participant recruitment and randomisation, endogenous bias, etc.). For more details, the following resources provide detailed insight into more generalised considerations [198, 199, 197]

The literature review covered the theoretical aspects of what will be discussed in this chapter and produced definitions that will be used for the remainder of this thesis. However, there remain further aspects of this research which are reliant on previous studies, such as the practical aspects of various concepts are important. These considerations have been noted throughout.

3.1. Design process

A process diagram is shown in Figure 3.2, which provides an insight into the design of the proposed types of experiment. The primary consideration is the ethical aspect of such experiments, and all elements of the design process both affect, and are affected by, the ethical considerations.

The ethical factors determine the limits of the stressor design, and therefore the ecological validity of the experiment. They are also a consideration in choosing the experimental paradigm, and therefore in the different potential measurements.

A pilot study, and presumptive consenting procedure can aid in establishing the ethical viability of an experiment, and also allow the investigation of certain experimental approaches (such as different stressor designs).

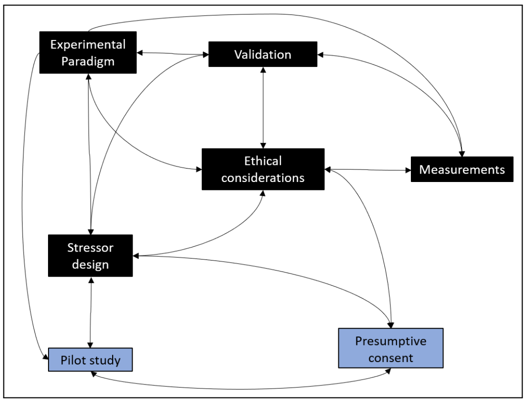


Figure 3.2.: Principles of experimental design

3.2. Ethical considerations

Ethical concerns are arguably the most important consideration when developing experimental procedures that investigate responses to terrorist attacks, as one high-profile element of this research is that, for ecological validity, experimental participants need to experience similar conditions to those undergone in hostile scenarios. It is specifically the risk of harm in these conditions, and the subsequent feelings of fear, anxiety, and stress, that run contrary to ethical guidelines. In order to satisfy ethical considerations, the participants must be protected from the risks linked to these scenarios. Furthermore, the possibility of longer-lasting harm experienced from hostile scenarios (such as PTSD) must be avoided during experimental research.

There are numerous resources available as guidelines for ethically acceptable experimental procedures. It is considered out of scope to detail the in-depth academic research on ethical concerns, and the author refers the reader to several sources for this purpose ([200, 201, 202]). The key considerations for these types of experiment are:

Risk to participant

The most important part of the ethical review is the risk to the participant, which must be as low as reasonably practicable (ALARP). This risk includes both physical and psychological harm, and any risk above day-to-day levels must be justified for the experimental procedure. There are processes available to ensure that the risk to the participant is both understood and minimised, varying from performing risk assessments (a minimum requirement for all experimental activities) to performing prior tabletop experiments and pilot studies.

For the ambulatory hostile experiments discussed in this chapter, the typical risks are:

- Slips, trips and falls
- Negative physical and mental responses to increased fear and stress

It is also necessary to consider if there any unlikely side-effects of such an experiment, as they will be expected to be able to minimise the risk all such events. These might include:

- Aggressive responses to any hostile actor
- Competitive behaviour between participants
- Extreme medical emergencies in response to heightened fear

Risk to the institution

The experimental design must also consider how the research is perceived from an external point of view. The experiment should not reflect negatively on the authorising institution, either from the perspective of the public or the participant, as the resulting damage to the reputation of the institution can be significant and have long-lasting material effects.

Benefits of data

Some data is only obtainable with the acceptance of some risk, as is commonly the case during initial medical trials on humans. In these instances, the data obtained needs to be valuable enough to justify this risk, that is, the benefits of the data must outweigh any risks. The trade-off between the risk to the participants and the potential benefits of the obtained data governs the ethical viability of the experiments.

When examining the potential benefits of the data, it is necessary to consider if the data is obtainable using a different methodology, and if so, whether the resultant risk to the participants could be reduced. Finally, for ethically viability the experimental design must provide a high probability of obtaining useful data, to avoid exposing the participants and institution to risks that result in unusable data.

3.2.1. Informed consent

All experimental processes with human subjects must include a consenting procedure, whereby participants are informed of the nature of the study prior to participating and are given the opportunity to refuse to participate. However, it is sometimes the case that fully informed consent will bias any obtained data, therefore limiting the usefulness of the experiment itself. This is often the case during psychological experiments, and in these cases the consenting procedures must be designed to inform the participants of the nature of the experiment as much as is practicable without introducing systematic errors into the experiment. This may include the generalising or omission of experimental details. In these instances, the experimental design must provide sufficient consenting procedure in order to satisfy internal ethical review guidelines.

As an extension of this type of consenting procedure, the omission of experimental details is considered different to providing misleading information that deliberately deceives participants of the true aim of the experiment, which would require a far higher level of justification. This approach will not be undertaken during this PhD, as in this instance it provides no further benefit to the experimental design, while requiring a higher level of ethical justification.

The types of hostile emergencies investigated within this PhD are unannounced and unexpected (for further detail, see Section 2.5), necessitating this form of limited consenting procedure. In subjecting participants to conditions designed to evoke feelings of fear, the experiment exposes participants to potential psychological harm, as well as any physical risks inherent in the experiment.

Therefore, the experimental design must have the minimal amount of risk, and must limit the information provided to the participants such that they are not biased by the experimental design, and primed to produce specific responses. This approach needs to be balanced with still providing enough information for participants to provide informed consent, and while still obtaining relevant data from the procedure.

Presumptive consent

Presumptive consent is a powerful tool for confirming that research involving stressing agents can be performed while still ensuring that participants are not put at unnecessary or inappropriate risk. Presumptive consent also permits the experimental design's ethical acceptability to be checked, from both the public and participant perspectives.

To obtain presumptive consent, members of the target demographic (usually sourced from the general public) are provided with information on the entire experiment, including any risks, the methodology and the desired data. These individuals are then asked if they would hypothetically take part in the experiment. Further presumptive consent procedures can ask these individuals if they would suggest any changes to the experimental methodology, especially if it means they would otherwise refrain from taking part. If, at the end of the presumptive consent procedure, these individuals agree that they would take part in the experiment, the investigators can presume that the experimental participants would do so as well.

3.2.2. Pilot study

This research suggests utilising an adapted presumptive consenting procedure and pilot study as a design tool for these types of experiment. This builds on prior work performed by Lawson [129], which used open narratives as a method of generating data in emergencies.

Lawson provides a method of understanding human behaviour in emergencies by asking participants to describe a hypothetical scenario of a fire in their house. The investigators in these experiments provide minimal prompting, and only add further details where necessary, or where certain trigger conditions have been met. This method provides useful insight into the thought processes that occur during these scenarios but lacks the necessary stressor to provide ecological validity. It is intended as a method of quickly obtaining results, using minimal resources, to help identify areas that require further analysis.

This PhD used this approach as a pilot study to understand the potential impact of the experimental design on the participants, and their interpretation of the instructions and information communicated to them. This achieved two aims: firstly, ensuring that any given instructions are not open to interpretation by the participants, improving the efficacy of the experiment. Secondly, it allowed an iterative experimental design, updating it to optimally investigate responses to specific stressors. Finally, this PhD performed a presumptive consenting procedure, to determine the level of ethical acceptability of the experiment.

A full description of this pilot study is detailed further in the next chapter Section 4.3.

3.3. Stressor design

When designing an experiment to measure human behaviour in response to stressful scenarios, it is necessary to consider the nature of the stressor, as it is expected that different responses will present in the face of different stressful stimuli.

This is a significant part of the experimental design, as it is usually the case that experiments cannot simply recreate the conditions of a hostile emergency; ethics committees will not allow experiments where participants believe they are at risk of significant harm. Therefore, the stressor utilised in these experiments must be designed to recreate these conditions as faithfully as possible, while still remaining ethically viable.

For the purpose of designing a stressor to mimic hostile scenarios such as terrorist attacks, it is necessary to define the attack vector and any associated characteristics. A complicating factor is that modern terrorist attacks are often comprised of multiple attack vectors ([203]); however this PhD will only consider individual vectors for experimental work. This was decided as the combination of multiple vectors would make it impossible to determine the responses to each individual stressor, and it is considered that vectors that are characterised separately (such as a gun and a knife) would result in very different behavioural responses.

This PhD defines the characteristics of terrorist attacks below in Table 3.1, followed by a characterisation of typical attack vectors.

3.3.1. Attack characteristics

Characteristic	**Vector description**	**Key examples**
Intensity	Defines the level of intensity of the emotional and physical responses.	High intensity: blade, firearm, bomb, vehicle Low intensity: non-painful forms of chemical/ biological/ radiological/ nuclear (CBRN)
Unannounced or declared	Notice is provided of the attack ahead of time, or not.	Almost all modern terrorist attacks are unannounced. Previously, there was often warning (e.g. bombings during the Troubles in Northern Ireland).
Physical nature: tactile or remote	Is transmitted by touch or from a distance.	Tactile: knife. Remote: gun.

Characteristic	Vector description	Key examples
Physical nature: painful or pain-free.	Causes immediate pain, or does not immediately affect the people involved.	Painful: almost all modern terrorist attacks. Pain-free: some CBRN, some gas-based attacks.
Mobile or stationary	The attacking agent is either able to move or not.	Mobile: vehicle. Stationary: Bomb.
Binary or progressive	The danger posed is either constant or incremental.	Binary: Knife/gun/vehicle. Progressive: CBRN.
Number of stressing agents	Attacks can have any number of hostile agents.	Individual: Bombing, single Marauding Terrorist Firearm Attack (MTFA) Multiple: multiple MTFA.
Number of targets: individual or group	Can target either selected individuals, or all members of a group.	Individual: assassinations, kidnapping. Group: bombing.
Non-selective or targeted	For group targets, either specific subsets of a group, or all members are equally at risk.	Non-selective: bombing, arson. Targeted: race-based attacks.
Temporal scale: short or sustained	The timescale over which the incident occurs.	Depends on the incident: see below.
Non-competitive or competitive	The survival capacity of individuals doesn't come at the cost of others, or does.	Depends on the incident: see below.

Table 3.1.: Attack vector characteristics

Intensity

The terrorist attacks considered in this PhD consist of aversive phenomena (such as explosions or firearm use) that induce particularly high levels of stress and fear in both the people involved and the wider public. This is typical of modern terrorist attacks which seek media publicity through high impact attack vectors [30].

The induced levels of fear and the consequent responses are partially a product of the potential for pain and injury through these attack vectors. Therefore a key counter-example of the high intensity terrorist attack is a specific use of non-visible and pain-free CBRN agents such as A-234 Novichok, used in Salisbury, UK in 2018. The responses in nearby people during this attack were very different to those seen during prior knife-based attacks.

Unannounced or declared

Modern terrorist attacks are typically unannounced, with the aim of producing as many casualties and causing as much destruction as possible. However, as discussed earlier in Section 2.5, there are some key counter-examples, including the tactics often utilised by paramilitary groups in Northern Ireland, where authorities were often notified of upcoming attacks with enough time to evacuate civilians.

A further aspect, which is out of scope of this PhD, is the concept of a decoy declaration, whereby attacks are announced in order to divert the attention of the emergency and security services to a location that presents no threat, while attacks are carried out in other locations.

Physical nature of the stressor: tactile or remote

Tactile stressors (e.g. a bladed attacker) are characterised by the aggressor needing to be within reaching distance of the people involved, whereas remote stressors (e.g. a gun) can affect people from a distance. It is suggested that tactile threats produce significantly different behavioural responses to remote threats.

Physical nature of the stressor: painful or pain-free

The physical nature of these stressors can be further subdivided into pain-inducing or pain-free. For example, some CBRN agents are not immediately painful, whereas other attack vectors (e.g. knives) are. It is suggested that these different stressors produce significantly different behavioural responses.

It is possible to induce a pain stimulus in experimental work, for example through electric shocks [157]. However, it is not ethically possible to do this without first informing the participants, and calibrating this to the participants' own pain thresholds. To introduce a painful element to the experimental paradigm will therefore inevitably remove the unannounced element of the experimental design.

Mobility of the stressing agent: mobile or stationary

The mobility characteristic varies from immobile (e.g. a bombing, or a stationary aggressor), to non-sentient mobile (e.g. CBRN) to sentient mobile (e.g. MTFA), which are capable of targeting and adapting their movement responses. The mobility of these stressors has a significant affect on the movement responses of the people involved.

Binary or progressive

Another consideration which will impact the stressor design is the level of interpreted threat, and how it develops through time. For example, some threats (e.g. MTFA) are clear and obvious, resulting in more immediate behavioural responses. However, other threats are less obviously present (e.g. pain-free CBRN), allowing far more decision time for the people involved.

It is suggested that this difference will result in significantly different behavioural responses, as well as result in these responses being produced in a far shorter time span.

Number of stressing agents

One of the major features of modern terrorism [30, 203] is that attacks are more often carried out by 'lone-wolf' agents (for example Nevada, USA, 2017). These agents are far harder to detect, track and ultimately prevent, and typically have access to vast resources from the internet. In rarer cases large groups of agents take part in a simultaneous attack (for example Paris, France, 2015). This PhD only considered individual aggressors.

Scale of the stressor: individual or group

Depending on the vector, attacks can also be characterised by their ability to target either single individuals (e.g. an assassination, Salisbury, UK, 2015), or multiple individuals (e.g. a weaponised plane, New York, USA, 2011).

This characterisation might result in very different behavioural sets, as was seen in the Woolwich, London, 2013 incident, in which observers filmed the attack, in contrast to the responses seen by pedestrians caught in the Paris, France, 2015 attacks, who typically hid or evacuated.

Temporal scale: short or sustained

The duration of the incident can have a large effect on the responses seen, and can also impact the other characteristics. For example, the Westminster Bridge, London 2017 incident lasted 82 seconds, whereas the Salisbury, London, UK, 2018, incident required several months of decontamination before the area was declared safe [204].

Targets of the stressor: targeted or indiscriminate

It is also possible to characterise a stressor as targeted or indiscriminate, where the attack vector will either represent a threat to a subset of a population, or the whole population. Following the definition of terrorist attacks (see Section 2.5), contemporary attack vectors are typically indiscriminate, with a few counter examples of targeted attacks. While all incidents will necessarily have a secondary aim (often political), the targets of the incident may be either the public (indiscriminate) or a subset of the public (targeted). A further relevant sub-characterisation is the sentient nature of the stressor, indicating whether the stressor can decide which target to approach next.

Recent examples of indiscriminate group-based attacks have occurred worldwide, in Oslo and Utøya Norway (2011); Christchurch, New Zealand (2019); Toronto, Canada (2020); Paris, France (2015); and Borough Market, London (2017). However, there are also examples of group-based, targeted incidents in white-supremacist terrorist attacks in the USA (e.g. South Carolina, 2015). Example of a non-sentient vectors include fire, or aerosol-based CBRN material (e.g. Tokyo, Japan 1995, or Salisbury, UK, 2018) .

Competitive nature of the stressor

The literature review on sociological theories in emergencies detailed how immediate competition arises from a perceived “zero-sum-game” nature of a scenario, in which the targets of an attack consider that the survival of other targets represents a decrease in their own chances for survival. This is expanded in entrapment theory (see Section 2.2), where competitive and panic behaviour results from:

1) An immediate, pressing threat to the individual's life.

2) An opportunity to escape, which is diminishing quickly.

3) A feeling of helplessness in otherwise dealing with the threat.

There are further examples of competition within these events, from civilians wanting aggressors to focus their attention on other people, to them wanting to reach a place of safety before others. It is worth noting, that even with these forms of competition, helpful behaviours are commonly observed within these scenarios, which has led to the development of self-categorisation theories.

To summarise, it is arguable that there is always a level of competition between civilians within modern terrorist attacks. However, this PhD will not explicitly create competitive stressors due to the rarity of events that are accurately characterised by a zero-sum-game stressor. Instead this research will allow all participants an equal and constant opportunity of successfully completing the experiment.

3.3.2. Typical attack vectors and associated characteristics

Following the taxonomy developed by Chaliand and Blin [2], and the definition of modern terrorism provided by Neumann [30] in Section 2.5, this PhD will consider the following single attack vectors for terrorist attacks: blade, firearm, vehicle, bomb and CBRN.

These attack vectors can be described using the characteristics described above, as shown in Table 3.2. In this description, the different attack vectors can be characterised in differing ways. For example, a bladed attack is a high intensity, tactile, mobile and binary attack vector. However, it can be carried out by individuals (e.g. Oslo and Utøya, Norway, 2011) or groups (e.g. Borough Market, UK, 2017), and can be targeted (e.g. Woolwich, UK, 2013) or non-selective (e.g. Manchester, UK, 2017).

The taxonomy above is aimed at classifying modern terrorist attacks, and is therefore not exhaustive when considering incidents that do not fall under the definitions adopted in this PhD (see Section 2.5). This taxonomy might lead to ambiguity or incompleteness in cases such as kidnapping or hostage-taking of individuals, in which scenarios the aggressors' primary aim is not the harming of individuals, but rather the coercion of third parties. As Neumann

Attack vector	Description	Typical characteristics	Possible characteristics	Key examples
Blade	One or more stressing agents armed with blades.	High intensity. Tactile. Mobile. Binary.	Individual or group. Non-selective or targeted. Non-competitive or competitive.	Woolwich, London, UK 2013. Borough Market, London, UK 2017.
Firearm	One or more stressing agents armed with firearms.	High intensity. Remote. Binary.	Individual or group. Mobile or stationary. Non-competitive or competitive.	Paris, France 2015. Las Vegas, Nevada, USA 2017.
Vehicle	One or more stressing agents in vehicles.	High intensity. Tactile. Mobile. Binary Group	Non-selective or targeted. Non-competitive or competitive.	Nice, France 2016. Charlottesville, Virginia, USA 2017.
Bomb	One or more explosive stressing agents.	High intensity. Stationary. Remote. Binary. Group. Non-selective	Targeted (e.g. mail bombs)	Paris, France 2015.
CBRN	Non-distinct widespread stressing agent.	Remote. Group Non-selective.	High or low intensity. Binary or progressive. Mobile or stationary. Competitive or non-competitive.	Salisbury, UK 2018.

Table 3.2.: Characterisations of typical attack vectors

[30] notes, this is an example of 'old' terrorism, and will not be considered in this PhD.

3.3.3. Possible proxy stressors

As detailed in Section 2.4, there are numerous experimental methods of eliciting emotions, with face-to-face methods considered the gold standard when establishing stress and fear. This contrasts with previously accepted methods, such as pictures (International Affective Picture System [105]), computer screens, and audio methods. Additionally, there has been significant previous work examining responses to multiple types of stressor, including physical (e.g. electric shocks [157]), psychological (e.g. imprisonment [205] or coercion [206]) and financial ([127]).

A common experiment used to evoke stress is the Trier Social Stress Test ([207]), where participants are told they are to take part in a public-speaking exercise, and then provided with a surprise task from which their performance is measured, and often provide measures of anxiety including self-assessed and ANS measures [133]. However, this test is meant to understand the effect of anxiety-led stress on mental acuity, and therefore is not appropriate for investigating movement responses to fearful stimuli.

Further work has examined how participants respond to financial incentives, which is a commonly utilised in combination with other stressors (including game theory and electric shocks [157]). However, there is sparse literature on the potential implications of the design of financial stressors on social interactions. It will be argued in Chapter 4 that an appropriate stressor when investigating a blade-based attack is the use of an unexpected hostile actor, who represents the potential for financial loss by way of tactile contact. This financial stressor will be binary and independent of the performance of other participants.

3.3.4. Summary

To create a valid experiment, it is important to accurately define the stressor under consideration, which must also mimic the characteristics of the terrorist attack under investigation. It is possible to use proxy stressors (indeed, it is necessary in most physical experiments), and as such the proxy stressors should satisfy the same component characteristics as the attack vector under investigation.

3.4. Experimental paradigms

3.4.1. Physical experiment design

The following general considerations

Setting

When designing an experiment with groups of participants, the environment needs to be controlled and well defined. It needs to be large enough to accommodate the entire group, and ideally free of extraneous effects that could distract or impact on the measurements (e.g. members of the public). This could be a single room, or multi-room environment, or indeed outdoors. This will therefore also be dependent on several further aspects, such as the weather or other extraneous effects such as transport access.

Timing

The time of day is an important factor when considering participant attendance levels, as well as when considering stress level measurements. For example, cortisol levels are strongly dependent on an individual's circadian rhythm, and therefore the results of an experiments may depend on the time of day that measurements are taken. It is suggested that any repeat experiments are performed at the same time of day, in as short a time period as is practicable. This will also impact the visibility levels of an experiment if it is performed outdoors, which will have a significant impact on participants (especially visually impaired participants).

Distractor tasks

If the stressor is unannounced, the participants will need to take part in a distractor task during the initial part of the experiment. Depending on any financial reward system, the participants must be able to readily stop focusing on this distractor task. One option is to provide activities for the participants, but explain that they have no ultimate importance for the experiment. For further information and a key example, the reader is directed to [208].

Size and demographic of group

The demographic of a participant can be defined a number of ways, including age, gender, profession, nationality, and many more. While the size and demographic of the experimental group is limited by recruitment, experimental setting, funding, and attendance, it is important to record as much information as possible to control for any effects seen as a result of participant demographic.

For large group sizes, it is possible that endogenous bias can be introduced by the non-attendance of participants. An expensive solution to this would be to use one participant, with the remainder of the group consisting of actors. This would furthermore allow the investigation of specific variables such as the demographic of the group, and has been performed in previous psychological research (e.g.: [209]).

Logistics

To ensure that the experiment is repeatable, the standard practice is to perform rehearsal experiments, as well as ensuring that all communication to participants is identical across experiments (e.g. by using scripts).

3.4.2. Virtual reality experiment design

Virtual reality (VR) provides an opportunity to perform highly realistic, repeatable experiments for emergencies that would be difficult or impossible to achieve in physical scenarios. There are multiple types of VR, including head-mounted displays (HMDs), which will be discussed below.

VR experiment development is similar to video game development, and as such it is necessary to consider numerous factors to ensure a level of realism and immersion. There are several unanswered research questions surrounding the use of VR as a data-generating tool, especially when considering behaviour in emergencies, which were detailed in a Strengths-Weaknesses-Opportunities-Threats (SWOT) [1] analysis. This analysis provides powerful insight into the use of VR as a research tool for investigating human behaviour in emergencies.

Further experimental considerations are detailed below, and for further information the author encourages readers to view previous work on Serious Games (SGs) in VR, such as [144, 208, 210].

Game engine

To start developing a VR experiment it is first necessary to select a game engine that can be used to create the virtual environment and run the scenarios. There are many options available with VR capability, with the most well-known open-source engines being Unity Engine [211] and Unreal Engine [212]. Both of these have active communities that provide resources and documentation for developing experiments, as well as online tutorials and free licencing options for researchers and individual users.

Storyboard

When developing a VR experiment, it is necessary to ensure that the user is aware of what goals they are being asked to achieve and the actions available to them. This can be implemented in several ways, but the primary method is to have an initial 'training' section in which participants are introduced to the environment, the controls and the communication methods (typically either verbal, visual or recorded audio).

This section should allow the participant to practice any planned interactivity within the experiment, including via controls or movement triggers. This training session has a further added benefit of allowing the participant to accommodate themselves in the virtual environment, and therefore minimise any motion sickness they might otherwise experience (see below for further details).

Non-player characters

Non-player characters (NPCs) are avatars within the virtual environment which are controlled by the computer rather than the participant. NPC behaviour can be extremely complex, limited only by the coding of the environment and the computational cost. There are multiple open-source options for developing avatars [213] that can be used within VR, including body type, clothing, and animations. There has been significant progress on automating various aspects of NPC behaviour, including locomotion (e.g. using the Unity Engine NavMesh Agent functionality), animation blending (e.g. a waving and running avatar) and 3D-audio effects.

The use of NPCs within the storyline of an experiment can be a very complex issue, as social effects within VR are poorly understood (see below for further information), and the development of the interactions is difficult to implement properly. As such, NPCs are commonly used as non-interactive dynamic elements of a VR environment.

One further element to consider with NPCs is the response they generate in the user as a result of the level of realism of their design. This is known as the "uncanny valley", which describes how participant responses change as the realism of the NPC avatar increases. As a result of this, there is a level of human likeness that should be avoided during experiments (see [214] for more details).

Multiplayer

There are significant benefits to having participants believe that the other avatars are controlled by participants rather than by a computer, as the participants' consequent behavioural and social interactions will be far more reliable and ecologically valid [208]. This is achievable by having multiplayer VR. However, implementing a multiplayer VR environment is more technically complex than doing so for an environment designed for individuals. This must be considered against the skillset of the experimental team. Additionally, multiplayer VR can be expensive as it requires extra hardware (e.g. through two or more HMDs).

Due to these logistical and resource constraints, this PhD instead aimed to achieve reliable social interactions by simply informing the participants that the experiment was multiplayer. This will be discussed in further detail in Section 4.5.

Motion sickness

Motion sickness can occur within VR environments, where participants feel nauseated as a result of conflicting information within their internal senses. Where visual information conflicts with internally sensed movement. This risk is minimised by having a low 'photon-to-motion' time, which is defined as the time difference between the movement of an individual's head and the consequent updated graphical information on the VR screen [215].

There is a risk of this effect for all participants, while those who are unaccustomed to VR can be more susceptible to motion sickness [216]. This can be alleviated by having a period of acclimatisation in the virtual environment while also limiting the overall exposure time and ensuring a high refresh rate on VR screens (minimum 20 Hz, ideal 50+Hz) [217, 218, 219].

Locomotion techniques

Movement in VR is an ongoing area of research ([220]), with typical methods detailed in Table 3.3. Any experiment investigating movement responses should consider how each of these areas produces the desired datasets (for example, continuous datasets preclude teleportation as a locomotion method). There is also evidence to suggest that specific movement methods can mitigate the impact of motion sickness. [216]

The ability for the user to move in VR has a significant impact on the functionality and validity of the environment. The navigation methods cause the movement of the user's avatar in the virtual environment, consequently updating the display shown to the user. However, unrealistic movement techniques can lead to increased motion sickness. Additionally, to obtain valid results in an experiment, the positional data measurements will be strongly affected by the locomotion technique, especially if the movement technique is non-continuous.

As continuous movement responses are required for high resolution data and analysis, teleporting is not appropriate for these types of experiment. This is due to teleporting not providing continuous measures of position, and precluding measures such as velocity and acceleration. Similarly, although joystick or button approaches are an improvement over teleporting, they are also inappropriate. Free movement is an ideal approach, minimising the risk of motion sickness while allowing exact replication of desired physical movement in the virtual environment. This type of movement also includes physical limitations, such as fatigue, which are not present in any other method of locomotion. However, free movement is usually not an option when designing large virtual environments, especially in a multi-room or multi-floor experiments. Instead, arm movement and treadmill approaches are more appropriate, and have been commonly used in previous experiments.

Social effects

As discussed in Chapter 2, social effects in emergencies are an active area of research, and in VR these effects are further complicated by the introduction

Method	Description	Pros	Cons
Teleporting	One or more stressing agents armed with blades. The user points a controller and presses a button, causing their avatar to immediately move to the desired location. This can be used while standing or sitting.	Easy to understand and implement.	Increase in motion sickness. Not a continuous measure.
Joystick/button movement	The user presses a button or directs a joystick, causing their avatar to move through the virtual environment in the specified direction. This can be used while standing or sitting.	Easy to understand and implement.	Increase in motion sickness. Does not accommodate variable speed/ acceleration well.
Arm movement	The user moves their arms as though they are walking, causing their avatar to move the virtual environment in the direction of arm movement (or head direction).	Allows variable speed/ acceleration. Becomes intuitive for user after training. Lower motion sickness.	Harder to implement
Treadmill	The user is placed on an omni-directional treadmill, and their leg movement is tracked, causing a corresponding avatar movement in the virtual environment.	Allows variable speed/ acceleration. Lower motion sickness.	Harder to implement. More expensive equipment required. Long training period required.
Free movement	The user is tracked as they move in the real environment, and the tracked movement is applied to the avatar in the virtual environment.	Gold-standard. Lowest motion sickness. Most realistic approach, no coded movement interface.	Not feasible for majority of virtual environments.

Table 3.3.: Separate navigation methods

of virtual environments and avatars. It is relative poorly understood how participants will act as a group, especially when considering the reactions of a participant to NPCs and other group participants. Furthermore, it is suggested that there will be drastic differences when comparing the responses of a participant to a known NPC, and to those of a participant responding to another participant. Therefore, when considering investigating social effects in VR with the intent of using this data with real-world applications, participants should believe that the avatars they observe in the virtual environment are controlled by humans, and that there is a continuing social element to the experiment that extends beyond the extent of the experiment.

There are several possible ways to achieve this, each with different requirements and associated levels of complexity. The ideal method would be to implement a multiplayer VR environment, which satisfies the belief requirements. Furthermore, ideally the participants should be located in the same experimental venue (rather than, say, connected across large distances by an internet connection), so that there is also the ongoing social element to the experiment. However, as mentioned above, multiplayer VR environments can be complex and expensive.

A second method is to implement NPCs that are sophisticated enough to convince the participants that they are human-controlled. This is a very complex task, requiring the NPCs to pass a form of Turing test [221], and becomes more complex with every added element of functionality. For instance, if the experiment requires communication between participants, it might be possible to implement an NPC that uses responsive AI and natural language processing systems. Examples of this are currently commercially available as AI assistants (such as the Amazon Alexa, Apple Siri or Microsoft Cortana), or the widely available online chatbots. However, these are currently not of high enough quality to fully implement a convincing NPC.

A final method is to limit the interaction between participant and NPCs, while informing the participant that the NPCs are in fact human-controlled. The choice of specific interactions, which can be coded in advance, should be designed to increase the participants' level of belief in the NPCs being controlled by other human participants. For example, a simple waving animation can be implemented, or a pre-recorded audio file played. The participants' level of belief can be increased by implementing 'mistakes' in the interaction, for instance multiple waving animations, or a delay before the playing of the audio file.

3.4.3. Summary

Physical experiments can be performed in a variety of settings, which should be selected based according to the appropriateness for the attack vector in question. There are many further considerations for performing physical experiments, including the fact that there are usually time and space constraints on any procedures. This leads to a requirement for well established project management during the experiments, as any mistakes can prevent data collection.

Conversely, VR offers the opportunity for completely controlled and repeatable experiments, investigating extremely dangerous scenarios within a safe environment, potentially with a smaller logistical requirement. VR paradigm can show identical scenes to real-world scenarios, which will be repeated perfectly for different participants. There are logistic requirements in performing these experiments, including the time taken to develop the environment and the financial outlay for the equipment. However, this offsets the requirements for physical experiments, where experiment times are usually inflexible, and experiment venues can be expensive to hire.

The basic data (e.g. position) obtained by VR experiments can be identical to those obtained within physical paradigms, and additionally there will be minimal noise and interference within the data. However, there are limitations to the types of data obtained within VR experiment, as any complex behaviours need to be implemented before they can be performed by a participant. As an example, participants in physical experiments can choose to hide, or perform crypsis ('playing dead'), but it is much more difficult to allow participants in VR environments to perform the same actions.

However, before this research can be performed, validation work underpinning this paradigm must be carried out. In order to have confidence in the data generated from experiments carried out within VR, there must first be a comparison between data generated in both physical and virtual paradigms, understanding any differences that occur. This PhD performed this analysis, investigating data that was generated in identical experiments in both physical and VR paradigms, finding that there were minimal differences between the two. This provides some of the first evidence that VR is valid paradigm

Section	Measurement	Detail
Subjective	Subjective state	Track participant internal emotional state. Track continuously.
	Demographics	Obtain perfect demographic and experiential context.
Objective	Physiological state	Track ANS (ECG, RR, GSR, skin temperature). Track endocrine response (cortisol, adrenaline). Track continuously (100Hz+, 100kHz+ for ECG).
	Neurological state	Track participant neurological state (EEG, fMRI, NIRS). Track continuously (100Hz+ [222]).
External	Position tracking	Track position and rotation of bodies, limbs and head (x, y, z, orientation, t). Track all participants and hostile actors. Track continuously (50Hz+).
	Eye tracking	Track gaze direction. (x, y, z, t). Track all participants and hostile actors. Track continuously (50Hz+).

Table 3.4.: Theoretical datasets

for generating behavioural datasets for emergencies.

3.5. Measurements

3.5.1. Theoretical datasets

When beginning an experimental design, it is often useful to consider an ideal dataset. The suggested measurements that would represent an ideal dataset when investigating movement responses to hostile threats are detailed in Table 3.4.

Subjective state tracking

Self-assessed measures have been used extensively to measure internal cognitive states, and are used here to provide supporting evidence for participants' level of psychological stress.

Subjective state tracking can obstruct experimental procedures, as it requires the participant to focus on and record what they are thinking and feeling, rather than participating in the experiment. Therefore, while this data would be particularly useful, it will be necessary to measure these states before and after the experimental procedure.

Demographics

Obtaining a complete demographic dataset on participants is a complex problem. It takes time to obtain this data through questionnaires or interviews, and there are several examples of detailed demographic questionnaires available. However, owing to the lack of data regarding behaviour in emergencies, it is not immediately obvious what specific demographic details are important in behavioural responses. Furthermore, these details may have to be obtained after the experiment in order to avoid suggesting the nature of the experiment to the participants, which may bias any data obtained.

Physiological tracking

As examined in Section 2.3, humans have distinct physiological responses to fear-inducing stimuli, producing measurable changes controlled by the different responses in the body.

Key examples of these responses are provided by the autonomic nervous system (ANS), which produces responses in heart rate (HR), heart rate variability (HRV), galvanic skin response (GSR), skin temperature (ST), and respiration rate (RR). Further responses are produced by the endocrine system, which releases different hormones in response to these stimuli, including adrenaline, adrenocorticotropic hormone (ACTH) and cortisol. These hormones are commonly associated with the "fight-or-flight" (later termed the "fight-flight-freeze") response [70].

These response types are produced unconsciously, resulting in an unbiased source of quantitative data. As a result, physiological measurements are commonly used to infer specific emotional states. However, multiple emotional states can produce similar physiological reactions (a many-to-many relationship). These responses vary widely, depending on the individual [116], leading to difficulties when generalising analysis. To identify and compare quantified measures of emotions (such as fear), therefore, it is important to use multiple measures, along with self-assessment measures.

Previous research conducted with physiological measurements was able to distinguish between emotional states of stress in a controlled environment, but was unable to do so during ambulatory experiments [110]. This represents a significant limitation when using these measurements for emotion quantification during ambulatory experiments.

Physiological tracking has become a sophisticated area of industry and research, with multiple commercially available ANS systems (see below for details). However, there are limitations as to which measurements can be obtained throughout an experiment. This could be either because of restrictions on mobility (for example, a blood sample would interrupt the experiment), or a lack of developed tools (such as continuous endocrine system measurements which are not commercially available at time of writing).

Neurological state tracking

Neurological state tracking is particularly well suited to identifying subjective states, as unique areas of the brain will become active during specific emotional responses. This is an area of ongoing research, and there are many existing commercial systems. However, they are typically housed in specialised rooms (e.g. fMRI, PET), and not designed for ambulatory measurement, therefore limiting their utility in this type of experiment.

Mobile EEG and NIRS systems are being developed - indeed, there are some commercial sets available, e.g.[223]. However, these systems are relatively new, and are not yet accurate enough for emotional quantification.

Position tracking

High-definition and high-frequency location tracking is possible, but requires significant pre-experiment time, and is often prohibitively expensive. Examples include motion capture for computer-generated imagery (CGI) or biomechanics research. In these cases, markers are placed on participants' bodies to track, in detail, the movement of specific muscles and bones. These techniques, however, require the participants to wear specific markers on their bodies, line-of-sight to the participants from the cameras, and do not necessarily work particularly well in a multi-room format (for indoor tracking).

3.5.2. Achievable datasets

As detailed above, a perfect dataset is not possible at the moment. This section will describe an achievable dataset (detailed in Table 3.5), with gold-standard

Section	Measurement	Detail
Subjective	Subjective state	Pre- and post- experiment questionnaires
	Demographics	Relevant data obtained prior to experiment, including personality.
Objective	Physiological state	ANS (ECG, RR, GSR, skin temp), tracked at 10+ Hz. Endocrine responses: cortisol measured, baseline (pre-experiment), and 20, 40, and 60 minutes post-experiment.
	Neurological state	Insufficiently advanced technology for reliable ambulatory measurements for this context.
External	Position tracking	Position of head (x, y, z, orientation, t), tracked at 10+ Hz The type of positional measurement will depend on the experimental paradigm. See below for a discussion on both physical and virtual experiments.
	Eye tracking	Gaze direction (theta, phi, t, or x,y,z,t), tracked at 10+ Hz. Eye tracking also depends on the experimental paradigm. See below for a discussion on both physical and virtual experiments.

Table 3.5.: Achievable datasets

measures detailed where appropriate.

3.5.3. Self-assessed measurements

Self-assessed measures are used to quantify subjective states where objective measures are not appropriate or possible, such as measures of emotion, preference, personality, or identity. These self-assessed measures are necessary to perform some degree of quantification of emotional state and social response.

Personality measurements

Personality psychology is a wide-reaching area of research that defines the personality of individuals by multiple factors, including their preferences, motivations and reactions. While an individual's induced internal state change may differ drastically depending on personality in response to specific stimuli, an implication of self-categorisation theory is that personality should have a minimal effect on any elicited responses during hostile emergencies.

One well-established method of measuring personality is the Big 5.

Big 5

The Big 5 model [224], also known as the Five Factor model, or OCEAN, presents personality as a combination of five separate elements, each quantified on a scale of 0-5:

- Openness

The openness factor quantifies characteristics such as creativity, imagination and insight. People who score highly in this category are open to new challenges, while people who score low are more resistive to change.

- Conscientiousness

This factor establishes the level of mindfulness of the individual, quantifying characteristics such as attention to detail and organisation. People with low scores in this category are less organised and dislike schedules.

- Extroversion

Extroversion represents the individuals sociability. A high score in this factor indicates that a person tends to gain energy from being surrounded by others, while a low score indicates a level of reserve in groups.

- Agreeableness

This factor includes characteristics such as empathy, kindness and prosociality. An individual with a low score in this factor would take little interest in others.

- Neuroticism

Neuroticism quantifies the emotional stability of the individual. A high score in this factor suggests that the individual experiences higher levels of stress and is easily upset. A lower score instead represents a relaxed, emotionally stable individual.

The Big Five model quantifies these traits by asking individuals to rate themselves on a scale in response to self-descriptive phrases. This provides a measure of the five separate factors of personality (defined by the model), allowing these to be easily incorporated into quantitative analysis like multiple regression.

Another well-established example of a personality test is the Myers Briggs Type Indicator (MBTI) [225]. The MBTI states that personality will fall somewhere between the extremes of each of these scales, with a preference for one extreme or the other. An individual's responses to the MBTI questionnaire are then compiled into one of sixteen four-letter personality statements (e.g. ISTJ- Introversion, Sensing, Thinking, Judging). However, this is a discrete classification of personality, which can be incorporated into quantitative analysis (e.g. via a multinomial logistic model) but would potentially reduce the explanatory power of any model.

The Big Five personality questionnaire will be used in this PhD for personality quantification. For further details on the extensive academic research surrounding personality measurements the author refers the reader to several sources for this purpose [226].

Fear measurements

Self-assessed measures have been used extensively to measure internal cognitive states, and research continues in this area, with numerous different survey measurements available.

State Trait Anxiety Inventory (STAI)

The STAI [118] is a regularly used method of assessing anxiety, fear and distress, producing a single number. The survey is distributed in a STATE or TRAIT form, assessing either long term or transitory conditions. This is also available as a short form questionnaire, with fewer items for the participant to answer.

Positive And Negative Affect Schedule (PANAS)

The PANAS [119] is another commonly used survey method of assessing anxiety, fear, and distress, producing two numbers (one positive, one negative). This is also available as a short form questionnaire.

Discrete Emotions Questionnaire (DEQ)

The DEQ [117], developed in 2016, aims to identify specific emotions and their relative levels.

Self-Assessment Manikin (SAM)

The SAM [108] asks participants to rate their affective state as a product of three separate factors: valence, arousal and dominance. Specifically, participants are asked to indicate their self-assessed agreement with the emotions represented by set of pictures.

Social response measurements

Another observed effect from extreme emergencies such as terrorist attacks is the onset of a degree of group identification and shared identity amongst the people involved [227]. Measuring this is therefore particularly important for obtaining an indication of the level of 'group stress' perceived by any participants.

Research is ongoing in this area, with a key contribution coming from Leach et al. [228] in developing a group identification questionnaire. Further work has been performed on shared social identities [229].

3.5.4. Physiological measurements

As detailed in Section 2.3, physiological measurements are commonly used to quantify the emotional state of the individual. Several methods have been used to measure emotional states, which can be split into autonomic nervous system (ANS) responses and endocrine responses.

Autonomic Nervous System (ANS) measurement

When using ANS responses in experiments, the typical measurements are:

- Heart rate (HR)
- Heart rate variability (HRV)
- Galvanic skin response (GSR)
- Skin temperature (ST)
- Respiration rate (RR)

There are numerous commercial options available to consider for measurement of ANS responses. Below is a non-exhaustive list of potential options (Table 3.6), including descriptions of the positive and negative aspects of each piece of equipment.

Equipment	Measurements	Comments
e4 Empatica [230]	ST, GSR, Blood Volume Pulse (BVP, inferring HR and HRV). Wrist-based.	Easy to use wrist-based system. Less reliable data than chest-mounted systems.
Equivital EQ02 [231]	ST, ECG (2 channel), RR, GSR (additional). Chest-based	Dense data source with multiple different measurements. Good fit necessary- increases experimental set up logistics. Expensive system.
Polar H10 [232]	ECG (1 channel). Chest-based	Inexpensive system. Limited datasets available. Difficult to use in group format.
Shimmer [233]	ECG (multiple channel), EMG (2 channel), GSR, HR, HRV Multiple-sensor based	Dense data source with multiple different measurements. Expensive system.
Movisens [234]	GSR Wrist-based	Limited datasets available.

Table 3.6.: Example ANS measurement systems

Method	Details	Comments
Saliva	Non-invasive swab or salivette. Non-specialist method.	Commonly chosen for extra-hospital contexts.
Blood (Serum, Plasma)	Invasive and specialist method.	Commonly used within hospital contexts.
Sweat	Non-invasive. No standard method (Choi et al).	Used for specific tests.
Urine	Non-invasive. Non-specialist method.	Increased considerations (ethical, logistical, etc) over saliva.

Table 3.7.: Methods of obtaining endocrine system measurements

Measure	Methods	Comments
Cortisol	Saliva, serum, plasma, sweat, and urine.	Blood considered gold standard. Saliva commonly used in extra-hospital contexts. Salivary cortisol peaks between 20 and 60 minutes after the onset of a stressor.
Adrenaline	Serum and urine.	Blood serum considered gold standard.

Table 3.8.: Endocrine system measurements

Endocrine system measurements

As stated in the literature review, the body produces endocrine responses in the presence of a stressor. However, they are produced in many other scenarios as well, and there is a high variability in the inter-personal levels of different hormones, suggesting that cortisol should be used in conjunction with multiple other measurement sets, and that repeat measures should be used.

Typical endocrine system measurements for stress responses include cortisol and adrenaline. There are numerous methods for quantifying the levels of these hormones in humans (detailed in Table 3.7), which are used to quantify the measurements (detailed in Table 3.8).

When quantifying an individual's cortisol level, the gold standard approach is to sample it directly from the blood. However, this represents a much higher ethical requirement, and also would need trained personnel to obtain the measurements. In the experiments detailed in this PhD the most logistically appropriate measure is to use salivary cortisol, as it is non-invasive and can be readily performed multiple times in short periods by non-trained personnel, and has a well-understood, strong relationship with blood cortisol levels.

Owing to the between-participant variability in cortisol, pre-experiment samples should be obtained as a baseline measurement, and multiple samples should be obtained at regular time intervals after the experiment. Other considerations include the circadian cycle which significantly affects the level of salivary cortisol, so for appropriate between-group comparisons it is necessary to control the time of the experiment.

However, it is not possible to perform salivary measurements for adrenaline, owing to its relatively short life in saliva [235]. Therefore experiments that wish to measure adrenaline may need to use blood serum or urine measurements, both of which increase the ethical and logistical requirements.

Neurological measurements

Measurements of neuronal activity are particularly well-suited to indicating levels of fear or stress, as differentiated to other forms of emotion as these emotions activate unique regions in the brain, allowing their identification and quantification as specific emotional subsets. However, the viability of these measures is limited as most measures are not suited for ambulatory use, preventing them from being utilised effectively.

Typically, measurement of neuronal activity falls into two categories: electrophysical and metabolic. Electrophysical measurements detect the presence of specific phenomena (e.g. an EEG measuring electrical current), and use the position of those phenomena to infer neurological effects. Metabolic measurements detect physiological activity (e.g. a near-infrared-spectrometer measuring oxygen saturation in blood) at specific locations, and use these measures to infer neurological effects. Many different measurements are available, including magnetic resonance imaging (MRI and fMRI), positron emission topography (PET), and electrocorticography (ECoG), but in an ambulatory setting, the only available measurements are the near-infrared-spectroscopy (NIRS) and electroencephalography (EEG).

NIRS

Near-infrared spectroscopy neuroimaging is a method in which electromagnetic waves with a wavelength of 700nm-1100nm are directed towards the brain, measuring blood and oxygen saturation in distinct regions. As metabolic activity increases, the amount of blood in the area increases, and the resulting change in the frequency of the reflected light waves is used as an indication of the activity levels in these regions.

However, this approach can only be used to image to a depth of 1-3 cm into the brain [236] and does not penetrate sufficiently deep into the brain to provide useful information surrounding emotional states (typically measured within the amygdala). This contrasts with other methods (e.g. fMRI) which can image all areas of the brain.

EEG

Electroencephalography (EEG) is a method in which multiple electrodes are placed on the scalp of a participant, which then read electrical signals that come from neuronal activity. Pairs of electrodes are combined into channels, with a differential amplifier providing the difference in electrical activity between the two regions associated with the electrodes. With enough channels, this data can then be manipulated to produce a three-dimensional map of electrical activity within the brain. This map can be used to infer the activation of specific regions of the brain, and therefore provide insight into the emotional state of the participant.

This is the only neurological measurement method with a temporal resolution of milliseconds and better [237]. However, there are serious issues with deep-brain electrical activity localisation, and even in surface level activity there are limits to the spatial resolution.

Other systems

Several other systems exist for quantifying activity within specific regions of the brain. The gold standard approach is functional Magnetic Resonance Imaging (fMRI), which uses high magnetic fields to induce detectable radio frequency emissions from oxygenated blood. As the individual neurons require the glucose and oxygen contained within blood in order to function, the flow of this oxygenated blood indicates which neurons are active. This method has a high spatial resolution and accuracy, in the order of millimetres through the entire cross-section of the brain, but has a low temporal resolution, with

Feature	Minimum requirements
Positional	<1m accuracy in three dimensions.
Temporal	> 15Hz update rate.

Table 3.9.: Minimum requirements for positional measurements

a frequency of up to 10Hz [238]. Moreover, fMRI systems are very large, and therefore not suitable for ambulatory experiments and the large magnetic fields induced render these systems unsuitable for VR experiments.

3.5.5. Positional measurements

Requirements

For the experiments considered by this PhD, accurate, high-frequency tracking is required to ensure that all movement details are recorded, which is especially important when participants are moving at high speed or accelerating at a high rate.

For the purpose of performing experiments, the minimum suggested specification of a positional measurement system is shown in Table 3.9.

Physical experiments

Various technologies are available to measure trajectories, depending on the scale and temporal resolution required, as well as the potential for post-processing of data. Some examples are listed below in Table 3.10.

Ultra-wideband (UWB)

UWB uses large bandwidth electromagnetic signals, along with multiple possible localisation algorithms (TDOA, TOA, etc) to accurately measure the position of chips within an experimental zone. The accuracy achieved depends on numerous factors, including chipset, localisation algorithm, update rate and movement speed of the tracked participant. Gold standard accuracy is quoted as

- 10cm accuracy in x, y, z dimensions
- 1/2degree accuracy for pitch, roll and yaw
- 20Hz temporal resolution

Technology	Maximum accuracy (x,y,z)	Temporal resolution	Comments
Ultrawideband (UWB)	∼10cm	<50Hz	Potential for high quality localisation. Needs to be installed correctly.
Ultrasound	∼1cm	<50Hz	Potential for high quality localisation. Requires line-of-sight, and less widely used than UWB.
Bluetooth	∼1-3m	<20Hz	Potential for medium quality localisation.
RFID	∼10m	N/A	Passive system, only works when participant is near an anchor. Requires fully installed areas
. WiFi	∼3-5m	<10Hz	Potential for medium quality localisation.
GPS	∼5m	<10Hz	Only works outside- inappropriate for indoor positioning.
Camera	∼10cm	<60Hz	Potential for high quality localisation. Requires high quality installation.
Inertial Measuring Unit (IMU)	N/A	<60Hz	Used in most sensor-fused localisation systems. Low quality localisation when used as standalone system. Suffers from measurement drift.
LIDAR	<1cm	100+Hz	Require line-of-sight. Produces large datasets.

Table 3.10.: Typical values for positional measurement systems

The flexibility and potential accuracy of this system, coupled with the fact that numerous technologies (including mobile phones) are equipped with UWB chips (e.g. [239]), indicate that UWB could be a flexible and widely available source of position data.

Furthermore, an important positive indicator for UWB systems is the potential for it to work across a multi-room environment. While many tracking systems require direct line-of-sight between anchors and tags, UWB can theoretically track moving tags behind walls. However, this ability is limited since metal components in the wall will interfere with UWB signals.

Bluetooth

Bluetooth is a commonly used method for transferring information, but a further application is to estimate the distance between anchors and tags. Modern Bluetooth systems (e.g. Bluetooth 5.1 or later) have been predicted to give centimetre-level positional accuracy for indoor positioning [240]. However, the temporal resolution is lower than that of UWB, with a maximum frequency of approximately 10Hz [241].

Ultrasound

Ultrasound has a better potential spatial resolution than either UWB or Bluetooth methods, as well as high temporal resolution, but suffers from a low level of technological proof, and needs line-of-sight between multiple anchors and each tag.

Wifi-based ranging

Wifi-based (or WLAN-based) positioning systems function similarly to Bluetooth systems, but suffer from low reliability and temporal resolution, as well as lower spatial resolution (2-5m) than that of UWB.

RFID

Radio frequency identification (RFID) is a particularly low-cost solution, wireless tracking solution, with no batteries required for the tracked tag. However, it suffers from low temporal and spatial resolution. It is suited to low-cost, large-scale indoor movement tracking, but is inappropriate for the high-accuracy requirements of the types of experiments discussed in this PhD.

Satellite tracking

Global Navigation Satellite Systems (GNSSs, including GPS, GLONASS, BDS and Galileo) are widely available, with potential for high spatial accuracy (especially in differential GPS), and up to 10Hz temporal resolution. However, these systems are only appropriate for external positioning, and become unreliable in high-rise city contexts.

Inertial measuring unit (IMU)

Inertial measuring units (IMUs) provide local measurements of spatial and rotational acceleration, coupled with magnetic and gravimetric measures, to provide a displacement and orientation from a pre-set origin. A perfect IMU would theoretically be able to fully describe indoor movements, but real-world systems suffer from issues such as drift, resulting in compounded errors.

These systems are, however, used as extra data sources in most on-board tracking systems (e.g. UWB, mobile, inside-out VR) for simultaneous localisation and mapping (SLAM) purposes.

Cameras

Cameras provide a gold-standard solution for tracking individuals, owing to their high resolution, high frame rate and low cost. Open-source tracking solutions are available (e.g. PeTrack [178], OpenCV [242]). Numerous camera systems are available, ranging from CCTV to digital cameras and higher cost research tools (e.g. Codamotion [243]).

However, often some pre-processing is required, and the cameras should be placed in optimal positions to ensure line-of-sight in all regions of the experimental area.

Light Detection and Ranging (LIDAR)

Light detection and ranging (LIDAR) is a technology that produces a high accuracy point cloud of the surrounding environment, with centimetre accuracy and a high refresh rate (e.g. 20Hz). However, this technology relies on line-of-sight, requiring complete observation of the experimental area.

Furthermore, LIDAR technology typically generates large amounts of data in the form of point clouds, requiring significant post-processing, and the technology itself can be prohibitively expensive.

Virtual experiments

In VR experiments, it is necessary to track the position (headset position - x, y, z - and orientation) of the participant as well as that of all avatar bodies and any moveable scenery. This can be performed at an arbitrary update rate (limited by the computational cost), but 10-20Hz is suggested as acceptable, as it is comparable to the gold standard available for physical location tracking. Additionally, it is possible to track the position and orientation of any controllers, providing data on the hand position of any participants.

Some headsets (e.g. HTC Vive Pro Eye) also provide the ability to track gaze direction, which could be beneficial for further analysis on the areas in which the participants are looking during the experiment. In the absence of this ability, it is possible to record the observed display and audio for each participant. This would provide data on what information is available to experimental participants, but will be limited by computational cost.

Finally, recent developments in commercial systems have made it possible to track participant body pose and orientation [244]. This allows an improved data collection and user experience, as any physical movements can be replicated and shown on the participant's avatar in the virtual environment, increasing the levels of immersion.

3.5.6. Summary

This section has detailed the different measurements that are possible for the type of experiments considered in this PhD. Table 3.11 shows a comparison between the ideal datatsets, the achievable datasets, and the datasets that were obtained during the experimental procedures.

Section	Measurement	Ideal dataset	Practical dataset	PhD dataset
Subjective	Subjective state	Track participant internal emotional state continuously.	Pre- and post- experiment questionnaires	Pre- and post- experiment questionnaires .
	Demographics	Obtain perfect demographic and experiential context.	Relevant data obtained prior to experiment, including personality.	Relevant data obtained prior to experiment, including personality.
Objective	Physiological state	Track ANS (ECG, RR, GSR, skin temperature). Track endocrine response (cortisol, adrenaline). Track continuously (100Hz+, 100kHz+ for ECG).	ANS (ECG, RR, GSR, skin temp), tracked at 10+ Hz. Endocrine responses: cortisol measured, baseline (pre-experiment), and 20, 40, and 60 minutes post-experiment.	Two e4 Empatica and two EQ02 equivital physiological monitoring sets. Endocrine responses: cortisol measured, baseline (pre-experiment), and 20, 40, and 60 minutes post-experiment.
	Neurological state	Track participant neurological state (EEG, fMRI, NIRS). Track continuously (100Hz+).	Insufficiently advanced technology for reliable ambulatory measurements for this context.	None.
External	Position tracking	Track position and rotation of bodies, limbs and head (x, y, z, orientation, t). Track all participants and hostile actors. Track continuously (50Hz+).	Position of head (x, y, z, orientation, t), tracked at 10+ Hz The type of positional measurement will depend on the experimental paradigm.	UWB and VR position tracking at 10Hz, providing position (10cm accuracy or better) and orientation ($\pm 0.5°$ accuracy
	Eye tracking	Track gaze direction. (x, y, z, t). Track all participants and hostile actors. Track continuously (50Hz+).	Gaze direction (theta, phi, t, or x,y,z,t), tracked at 10+ Hz. Eye tracking also depends on the experimental paradigm.	None

Table 3.11.: Comparison of datasets

3.6. Validation

Validation is used to ensure that the results from experiments can reliably used to predict responses in future events. There are two relevant methods of validation that will be discussed here: internal and external.

Internal validation

The internal validity of the experiment is the extent to which it can accurately assess the desired cause and effect. That is, the internal validity is a measure of the level of credibility of any reported results and the relationships between independent and dependent variables.

For these experiments, internal validation can be provided by the level of agreement of the different measures, in this case the internal subjective, internal objective and external measures. For instance, for experimental participants in a stressful procedure, the expected results would include an internal subjective response suggesting a heightened level of stress (either through PANAS, STAI, or other), and an internal objective response suggesting an internal process responding to a stressor (e.g. heightened cortisol levels, increased heart rate, etc.).

As the data surrounding real-world terrorist attacks is sparse and incomplete, there is no equivalent dataset against which these measurements can be compared. For example, it is unknown how much an individual's heart rate will change during an attack. Therefore internal validation is the only possible option for the vast majority of these measures.

External validation

The external validity of an experiment is the ability to extend any conclusions to real-world scenarios. For the types of experiment described in this thesis, a difficulty arises owing to the lack of available real-world data as a comparison for experimental data. This restricts the ability of the experiment to investigate the external validity of any experiment. Furthermore, for the real-world data that is available, there is often a lack of specific information which would allow a full comparison. For instance, CCTV cameras are often sub-optimally placed, preventing a full picture of an incident.

In the absence of appropriate and freely available datasets for use in external validation, it is instead possible to infer external validation by relying on the validation of the other measures used (e.g. the validation for the STAI, PANAS and cortisol levels as a measure of stress). As external validation has already been performed for these measures, they can be used as validation of these experiments. However, this necessarily only refers to the measures which have been validated, which does not include the positional tracking. Thus any validation of movement responses will rely on the measures of internal subjective states.

Furthermore, partial external validation can be inferred by the use of theoretical constructs. If the experimental research produces responses that fit within the existing theories of human behaviour in emergencies, such as emergent norm theory (ENT) or self-categorisation theory (SCT), it is possible to infer that the experiment produced similar conditions. However, this approach must come with caveats. Firstly, such a validation attempt must include high levels of internal validation. Secondly, requiring a level of agreement between theory and experimental response precludes the possibility of using this data to develop new theories. Therefore, this method of validation should be viewed as secondary to the previous methods.

3.7. Chapter summary

This chapter has established the generalised methodology for investigating behaviour responses to terrorist attacks. The methodology includes establishing the ethical viability of any experiments, characterising the terrorist incident, and specifying the measurements that will be used during these experiments. This chapter has also detailed the different possible options for each of the considerations, including defining both an ideal and a practical dataset. Finally, this chapter also specified the differences between physical and VR paradigms, and how that affects the remaining considerations within the methodology.

The next chapter will apply these concepts to the design of a set of physical and virtual experiments which investigate how pedestrians respond to bladed attackers.

4. Methodology: Knife-based Attacks

The previous chapter discussed the numerous factors that an experiment must consider when investigating pedestrian responses to a hostile aggressor. This chapter will build on this to detail the design of an experiment that investigates movement responses to hostile aggressors armed with bladed weapons (a 'bladed attack'). This experiment was designed to be performed in two separate paradigms, virtual and physical, and was performed in December 2018 (physical experiment) and January-March 2020 (virtual experiment). The implementation process for the design of these experiments is shown in Figure 4.1.

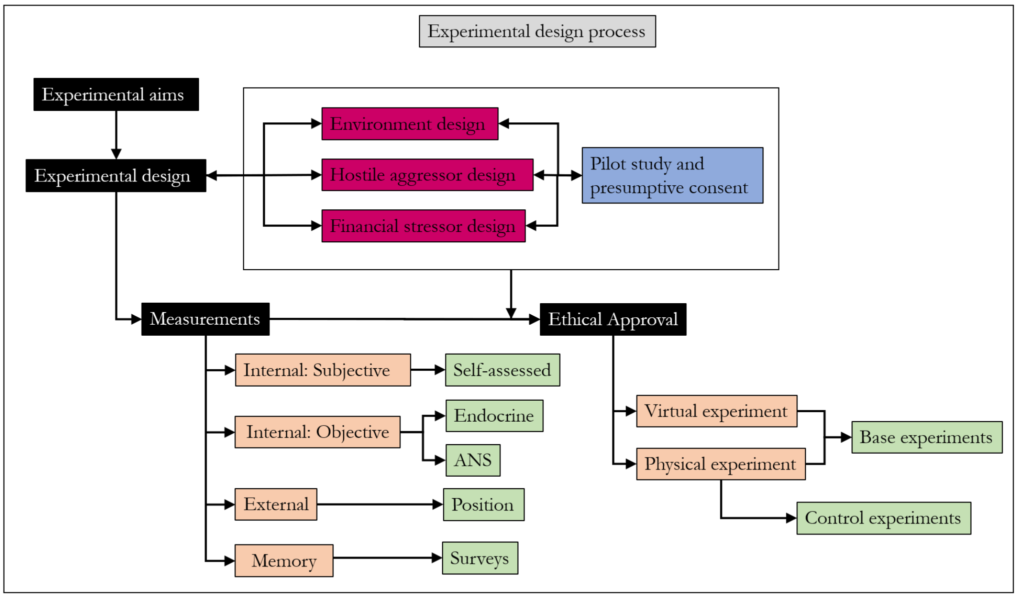


Figure 4.1.: Implementation of methodology

This chapter will detail the aims of this experiment, before examining the stressor design, taking into consideration the threat characterisation of knife-

Measurement	Example
Internal subjective	Surveys of experiences
Internal Objective	Autonomic nervous system responses
	Endocrine system responses
External	Position of participants throughout experiment
Recall quality	Survey of recall of experiences

Table 4.1.: Measurement subsets

based attacks. Next it will detail the results of a pilot study undertaken in conjunction with Public Health England (PHE) and the consequent impacts on physical area and hostile aggressor design.

The remainder of this chapter will describe the exact procedure followed in the two separate experimental protocols, including the measurements obtained, and will specify any differences between the virtual and physical experimental paradigms. When any distinction is required, the physical experiment will be referred to as PR (physical reality) and the virtual experiment will be referred to as VR (virtual reality). Finally, it will detail the investigation into recall quality after stressful scenarios.

4.1. Aims

As a result of the initial critical analysis of the state of the industry surrounding pedestrian dynamics and hostile emergencies (see Section 2.6), one of the objectives of this PhD was to design an experimental protocol to investigate bladed attacks.

These experiments were designed in both PR and VR paradigms, consequently providing a point of quantitative comparison between the experimental paradigms. The measurement obtained during these experiments are based on the descriptions in the previous chapter, and fall broadly into three categories: Internal-Subjective, Internal-Objective and External (Table 4.1).

4.2. Possible experimental designs

As discussed in the previous chapter, the design of the experiment is crucial in determining its crucial validity, and therefore any ability to generalise and extrapolate any results. The first task in the design of these experiments is to

Characteristic	Comments
High intensity	Bladed attacks induce high levels of stress and fear. This is considered the limiting ethical concern.
Unannounced	Following the terrorism definition adopted in Section 2.5, the attack considered will be unannounced.
Binary	Bladed attacks are binary stressors, rather than progressive.
Tactile	Bladed attacks are tactile-based, rather than remote.
Mobile	The hostile aggressor will be mobile.
Number of stressing agents	This PhD will only consider attacks carried out by single agents.
Group targets	This PhD will focus on group experiments as this is more relevant to 'new terrorism'.
Non-selective	All members of the experimental group will be equally likely to be targeted.
Non-competitive	There is no competitive element to these types of attack: the survival of one target does not depend on the non-survival of another, or vice versa.
Painful	This was not incorporated into the design, for logistical and ethical reasons.

Table 4.2.: Bladed attack characterisation

characterise the threat that is being investigated.

This PhD will examine a common example of bladed attacks, specifically incidents where an individual aggressor targets groups of pedestrians. The characteristics of a bladed attacks, based on Subsection 3.3.1, are detailed below in Table 4.2.

Owing to ethical and logistical restrictions, this PhD decided that, while such a bladed attack would be characterised as a painful stressor, this should not be incorporated into the experimental design. This decision was taken as a pain stimulus represents a higher ethical barrier and removes the potential for an unannounced stressor. Instead, it was decided that this experiment would use a financial stressor, combined with an unnanounced, hostile aggressor as a proxy for the bladed-attack vector. This decision was taken because unannounced physical stressors (such as electric shocks) would be unethical, and because the use of a financial incentive is a method of ensuring the imposed stressor is quantifiable.

Characteristic	Appropriate stressor	Comments
High intensity	Financial and aggressor	Limited by experiment funding and ethical concerns. Decided to create a medium-intensity aggressor, who obviously poses no physical threat.
Unannounced	Aggressor	Ensure information provided pre- and during experiment is controlled.
Binary	Financial	An all-or-nothing implementation, rather than partial financial losses.
Tactile	Aggressor	Aggressor must touch participants to cause the any financial loss.
Mobile	Aggressor	Aggressor is able to move.
Number of stressing agents	Aggressor	Single aggressor in experiment.
Group targets	N/A	Experiment looks at group behaviour.
Non-selective	Aggressor	Aggressor selects next target based on proximity only.
Non-competitive	Financial and aggressor	Successful payment does not depend on other participant failures.

Table 4.3.: Stressor characterisation

The previous chapter argued that the stressor must be designed to mirror the characteristics of the attack vector under investigation as far as possible. These characteristics will be accommodated in the design of two separate stressors (financial and aggressor) under consideration, as detailed in Table 4.3.

4.2.1. Financial stressor design

This PhD considered multiple options for financial incentives as stressors, whereby the participants would believe they had the potential to lose (or, more specifically, to not gain) financial earnings. These included:

Lottery systems

A lottery system was considered, in which successful completion of the experiment would allow participants a chance of winning a large amount of money. An extension to this lottery system was also considered, where all participants would have a chance of winning, but successful completion of the experiment would lead to a better chance of winning the lottery. This method would allow for the condensed use of the experimental funding, leading to far higher financial rewards, and therefore a far higher potential stressor. However, this system could potentially introduce an element of competition between participants, which this PhD previously ruled out (see Section 3.3).

Incremental earnings

Another financial incentive paradigm considered was an incremental earnings system, in which participants would accrue financial rewards based on experimental tasks prior to the introduction of a hostile aggressor. Participants would be informed that, if they failed the experiment, they would lose all accrued financial incentives, representing the binary stressor. In reality, all participants would have been paid an equal amount for participation. It was considered that this approach would increase the participation levels within the experiments, therefore inducing a higher sense of potential loss, and therefore a higher potential stressor.

Non-incremental earnings

The final paradigm considered was a binary stressor, in which participants would be informed that they will receive the maximum financial reward upon successful completion of the experiment but would receive no reward upon a failed performance. An extension to this would be the guarantee of a minimal financial reward for all participants, with a larger financial reward gained upon successful completion of the experiment. Again, all participants would in fact have been paid an equal amount for participation regardless of performance.

The lottery option was removed as a consideration, as it was deemed unfair on participants who did not successfully complete the experiment, and it had the potential to create elements of competition between the participants. Both incremental and non-incremental earning systems were taken forward into the pilot study for testing.

4.2.2. Hostile aggressor design

This PhD considered two separate options for the introduction of the hostile agent:

Participant-aggressor

In this paradigm, the hostile aggressor would initially be a member of the participants. They would participate in the experiment before revealing themselves as an aggressor and pursuing participants.

Unseen-aggressor

In this paradigm, the hostile aggressor would be a previously unseen member of the experimental team. They would remain out of sight of the participants for the initial section of the experiment, until after a set period they would enter the environment and reveal themselves as an aggressor.

Both of these options were taken forward into the pilot study for testing.

4.2.3. Environment design

Finally, this PhD also considered different options for the environment of the experiment:

Open room, entrances and exits

The environment was decided to be an indoor environment in order to avoid adverse weather conditions, with space for up to 15 participants to move around with ease. Several options were considered, including lecture halls and exhibition venues. The selected venue was a sports hall, as it was optimally designed for free-movement, and there was the option of semi-permanently installing the position-tracking system.

A 'safe' area

It was considered that a 'safe' area could be provided to allow participants to remove themselves from the experiment if they felt unsafe. This would be tested within the pilot study.

4.3. Pilot Study and presumptive consent

As part of the design process, a pilot study was carried out in order to understand the different interpretations of participants and to ascertain the ethical viability of the experiment via a presumptive consenting process. This pilot was based on previous research that studied data generation for emergencies using narratives [129].

As a result of the initial design process detailed above, the following options were taken forward for assessment in the pilot study:

- Incremental vs non-incremental earnings
- Participant-aggressor vs unseen-aggressor design
- Safe area provision

The pilot study was performed in conjunction with the People's Panel, a dedicated group at Public Health England (PHE). This group focuses on academic engagement and trials with the public, comprising a selection of over 1600 randomly recruited members of the public [245]. The pilot study procedure was developed over a period of several months, with the People's Panel recruiting participants.

Ten members of the People's Panel were interviewed in three separate group focus sessions, performed over Skype, on the 21st and 22nd May 2018. The focus sessions lasted approximately one hour each and took the form of a narrative-based interview, in which the participants were provided with chronologically ordered information on the experimental protocol. Participants would then answer questions and provide their impressions on the experimental design at various stages throughout the interview.

Finally, each participant would provide their decision on whether or not they would consent to be part of the experiment ("presumptive consent"). Unanimous presumptive consent was achieved, with multiple participants stating that an increased stressor would be both acceptable and desirable. The pilot study process is shown in a flowchart in Figure 4.2.

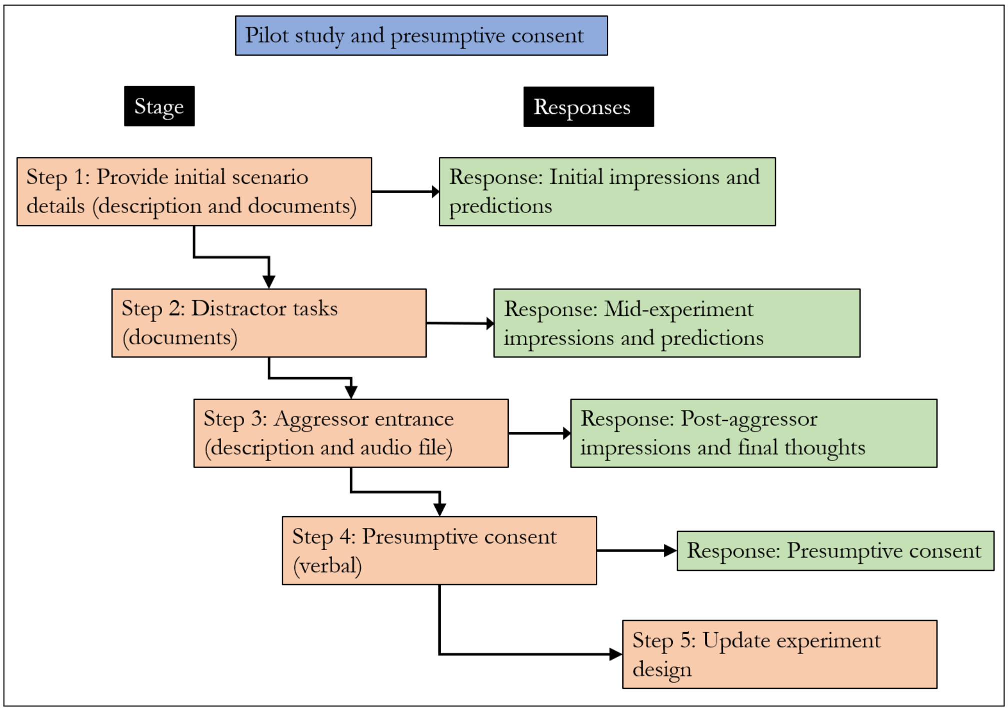


Figure 4.2.: Pilot study flowchart

By leading the participants through the entire experimental design, this approach improved the experimental design by understanding the potential interpretation of the various design aspects, as well as predict likely behavioural responses and therefore evaluate different experimental elements. The design of the experiment was updated after each interview, accommodating any unanimous comments from the interviewees.

Transcripts of each session are provided in Appendix B, and a thematic analysis is provided below in Table 4.4.

Int	Experiment composition	Interpretations and comments	Conclusions	Presumptive consent

Int	Experiment composition	Interpretations and comments	Conclusions	Presumptive consent
1	Participant-aggressor Incremental earnings A 'finish/safe' area is marked out	Shouldn't have the aggressor originally be a participant: too much confusion arises. Didn't understand why someone is being pulled out of the experiment. Should trade-off how much you tell people with how scary it is. Fear caused by aggressor. Proposed arguing against aggressive person. Wondering about whether someone is cheating.	Experiment is at the limit of the level of 'stress' caused.	4/4

Int	Experiment composition	Interpretations and comments	Conclusions	Presumptive consent
2	Participant-aggressor Incremental earnings A 'finish/safe' area is marked out	The experiment could increase the level of stress significantly without breaching ethical guidelines. Suggested that the aggressor could to bring in a knife or a pistol. Could increase the level of stress by having the actor could portray a far scarier persona. Ensure the actor is explicitly identified as having the ability to cause participants to lose financial reward.	Hostile aggressor should not be a participant, but rather should a previously unseen, identifiable member of the experimental team.	3/3

Int	Experiment composition	Interpretations and comments	Conclusions	Presumptive consent
3	Unseen-aggressor Incremental earnings A 'finish/safe' area is marked out	Participants would ignore aggressor in favour of continuing to earn financial incentive through task. Need to ensure that the 'threat' is identified and valid. Could increase level of stress, to ensure ecological validity.	Incremental earnings paradigm should not be used. Hostile aggressor should be should a previously unseen, identifiable member of the experimental team. The 'finish' area should be removed.	3/3

Table 4.4.: Thematic analysis of pilot study

4.4. Physical experiment design

As a result of the pilot study and the previous design process, the final PR experiment design used the components described below.

4.4.1. Experimental area design

The experimental area is shown below in Figure 4.3 and Figure 4.4, and was kept constant throughout all experiments. The individual elements of the experimental area are described in Table 4.5.

The experiment was held in a sports hall, approximately 18m wide and 30m long. There were several key elements to the experimental area design, including:

Item	Description
Positioning anchors	There were five positioning anchors used in the UWB positioning system. These were placed at the corners of the experimental area, and above the central column. These locations were chosen to maximise the accuracy of the positioning system.
Separating wall	The separating wall was introduced to the area to define the initial experimental area, as well as to provide a physical boundary on which the distractor task could be placed. The wall consisted of blue table-tennis tables.
Central column	The central column was introduced for increased accuracy of the UWB positioning system, as well as to provide a physical wall for the distractor task.
Entrance/exit	The entrance/exit was propped open permanently during the experiment. Storage area This area was used as storage for the personal effects of the participants.
Distractor tasks on the walls and central column.	The distractor tasks were a mixture of physical and mental tasks, provided as a method of inducting the participants into the experiment.
Actor-participant position	The actor-participant was in the same position in every experiment when the actor-aggressor entered the environment.

Table 4.5.: PR environment details

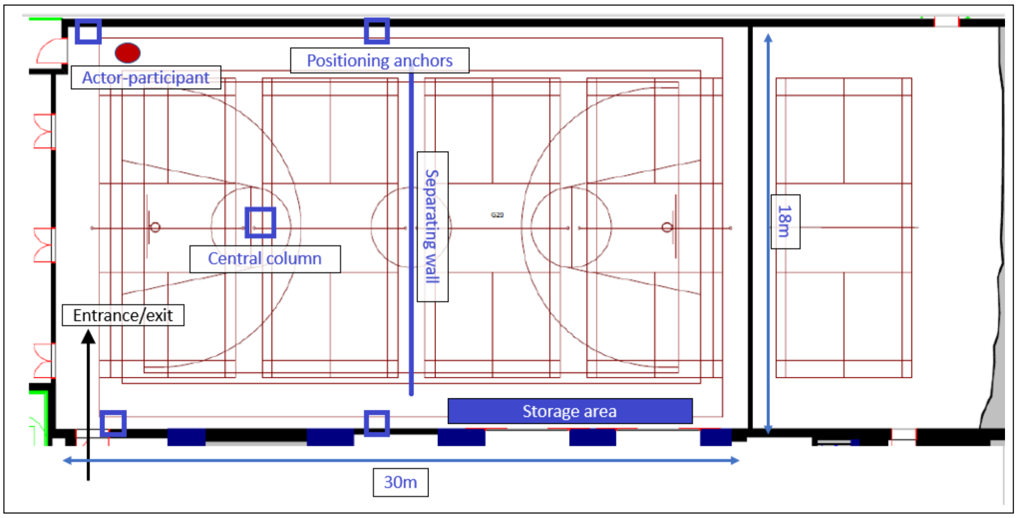


Figure 4.3.: Detailed view of the PR environment

Figure 4.4.: Photograph of PR environment during an experiment

4.4.2. Hostile aggressor design

As mentioned above, the design of the hostile aggressor is integral to the validity of the experiment. It was determined that the aggressor should be:

- Previously unseen, and explicitly identifiable as a member of the experimental team

- Initially shouting, and moving at a slow speed (up to a jog)
- Instructed not to approach participants who are visibly struggling with the experiment

In this experiment, the actor-aggressor would shout loudly upon entering the experimental area, approach the actor-participant first, and move at a jog. A full description of the actions of the hostile aggressor is available below, and a timeline of the final PR experimental design is provided in Figure 4.5, and of the final VR experiment in Figure 4.10.

4.4.3. Financial stressor design

The financial stressor design was binary incentive, using the following two principles:

1) A high financial payment upon successful completion of the experiment.
2) A low financial incentive for participants who fail to complete the experiment.

The amount each participant would be paid was primarily determined by the sample size and the total funding for the experiment. To determine the required sample size, a power analysis was performed on G*Power 3.1.9.7. The initial analysis (see Table 4.6 and Table 4.7) was performed to analyse how many participants were required for the four groups (Financial Control, Aggressor Control, Experiment: Prompt and Experiment: No prompt).

After having established that the total sample size needed to be 112 or higher, it was decided to recruit up to 135 participants, with a maximum payment of £40 per participant.

With the exception of the financial control experiment, all participants were informed of the risk to their payment. Specifically, they were told that they would be paid £40 upon successful completion of the experiment, but that if they failed the experiment, they would only be paid £5. The £5 minimum payment was provided as an incentive to ensure that participants would sign up to take part in the experiment. However, all participants were in fact paid the full £40 at the end of each session, regardless of what occurred during their experiments.

Input parameter	Value
Statistical tests	ANCOVA: Fixed effects, main effects and interactions
Type of analysis	Compute sample size, given alpha, power and effect size
Effect size	0.4
Alpha	0.05
Power	0.95
Numerator df	3
Number of groups	4
Number of covariates	2

Table 4.6.: Initial power analysis inputs

Output parameter	Value
Denominator df	103
Total sample size	112
Actual power	0.9511

Table 4.7.: Initial power analysis outputs

4.4.4. Stressor comparison

A comparison between the proposed hostile aggressor-financial stressor with a bladed attack is detailed below in Table 4.8. It can be seen that there is a large overlap between the two, with the caveat that the participant interpretation of the intensity of the financial reward is a limiting factor. This can be improved by increasing the financial reward offered, but is in turn limited by the funding available for the experiment.

4.4.5. PR experimental procedure

The PR experiments were performed over 17-21st December 2018. Two separate experiments were performed on each of the five experiment days. The experiments took approximately 90 minutes each, including the set-up and break-down of the environment. The experimental team consisted of:

- Primary investigator
- Three support team members
- Two professional actors

Characteristic	Bladed attack	Proposed stressor
High intensity	Yes	Yes - dependent on interpretation of the level of financial reward.
Unannounced	Yes	Yes
Binary	Yes	Yes
Tactile	Yes	Yes
Mobile	Yes	Yes
Individual stressing agent	Yes	Yes
Group targets	Yes	Yes
Non-selective	Yes	Yes
Non-competitive	Yes	Yes
Time scale	Short	Short
Pain stimulus	Yes	No

Table 4.8.: Comparison between stressor characteristics

Recruitment process

Recruitment focused on students and staff members at Imperial College London. Recruitment was performed by email and by electronic advertisements around the South Kensington campus. This restriction on participant type was primarily a result of the ethical and insurance aspects of the experiment, and represents an experimental limitation in terms of the participant demographic. This limitation was mitigated as far as possible by obtaining all relevant demographic data prior to the experiment. Further to this, several inclusion and exclusion criteria were required as an unavoidable ethical concern. The following restrictions were provided for applicants:

Inclusion criteria

- Fit and physically healthy (e.g. able to jog 100m without stopping)
- Able to abstain from caffeine for 24 hours.
- Non-smoker

Exclusion criteria

- Any stress related illnesses (e.g. PTSD, high blood pressure)

Input parameter	Value
Test family	t-tests
Statistical tests	Means: difference between two independent means (two groups)
Type of analysis	Compromise: compute implied power and alpha, given beta/alpha, sample sizes and effect size.
Tails	One
Effect size	0.8
Beta/Alpha	2
Group 1 size	17
Group 2 size	50
Number of covariates	2

Table 4.9.: PR power analysis inputs

Output parameter	Value
Df	63
Alpha	0.055
Power	0.890

Table 4.10.: PR power analysis outputs

- Any coronary disease
- Depression or anxiety related illnesses

A power analysis was performed to establish the required numerical split between control and experimental group splits. As this PhD established that there is insufficient data surrounding movement behaviour within hostile emergencies (see Chapter 2), it was important to prioritise the number of participants within the experimental groups. This analysis was performed in G*Power 3.1.9.7, with the inputs shown in Table 4.9, and outputs shown in Table 4.10.

These results directed the following predicted experimental group splits (Table 4.11).

The recruited participants were initially allocated using randomised stratified sampling, although the participants had the option to request changes to different experiment days and times, if they had scheduling clashes. Overall, 135 participants were recruited and allocated according to the above scheme.

Group	Number of participants
Financial control	17
Aggressor control	18
Experiment prompt	50
Experiment no prompt	50

Table 4.11.: PR experimental group splits

However, only 80 (50M, 30F) participants attended the PR experiments, representing approximately 41% attrition between recruitment and experiment.

Experimental branches

The ten separate experiments were split into the base experiment (with two branches) and two different control experiments. These are detailed below in Table 4.12.

The next section will describe the timeline and protocol of both prompt and no-prompt branches of the base experiment. For a further explanation of the control experiments, see below.

Base experiment: initial stage

After adapting the experimental design using the insights provided by the pilot study, the PR experiment was designed according to the timeline below (Figure 4.5). This timeline includes the two separate branches (prompt and no-prompt), indicated by the two separate boxes below the time axis.

Participants were brought into the experimental hall and asked to complete the initial paperwork and baseline saliva sample. The paperwork included a participant information sheet, a consent form, and a STAI-T questionnaire. They were asked to put on the UWB tags, and the selected participants were asked to put on the ANS equipment. After this, they were informed by the primary investigator:

> *"We'll be looking at how people move and perform while they do cognitive and physical tasks. These tasks will become more and more complex and stressful, so we will also measure physiological signs and check saliva samples. You will be paid up to £40 if you successfully complete this task, which we expect to take around 10 minutes. Please keep in mind that you fail the task or are removed*

Experiment	Description	Number of experiments	Number of participants
Financial control	Participants have no financial stressor, as they are told they are guaranteed their financial incentives prior to the experiment, and the aggressor is directed not to be hostile. Therefore, the hostile aggressor does not represent a potential loss.	1	8
Aggressor control	Participants are never exposed to the hostile aggressor, and consequently are not exposed to the financial stressor.	1	11
Experiment prompt	Participants are exposed to both financial stressor and hostile aggressor. The actor-participant stands still when approached.	4	31
Experiment no prompt	Participants are exposed to both financial stressor and hostile aggressor. The actor-participant moves away when approached.	4	30

Table 4.12.: Breakdown of PR experimental branches

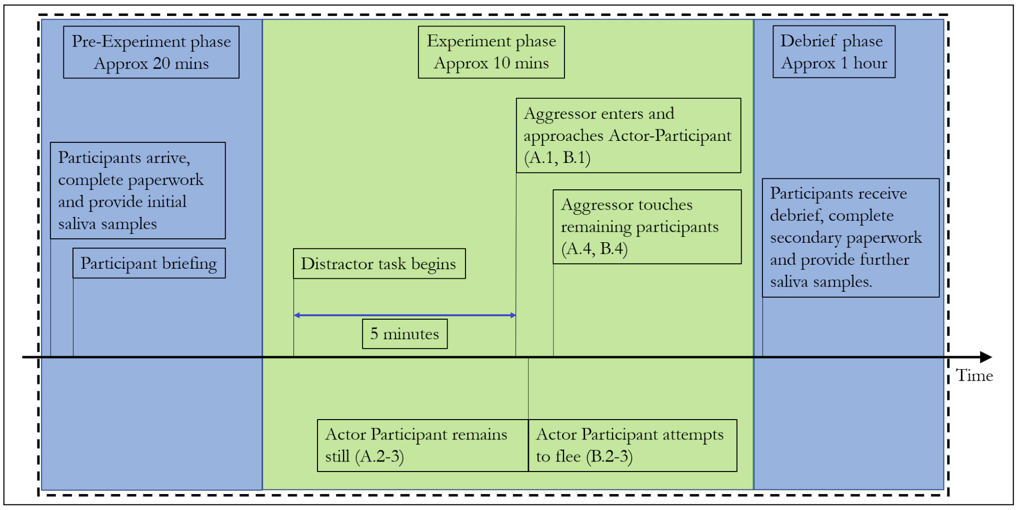


Figure 4.5.: Timeline of PR experiment branch protocols

> *for any reason you will only be paid £5. You must follow any instructions that you receive from myself or anyone wearing an 'instructor' t-shirt. You can go anywhere in the building, or indeed outside the building, but to start with there are some tasks on the walls. Please keep in mind that these have absolutely no bearing on your performance in the experiment, and therefore on whether or not you will be paid the full amount."*

Distractor tasks

During the PR experiments, the participants were asked to complete distractor tasks as a method of getting them to engage with the experiment, as well as a method of disguising the true aims of the experiment. These distractor tasks were comprised of cognitive and physical tasks, designed to be distracting but not mentally taxing. Example tasks are provided in Appendix C. This phase lasted a fixed duration of 5 minutes, after which the experimental intervention occurred.

Base experiment: intervention

As part of the secondary aims of this experiment, the procedure included an intervention to check for social responses, leading to two separate experimental branches. These branches are described below, with reference to the graphical descriptions shown in Figure 4.6.

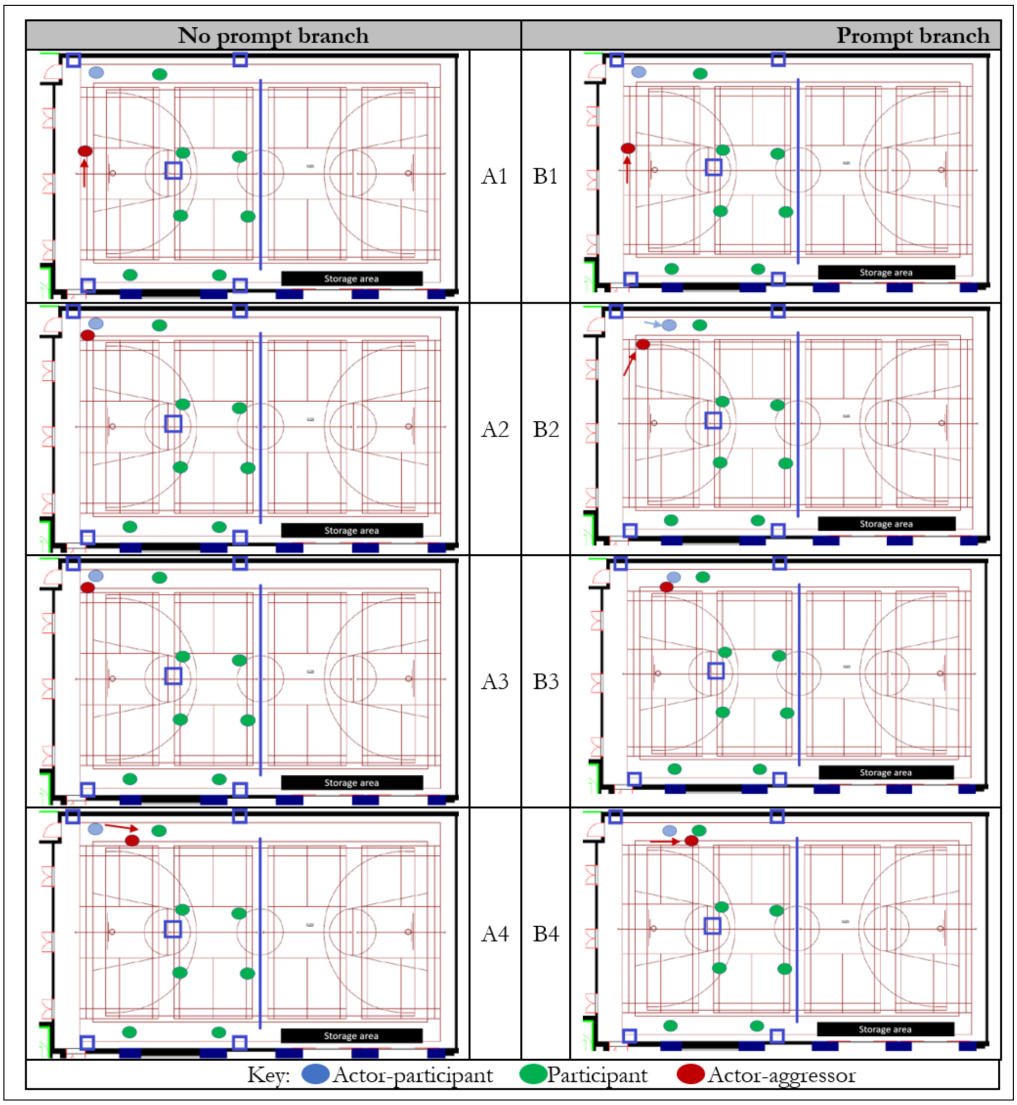


Figure 4.6.: Experiment branch protocols

In both of these branches, the actor-aggressor always enters at the same time, and targets the actor-participant first, who is always standing in the same position (the actor-participant position, see Figure 4.3).

Branch 1: (NO-PROMPT) In the first branch (four experiments total, coded as “NO-PROMPT”), the actor-aggressor walks into the room (A.1), shouts at, approaches and touches the actor-participant on the shoulder (A.2). After touching the actor-participant, the actor-aggressor loudly shouts (A.3):

"HEY, YOU, FUCKER! STOP! YOU'VE BEEN REMOVED FROM THE EXPERIMENT. YOU HAVE FAILED THE TASK AND YOU LOSE ALL YOUR MONEY. SIT DOWN AND DON'T MOVE."

The actor-participant was directed to act as if confused and irritated, before sitting down. The actor-aggressor then announces loudly to the rest of the group:

"ALRIGHT EVERYBODY! I'M GOING TO TRY TO TOUCH YOU NOW. IF I SUCCEED, YOU WILL ALSO FAIL THE TASK AND LOSE ALL YOUR MONEY."

At this point, the actor-aggressor approaches the remaining participants, selecting the next target by locating the nearest participant, and touching each of them in turn (A.4), moving no faster than a quick jog.

Branch 2 (PROMPT): In the remaining four experiments (coded as "PROMPT"), after 5 minutes of initial distractor task performance, the actor-aggressor walks into the room, points at the actor-participant and shouts (B.1):

"HEY YOU FUCKER, AND EVERYONE ELSE! I'M GOING TO TRY TO TOUCH YOU NOW. IF I SUCCEED, YOU WILL ALSO FAIL THE TASK AND LOSE ALL YOUR MONEY."

At this point the actor-aggressor then approaches the actor-participant and attempts to touch him. The actor-participant was directed to attempt to evade the aggressor by running a few steps away (B.2). After the actor-aggressor touches the actor-participant (B.3), he goes on to approach the other participants in an identical fashion to the no-prompt branch (B.4).

These experimental branches formed the basis for the logistic model that will be reported in the results section. This description is shown graphically in Figure 4.6, and as the timeline above in Figure 4.5.

If at any point a participant leaves the initial experimental area (either through the entrance/exit, or by going behind the separating wall), the actor-aggressor will stop pursuing them and move to another participant. If a participant leaves the environment through the entrance/exit, they will be taken aside by a member of the experimental team and informed that they have completed the experiment, that they will be paid the full amount, and that they should remain where they are until the experiment is completed.

Base experiment: final stage

After all participants have either been touched by the actor-aggressor, or have left the experimental area, the experiment is called to a halt by the primary investigator. At this point the primary investigator says:

> *"Ok, thank you everyone, the experiment is over. Can everyone move over to the side please?*
>
> *"I would just like to start by saying that there will be no financial penalties on anyone who was touched during this experiment. I will now take you all through a debrief, after which we will take the final saliva samples.*
>
> *"First of all, our experiment was looking at movement behaviour, but during stressful scenarios. Specifically, we are looking at how people move when confronted with a bladed attacker – here, our hostile member represented a person with a knife. The initial tasks were there as a distractor, and we introduced financial incentives to make you think that you had something to lose if you were taken out of the experiment, and then an actor was used to imitate an aggressor. However, all of you will be paid the full amount, as there will be no penalties taken.*
>
> *"The saliva samples are to test for cortisol levels, which a strongly correlated with a stress response. We measured your position throughout this experiment with the tags in your helmets. These will provide us with quantitative measurements of how people move when they are performing tasks.*
>
> *"We would ask that you read and complete the following documents: the post-test information sheet, a reconsenting form, STAI-S, PANAS, and group identity questionnaire. These will let us know what you were thinking and feeling during the experiment. We are going to hand out some contact details for you if you need to contact any health services as a result of this experiment. If you experience any lasting effects, please either get in contact with your GP, any service on this sheet, or email myself and I will get you in contact with the appropriate services."*

All participants were also asked to produce saliva samples at 20, 40 and 60 minutes after the initial entrance of the actor-aggressor, and finally were provided with the payment for the experiment.

4.4.6. Control experiments

As was detailed earlier in Table 4.12, two separate control experiments were developed, to aid in understanding the impact of the various experimental interventions. For the purposes of the analysis, both of these experiments were coded as "NON-STRESS".

Control experiment: Aggressor control

In this instance, the participants were presented with the same initial stage instructions as the main experimental group, but the hostile aggressor was not introduced at any point. The experiment was called to an end after ten minutes, and the debriefing was performed.

Control experiment: Financial control

In this scenario, the participants were informed that they had a guaranteed £40, removing any financial stressor. The participants were told during the briefing that there would be some tasks to perform, but they were guaranteed the full £40 payment.

During the experiment, the aggressor entered the room and explained that he was going to try to touch them, and if they were touched, they should remain still (as if in a game of 'tag'). No actor-participant was used in this experiment.

4.4.7. PR experiment ethics

It took approximately nine months to design this experiment and obtain permission from the Imperial College Research Ethics Committee (ICREC). The certificate of permission (ICREC reference: 18IC4637) was granted on 6th August 2018 and a copy is provided in Appendix C.

4.5. Virtual experiment design

Between 14th January - 24th February 2020 this PhD performed 55 separate VR experiments, with the participants recruited from Imperial College London student and staff populations. Each experiment lasted approximately 90 minutes, including set-up and break-down of the equipment, and required the presence of a single investigator to perform.

4.5.1. VR equipment

This experiment was performed using an HTC Vive Pro Eye, in a room that measured approximately 5m x 5m. The HTC Vive Pro Eye is an outside-in VR device, using a minimum of two base stations. The VR equipment was supported by an Alienware Aurora R8 desktop computer, with an Intel Core i9 9900K CPU, and a Nvidia GTX 2080Ti GPU. The HTC equipment is shown below in Figure 4.7.

Figure 4.7.: VR equipment

4.5.2. Environment

The environment was built in Unity version 2019.1.10f1. The environment was designed to mimic that of the sports hall used in the PR experiment. To this end, floorplans from the sports hall were imported into Unity and the central column and dividing wall were both placed in an identical position as they had been during the PR experiment. The design of the environment is shown in Figure 4.8.

4.5.3. Locomotion technique

The locomotion technique used in this experiment was based on arm movement, provided by the open-source Virtual Reality Tool Kit (VRTK). In order to move, the participants held down the trigger button on the controller, and move their arms as though they were walking. The speed of movement in

Figure 4.8.: VR environment

the environment was dependent on the speed of the arm movement, and the direction of movement was defined by the average direction of the controllers. This locomotion technique allowed the participants to move at variable speeds and in any direction, while being decoupled from their gaze direction. This allowed the participants to move through the environment by swinging their arms, in a similar manner to walking and running.

Further movement could be achieved by the participant physically walking around the environment. The investigator was present at all times to ensure that the participant was not at risk of harm by walking into walls or tripping over.

4.5.4. Non-player characters

The non-player character (NPC) avatars were produced using the open-source online tool Blender, with animations such as walking, running, strafing, and waving produced from the similarly online open-source tool Mixamo. A further open-source toolkit, the Unity HeadLookController, was used to provide gazing animations for the avatars.

Four separate NPC avatars were used in this experiment (examples are shown below in Figure 4.9), comprised of two female avatars and two male avatars. The NPC participants used one male and one female avatar, the aggressor was a separate male avatar and the instructor was a separate female avatar.

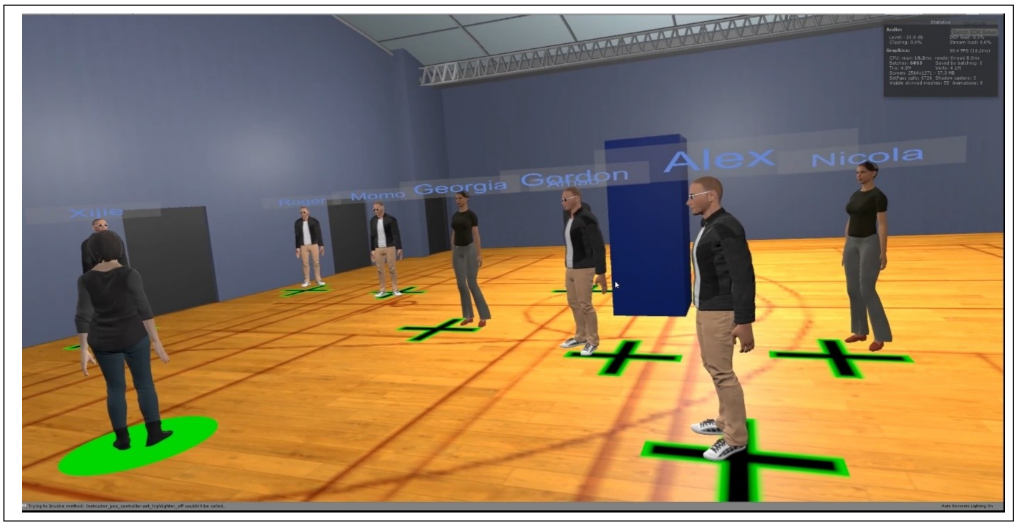


Figure 4.9.: VR NPC avatars

4.5.5. Group responses

The VR experiment was designed to mimic the PR experiment as far as practicable. However, an obvious difference between the two experimental paradigms was that the participants were not taking part as a group, but rather as individuals. However, as the PR experiment had indicated that individual behavioural responses are strongly affected by social influences (see Section 8.2 for further details), it was considered necessary to convince the participants that the other NPCs within the experiment were controlled by other 'live' participants. Therefore, over the experimental process, several elements were developed to convince the participant of this. The participants were all:

1) Informed that the other avatars in the environment were controlled by other participants

2) Informed that they would be meeting the other participants after the experiment

3) Asked to type their names into the computer, after being informed that it would be displayed above their avatar. They were told that this was so

that they could identify the other avatars, and that they would in turn be identifiable

4) Asked to record an audio introduction, detailing their name what they were studying and which room they were in During the experiment introduction section:

5) Invited to play their own introduction audio file. The NPC audio files were all unique and each experiment provided the same avatar introductions.

6) Invited to 'wave' to the other NPCs. At this point the NPC avatars were animated as waving to each other several times (ranging from one to three waves, separated by a fixed duration ranging from 0.2-1.0 seconds). As the NPC waves were predetermined, each experiment was identical in this respect.

The level of participant belief in the humanity of the observed avatars was measured after each experiment, after they had completed the final pieces of data collection. Further discussion and analysis of this belief level is provided in Section 5.1.

4.5.6. VR experimental procedure

Recruitment

The VR experiment recruitment process was identical to that of the PR experiment detailed above, with a total of 55 experiments performed (37 male participants, 18 female). As the VR experiment was designed to be as similar to the PR experiment as possible, the financial incentive was also determined to be £40. Prior to the experiment, each participant completed an identical questionnaire to the previous experiment, including demographics and personality measures. The analysis of this questionnaire is detailed in Section 5.4.

Experimental branches

As with the PR experiment, there were two VR experimental branches, aimed at understanding the potential social influence on movement decisions. The two branches, coded as "VR-MOVEMENT", and "VR-NO-MOVEMENT", showed differing levels of NPC movement after the aggressor had entered. This

Branch	Details	Number of participants
Movement	Both the hostile aggressor and the NPC avatars seen by the participant follow routes defined by the data taken from a PR experiment. At the end of the predetermined route, the aggressor targets the participant, if they are still in the experiment.	27
No movement	The avatars seen by the participant do not move in response to the aggressor. The hostile aggressor approaches avatars (including the participant) based on proximity.	28

Table 4.13.: VR experiment branches

is detailed below in Table 4.13. The participants were split equally between two experimental branches, with the branch allocated to each participant chosen randomly at the beginning of each experiment.

Phases

The VR experiment was performed with one participant at a time, with each individual experiment taking approximately 90 minutes. The experiments were split into three phases, shown in Figure 4.10.

Pre-experiment phase

During the pre-experiment phase, the participants were provided with an information sheet and taken through a consenting process, and were asked for initial self-assessed survey responses. They then provided baseline saliva samples, put on the e4 Empatica watch (and for several participants, the Equivital EQ02 chest monitor), and were introduced to the VR system. At this point the participants were asked to record the introduction that would be played during the experiment. This introduction lasted for seven seconds and could be re-recorded if necessary. Experiment phase

At this point the participants placed the HMD on their heads, and they

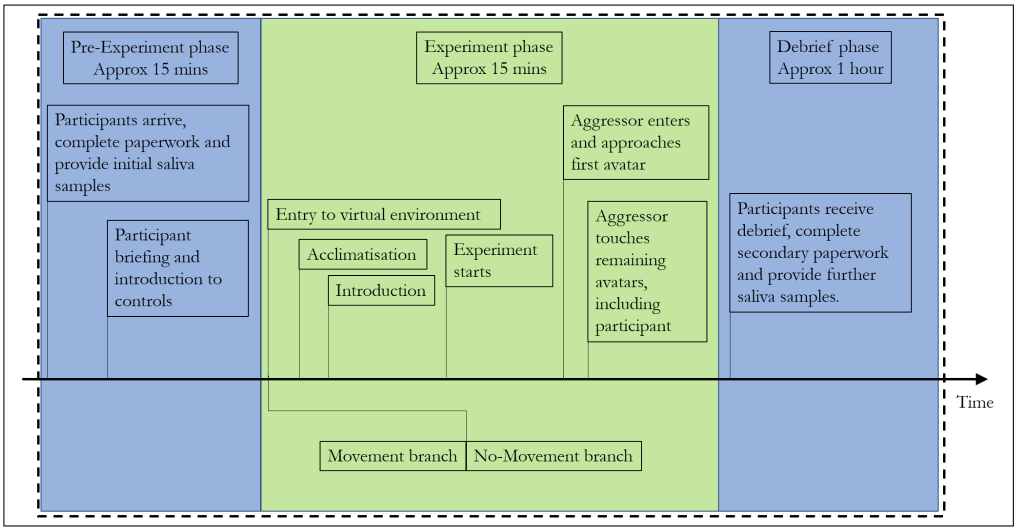


Figure 4.10.: VR experiment timeline

were placed into the experimental area. Simultaneously, the allocation procedure was performed, with a random number generator in the VR environment determining the branch in which to place each participant.

The participants were asked to perform some acclimatisation exercises in an introduction phase, followed by the actual experiment (see below for more details). Debrief phase

During the debrief phase, the participants were asked to fill in similar questionnaires to the ones provided to participants in the PR experiments. They were also provided with a further, post-experiment information sheet, detailing the true aims of the experiment and a re-consenting form. Finally, during this period they also provided saliva samples at 20, 40 and 60 minutes after the aggressor's entrance into the experiment.

Experiment phase

Storyboard

The experimental storyboard is detailed below in Figure 4.11.

Acclimatisation

As discussed earlier in Section 3.4, VR experiments typically require an acclimatisation period, where the participant is able to:

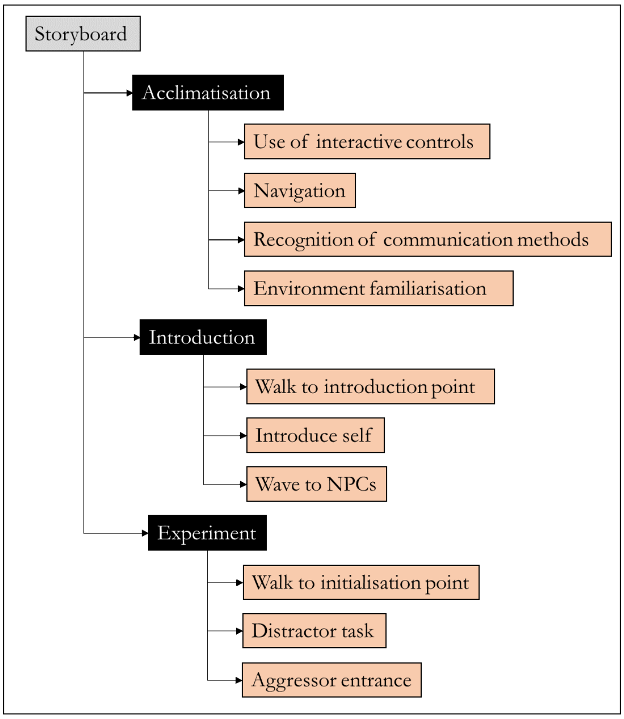


Figure 4.11.: VR experiment storyboard

- Understand typical prompt methods
- Familiarise with the whole environment
- Familiarise with navigation techniques
- Familiarise use of interactive controls.

To familiarise the participants with the controls, they were placed into the environment and presented with a prompt in their field of view. This prompt informed them of the navigation technique and asked them to explore the environment to find red crosses on the floor.

After the participants reached the first red cross (see Figure 4.12), the prompt is removed and replaced with an interactive onscreen button which they were asked to press with a pointer activated by touching the touchpad on their controller. A further prompt was then presented, asking the participant to press the 'grip' button on their controllers. Finally, the participant is asked to find another red cross on the floor in order to begin the experiment.

<u>Introduction</u>

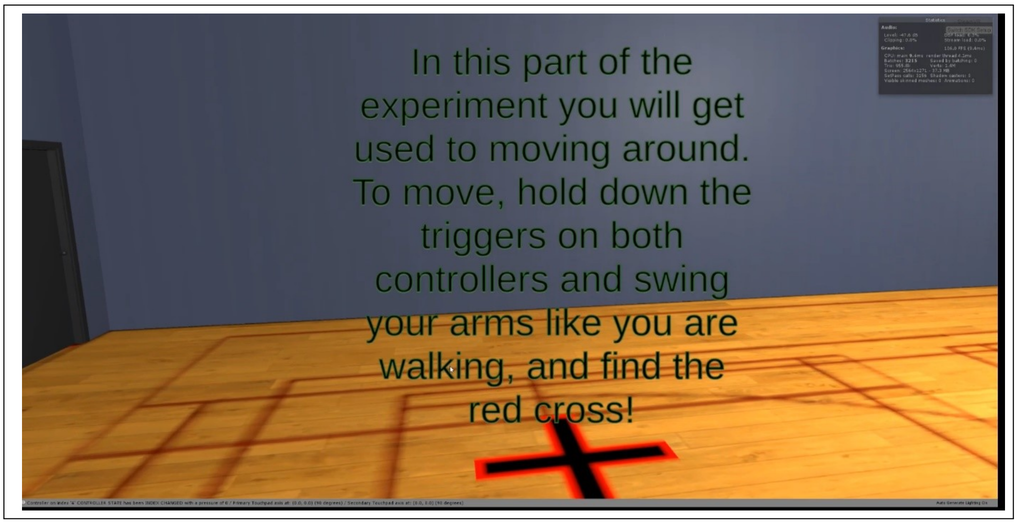


Figure 4.12.: VR initial instructions and red cross

When the participant has completed the acclimatisation section and moved to the introduction point, the introduction phase of the experiment starts. At this point the NPC 'participants' enter the virtual environment, followed by an investigator. The investigator moves to the centre of the environment, and is identified with a green ring around their feet (see Figure 4.13), indicating that they are talking. The investigator then says the following:

> *"Hi everyone, can I have your attention please?*
>
> *"Thanks for coming to be part of this experiment, we're going to start by getting everyone introduced, and there will be further instructions later on. Could everyone please make their way to a red cross."*

At this point the NPCs are automatically routed to pre-defined red crosses that appear on the floor. Once the participant has also moved to a red cross, the investigator then says:

> *"Thanks everyone. We'll start over here. Please could you introduce yourself to the group, and then we'll go around the room."*

At this point the NPCs and participant are identified one by one as a green highlighted circle appears at their feet and are asked to introduce themselves by playing pre-recorded audio files. They introduce themselves in order of

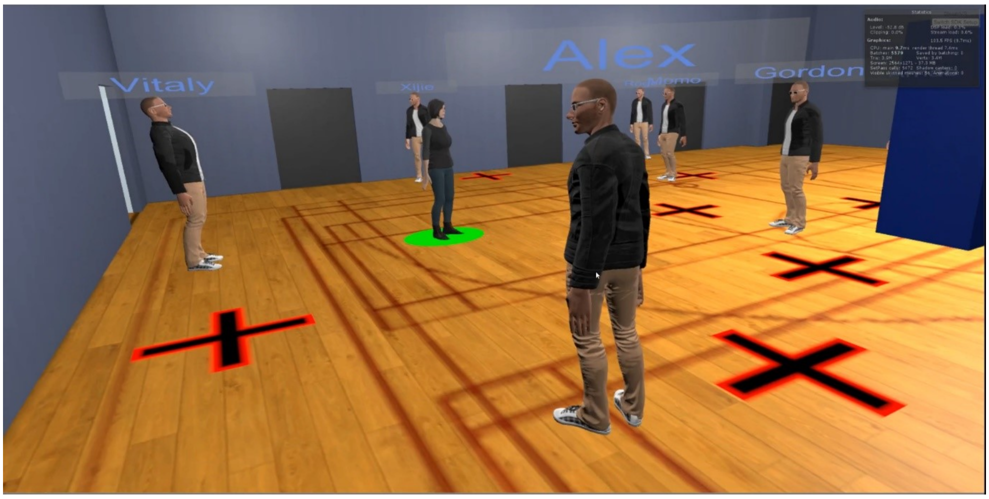


Figure 4.13.: VR instructor entering the environment

distance from the investigator, with the nearest avatar beginning. The NPCs introduce themselves after a fixed delay ranging from 1-3 seconds. Once the participant is identified, a button appears on the participant's screen, asking them to click a button to introduce themselves. Once the participant and all NPCs have introduced themselves, the investigator says:

> *"Thanks for that everyone. You are all here today to go through this experiment looking at how people move in virtual reality, and also specifically looking at how groups interact together in relation to cognitive tasks. Please keep in mind that you'll all meet in person for a debrief later. As you're moving around this arena, feel free to go wherever you want. As part of this experiment you can wave at each other. You just wave your controller, and click the grip button at the side, and the other participants will see your avatar waving."*

At this point the participant is presented with a prompt on their screen, asking them to click the grip button on their controllers, and wave their hands (Figure 4.14). At this point they see the NPC avatars wave at each other. After the participant presses the grip button, the investigator says:

> *"Excellent – now we're ready to start the experiment. The aim is to successfully complete the experiment and further instructions*

Figure 4.14.: Waving in VR experiment

> *will be given to you at specific points. You'll all either be given instructions up on your screen, or told what to do by an instructor. You'll be paid up to £40 if you complete this experiment, which we expect to take around 10 minutes. Please keep in mind that if you fail the task, or are removed for any reason, you will be only paid £5. Now, let's start the experiment."*

After the investigator finishes giving out these instructions, the participant is presented with a prompt on their screen, asking them to find a red cross on the floor in order to start the experiment.

<u>Experiment</u>

Once the participant has found the red cross, they are given the distractor task of waving at the nearest avatar. After a fixed-time delay of two seconds an aggressor avatar enters (Figure 4.15) and loudly explains his role, saying:

> *"HEY EVERONE. THAT'S RIGHT, FUCKING LISTEN UP. IF I TOUCH YOU, YOU WILL LOSE THE EXPERIMENT, AND LEAVE WITH NO MONEY. YOU UNDERSTAND? YOU CAME HERE WITH £40, AND IF I TOUCH YOU, YOU WILL LEAVE WITH £5.[Pointing at avatar in starting position*] RIGHT YOU, I'M GOING TO TOUCH YOU"*

* This position is identical to that of the actor-participant in the PR experiment.

Figure 4.15.: Aggressor entering the VR environment

At this point, the aggressor then pursues all NPCs and the participant, selecting the next target by proximity. If the aggressor moves into close contact (<1m) of the NPCs, they are prevented from moving further, placed into a crouch position, and highlighted in red (Figure 4.16). If the aggressor moves into close contact of the participant, then the participant is removed from the environment and presented with a prompt to take off the headset.

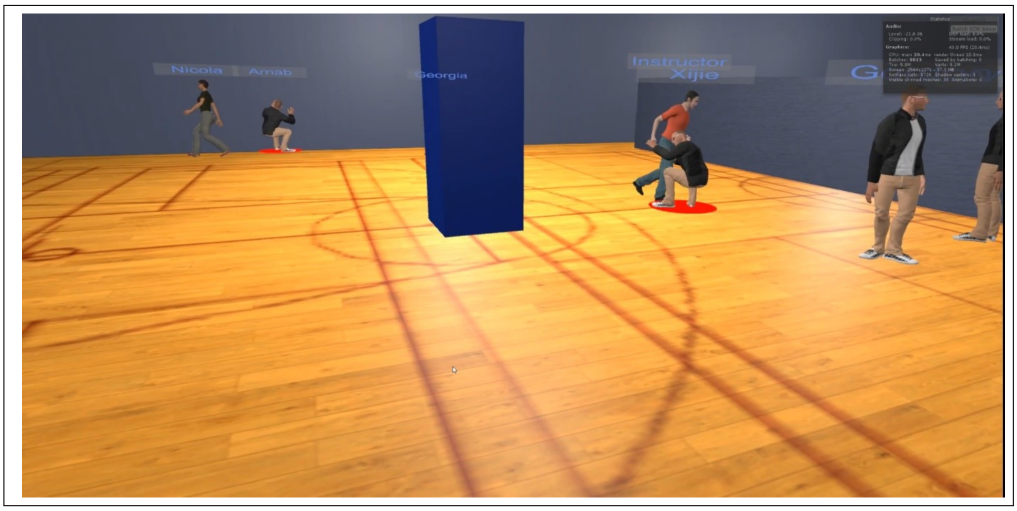


Figure 4.16.: Avatar removed from VR experiment

The experiment was limited to two minutes, at which point it would end and the debrief process would take place.

4.5.7. VR Ethics

It took approximately six months to design this experiment and obtain permission from the Imperial College Research Ethics Committee (ICREC). This experiment was considered lower risk than the PR experiment, owing to the lack of movement and the virtual nature of the experiment. The certificate of permission (reference: 19IC5216) was granted on 22nd September 2019 and is provided in Appendix C.

4.6. Differences between experimental paradigms

While significant efforts were made to ensure the PR and VR experimental paradigms were as similar as possible, it was inevitable that there would be some differences. The major differences are detailed below.

Real neighbour participants

In the PR experiment each individual session consisted of several participants who underwent the experiment at the same time. However, given time, logistical and equipment constraints, this was not possible within the virtual experiment and therefore the VR experiment participants were not taking part in an experiment with other real participants, but rather with automated versions of participants.

Although measures were taken to convince the participants that the avatars they observed were real (see above for details), it remains a significant difference between the experimental paradigms. As a result of this difference, the interactions between participants and observed avatars may be distinctly different from those from an experiment in which the avatars had been controlled by real participants. The post-experiment surveys for the virtual paradigm asked participants about their belief in the controlled nature of the avatars, with the majority of participants stating that they had been convinced that the avatars were real participants. This level of belief is investigated fully in Section 5.1.

Future experiments should perform experiments with groups of participants taking part simultaneously within shared environments. This will be a significant improvement to the experimental design, as it reduces experimental deception, and increases the parallels with the PR experimental paradigms.

Possible activity sets

Owing to the automated nature of the environment, there were limitations on the available participant actions. For example, during the stressful period of the experiment, there were several activities available to the PR participants that were not available to the VR participants, including crypsis ('playing dead') and altruistic behaviour.

There were also limitations on the possible interactions between the participants and the NPCs; these interactions were specifically limited to waving, and one instance of audio communication through a recorded message

These differences in available actions were unavoidable. However, as the VR environment was built after the PR experiment had been completed. This led to the VR environment being tailored to investigating the effects observed in the PR experiment, specifically the impact of avatars moving on the participant's choice to move, and therefore the lack of the above behaviours had a minimal effect on the overall results.

Access to the full building

Access to the surrounding areas of the VR experimental environment were very different to that of the PR experiment. The PR environment was located within a sports hall and had access to the outside world through a series of doors, whereas the VR environment had limits on where the participants could move. In both paradigms the participants were told that they could go anywhere inside or outside the building. This was done to ensure the participants felt able to leave the room if they wanted to. If the participant attempted to leave the room in a VR experiment, their avatar would not be able to move through the door. However, this difference is not considered large, and over the next chapters any behaviour disparities between experimental paradigms will be investigated fully.

Distractor tasks

Finally, the participants were provided different distractor tasks during the experiments. The PR experiment asked the participants to perform cognitive and physical tasks for a fixed duration of five minutes prior to the entrance of the aggressor. The virtual experiment instead asked the participants to perform an introductory session, including waving at other avatars and playing a pre-recorded introduction message. This difference was unavoidable, but was not considered significant, as both task types were simple and low intensity, and participants were aware that these tasks were not part of the experiment.

4.7. Measurements

The measurements obtained during these experiments are detailed below in Table 4.14 and shown graphically in Figure 4.17.

Subset	Method	Approach	External reliability
External	Position Orientation	Separate methods use for PR and VR experiments, obtaining the same data.	Physical position obtained using state-of-the-art UWB system.
Internal subjective	Demographics Personality STAI-S, STAI-T, PANAS Social responses	Both PR and VR experiments obtained these measures for all participants.	Standard demographics obtained. Established surveys were chosen based on a large basis of evidence for validity.
	Recall quality	In-house survey, obtained three months after the experimental date. See below for more details.	Survey developed in conjunction with psychologists at University College London, and improved between distribution for PR and VR participants.

Subset	Method	Approach	External reliability
Internal objective	Autonomic nervous system responses	ANS measures were obtained for two participants per PR experiment, and all virtual experiments.	ANS measures obtained using two separate methods (chest- and wrist-based monitors), using state-of-the-art equipment.
	Endocrine system responses	Endocrine measures were obtained for all participants. These were one baseline sample and three post-experiment samples.	Endocrine measures analysed at Wythenshawe Hospital Biochemistry lab, using gold-standard methods.

Table 4.14.: Summary of measurements obtained

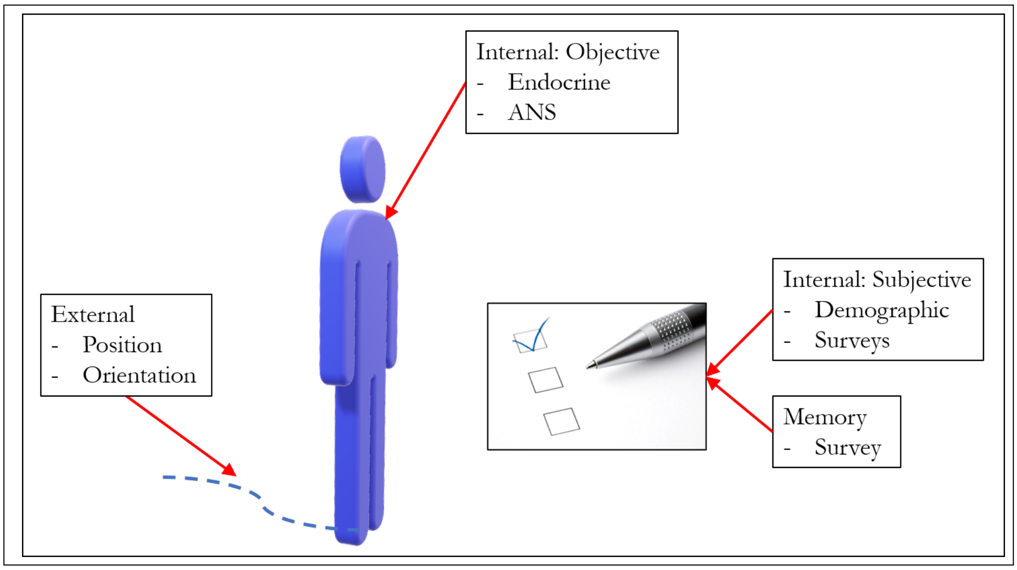


Figure 4.17.: Experimental measurements

4.7.1. Position measurements

PR experiment

For the PR experiments the Pozyx Ultra-wideband (UWB) system was chosen as the primary positional measurement system. Each participant was issued with a helmet and tag to wear (see Subsection 4.7.1), which provided positional measurements with the following specifications:

Figure 4.18.: UWB tags used in PR experiments

- 10cm positional accuracy in three dimensions
- 0.5° rotational accuracy around three axes (pitch, yaw and roll)
- 10Hz time resolution

These tags provided the following measurements for all participants, and for the actor-aggressor and actor-participant:

- Head position coordinates (x, y, z)
- Head orientation as a quaternion (a, bi, cj, dk)

Multiple cameras were also used to record the experiments as a secondary positional measurement system. These were used as a secondary positional tracking system in case the UWB system failed.

VR experiments

Within the VR experiments it is necessary to track the participant, avatars and aggressor. The experimental procedure tracked the following measurements of the participant and all avatars, including the aggressor:

- Head position coordinates (x, y, z),
- Head orientation as a quaternion (a, bi, cj, dk)

These measurements were tracked at 10Hz time resolution

4.7.2. Self-assessed measures

Each participant filled in several different self-assessed survey forms. The timeline for these forms was as follows:

Pre-experiment

Demographics

A full demographic assessment was obtained as part of the recruitment process. This was distributed as a spreadsheet to all participants a week before the experiment, in order to minimise the time taken during the experiment.

Part of the demographic assessment was a Big-5 personality questionnaire, which quantifies personality in five separate factors. An individual's scores are calculated from their responses to fifty questions, where they rank their agreement with a phrase (e.g. 'I am full of ideas') out of 5. The responses to these questions are then combined into five separate scores, one for each personality element.

The personality elements are: Personality A: Agreeableness, Personality C: Conscientiousness, Personality E: Extraversion, Personality N: Neuroticism, Personality O: Openness.

STAI-T

The cognitive state of the participants was assessed prior to the experiment using the short form State-Trait Anxiety Inventory (STAI), which was given out alongside the consent forms, participant information sheets and the initial saliva sample at the beginning of the experiment.

Post experiment

After the experiment several further surveys were distributed, including the state form of the STAI questionnaire, the Positive And Negative Affect Schedule (PANAS) and group identification questionnaires.

STAI-S

This survey was distributed immediately after the experiment, and provides a direct comparison with STAI-T, and therefore is a direct measure of the negative emotions induced by the experiment. This change between pre-experiment and post-experiment measures is defined "Delta-STAI".

PANAS

This survey provides two separate measurements, one for positive emotion and one for negative emotion. The responses to this survey can be used to check the level of emotion induced by the experiment, as well as check correlations between the PANAS and the Delta-STAI.

Social responses: Group Identity and Shared Social Identity

An implication of self-categorisation theory is that in extreme emergencies, the level of group identification and shared-social identification by participants is expected to increase. Following consultation with subject matter experts, measures of group identification (from Leach et al. [228]) and shared social identity (adapted from [229]) were adopted. These questionnaires ask Likert-scale questions, combining the answers into scores for several unique factors. These are:

- Solidarity
- Centrality
- Satisfaction
- Individual self-stereotyping
- In-group homogeneity
- Shared identity

The full questionnaires are provided in Appendix C.

Open questions

Further open questions were asked to gain further insight into the participants' interpretation of the experiment. These were:

- What were you thinking when the aggressor walked in and approached first participant? (e.g. I thought that he was..., My impression was..., I thought that the first participant was...)
- What did you think was the best course of action when the aggressor started pursuing participants? (e.g. I thought that I should..., The sensible thing to do would be to..., What I wanted to do was...)

4.7.3. Physiological measurements

The experiments were designed to incorporate both endocrine and autonomic nervous system (ANS) measurements.

Endocrine system measurements

Each participant provided four separate saliva measurements, using the standard protocol for Sarstedt salivettes. This protocol required participants to only ingest water for an hour before the experiment and avoid caffeine on the day of the experiment. In order to provide a sample, the participant would chew the salivette for approximately one minute, before then storing the sample in a plastic container. Participants provided the samples at the following times:

- Baseline at the beginning of the experiment
- 20, 40, and 60 minutes after the aggressor's entrance in the experiment

These saliva samples were then analysed by Wythenshawe Hospital. This analysis used HP-LC mass spectroscopy to quantify the levels of salivary cortisol following the methodology detailed in [246].

ANS measurement

The ANS measurements were provided by the:

- e4 Empatica watch

which measures BVP, GSR and ST. Using BVP, HR and HRV can be inferred.

- Equivital EQ02 chest-mounted monitor

which measures HR, HRV, GSR, ST and RR.

Logistical and financial limitations meant that only two participants could provide these physiological measures per PR experiment. Further to this, while the e4 Empatica was used for every VR participant, equipment limitations meant the Equivital EQ02 was only used for four VR participants.

4.7.4. Recall quality

Chapter 2 detailed how the theoretical frameworks surrounding human behaviour in emergencies (including terrorist attacks) are based on empirical findings primarily obtained through interviews with individuals who were present during these scenarios. These datasets may be the most common way of increasing the understanding of human behavioural responses to life-threatening situations. However, these datasets are subject to a number of biases including:

- Leading questions
- Social desirability
- Maintenance of a positive image of oneself [247] and one's own group [248]

These biases are often explicitly acknowledged in research (e.g.:[249]), but remain an issue to such an extent that they lead to researchers being unable to confirm the accuracy of the data. Acknowledging the existence of such biases does not reduce the possible benefit of information gathered in these datasets, but does highlight the areas where researchers are uncertain about how the reported information diverges from ground truths. Therefore, it is necessary to quantify these deviations in order to best evaluate the theories of how people behave in emergency situations. Without this quantification, it is not possible to ascertain the disparity between an interviewee's transcript and what actually occurred.

While methodological measures have been developed to mitigate against some of the biases associated with self-reports (e.g. triangulation to ensure independent sources converge with reported facts), biases stemming from the overall reliability of testimonies (i.e. establishing whether there is reasonable correspondence between what happened and what is reported) are more difficult to accommodate.

Current understanding surrounding an individual's recall of threatening events is inconclusive at this stage, and primarily suggests that memory is a constructive process, which can be particularly susceptible to the delivery of new information [250]. During exposure to a stressor, memory of specific and contextual details ([251, 252, 253]) can be impaired, making eyewitness reports probably unreliable, and at least highly questionable ([254, 255]). Review papers are available in [256, 257]. In contrast to this impairment, research has also suggested that attention highlighting that occurs during exposure to a stressor can result in the improvement of specific elements of an individual's memory [258]. This dichotomy of results is unexplained.

Additionally, further memory distortion can occur during time between the stressor and the individual reports as memories undergo natural decay [259] and time generally makes people more likely to encounter external information which can degrade their recall quality [260]. Finally, if the stressor is sufficiently intense, major psychiatric symptoms can occur that may also affect recall of the event ([261, 262]).

Although there is controversy concerning the reliability of testimonies for numerous different scenarios, there is a particular focus on hostile emergencies, such as mass shootings and knife attacks, since the majority of previous data on these incidents has relied on interviews rather than other data sources, as detailed in Chapter 2. There is consequently a significant need for investigation in this area to progress the understanding of human behaviour under these stressful circumstances.

Given this lack of clarity, and the lack of data, it is necessary to gather data to investigate any potential disparity between report and truth. Therefore the final dataset obtained by this PhD was a survey aimed at assessing participant recall of their experiments. This survey methodology was chosen owing to its lower logistical requirements while still providing useful insight into the ability of interviews to investigate emergencies.

This survey was performed to provide further insight into the knowledge base surrounding the impact of stressful events on memory. This could help quantify how much an interviewee's memory might be unreliable based on several domains, and whether this is dependent on the levels of stress experienced. The survey questions were allocated to different domains, allowing the investigator to understand how the recall of an event changes based on these different domains. These domains are described in Section 9.4.

This dataset was obtained by distributing a survey to all participants three months after their respective experiments, asking them to recall their experiment. This was considered as sufficient to act as a representative period for typical interviews of emergency, while still remaining viable within the time constraints of the PhD. Furthermore, this dataset was a cross-sectional investigation aimed at understanding the difference in recall quality between control, PR and VR participant groups at a single moment in time, instead of a longitudinal investigation which would provide insight into how recall quality might change as a function of time.

4.7.5. Example datasets

Examples of the full datasets are provided in Table 4.15, Table 4.16, and Table 4.17.

ID	Timestamp	Position	Orientation	Heart rate	GSR	Skin temp
1	0.0	(1,2,3)	(1,1,1,1)	120	10	20
2	0.0	(2,3,3)	(0.5, 0.5, 1, 1)	80	12	25
Aggressor	0.0	(5,6,3)	(1.2, 4.2, 5, 1)	-	-	-
...	...	...	...	...	...	...
1	0.1	(1.1, 2.1, 3)	(1.2, 1.2, 1.2, 1.2)	120.1	10.1	20.1
2	0.1	(2, 3.2, 3)	(0.5,0.5, 1, 1)	80.1	12.1	25.1
Aggressor	0.1	(5.1, 6.3, 3)	(1.2, 4.2, 5, 1)	-	-	-
...	...	...	...	...	...	...

Table 4.15.: Example PR dataset from a single experiment

Virtual component	ID	Timestamp	Position	Orientation	Heart rate	GSR	Skin temp
Participant	3	0.0	(1,2,3)	(1,1,1,1)	120	10	20
Aggressor	-	0.0	(2,3,3)	(0.5, 0.5, 1, 1)	-	-	-
Avatar 1	-	0.0	(5,6,3)	(1.2, 4.2, 5, 1)	-	-	-
...	...	...	...	...	...	...	...
Participant	3	0.1	(1.1, 2.1, 3)	(1.2, 1.2, 1.2, 1.2)	120.1	10.1	20.1
Aggressor	-	0.1	(2, 3.2, 3)	(0.5,0.5, 1, 1)	-	-	-
Avatar 1	-	0.1	(5.1, 6.3, 3)	(1.2, 4.2, 5, 1)	-	-	-
...	...	...	...	...	...	...	...

Table 4.16.: Example VR dataset from a single experiment

ID	**Paradigm**	**STAI (T,S)**	**PANAS (P,N)**	**Social responses**	**Cortisol (baseline, 20, 40, 60)**	**Survey responses**	**Demographic**
1	PR	(36, 50)	(35,20)	(10,5, 2, 45,6)	(0, 10, 20, 0)	...	...
2	PR	(36, 40)	(25, 20)	(22, 8, 7, 30, 9)	(0, 5, 10, 0)	...	...
3	VR	(33, 38)	(20, 19)	(24, 4, 1, 33, 10)	(0, 6, 11, 2)	...	...
...	...	...	...	...	...	...	...

Table 4.17.: Example combined datasets

4.8. Chapter summary

This chapter has detailed an implementation of the generalised methodology that focuses on measuring behavioural responses to knife-based hostile aggressors. This chapter initially focused on defining the aims of the experiment, before detailing the threat characterisation and proposed proxy stressors. After having established these, this chapter described the pilot study performed in conjunction with PHE, interviewing 10 members of the public over three separate sessions using a narrative interview technique to understand the participant interpretation of specific experimental aspects.

The next sections of this chapter described the proposed implementation in two separate experimental paradigms, PR and VR. These implementations needed to be as similar as possible, and therefore the experimental designs were almost identical. This overlap included the proxy stressor design and the measurements obtained. The major difference between these two paradigms is the presence of social groups, which was mitigated by making a believable set of NPC and participant interactions. Further to this, the movements of the NPCs during the VR experiments were taken from two separate PR experiments, consequently increasing the validity of the experimental procedure.

The final section of this chapter detailed the different proposed measurements obtained in the separate paradigms. The positional measurements were obtained by UWB in the PR experiments, and were recorded directly from the environment in the VR experiments. The psychological measurements were obtained by self-assessed surveys, using an identical method in both paradigms. The physiological used ANS measurement equipment and salivary samples for endocrine measures. The recall quality dataset was obtained by a survey obtained three months after the experiment in both paradigms, while the question sets were updated between the experimental procedures.

This PhD performed both of these experiments. The PR experiments were performed in December 2018 and the VR experiments were performed in January-February 2020. The next chapter will quantitatively describe the datasets obtained by these experiments, including a comparison of the participants. This comparison will perform normality tests on all datasets and quantify any missing datasets. The next chapter will also describe the different analysis techniques that will be used on these datasets in the later chapters.

5. Data Preparation and Analysis

The previous chapter detailed how two separate experimental procedures were designed and performed, and the measurements that were taken during these experiments. This chapter will summarise the data obtained from these experiments, including performing initial assessments on basic data properties such as mean, range, standard deviation (sd) and normality, as well as identifying any missing data. This chapter will then detail the data preparation and the analysis techniques that were carried out.

For the remainder of this thesis the acronym VR will be used to refer to the virtual reality paradigm, and PR will be used to refer to the physical paradigm.

5.1. Data summary

5.1.1. PR experiments

The PR experiments were performed over the period Monday 17th to Friday 21st December 2018, over 10 separate sessions. These sessions were carried out with participants in groups, ranging in size from 5 to 11 participants (mean:8, SD: 1.789). In total 80 participants (26F) took part in these experiments. Furthermore, as the PR experimental paradigm was split into 10 separate group experiments, the demographics of each group are indicated below. The timetable for these experiments is shown below in Table 5.1.

The datasets obtained in the PR experiments are detailed in Table 5.2.

5.1.2. VR experiments

There were 55 separate VR experiments (37M, 18F) performed between January-February 2020. The VR experiments were performed at the following specific times: 09:15, 11:15, 13:15, 15:15, and 17:15. The datasets obtained in the VR experiments are detailed in Table 5.3. The timing difference between the VR

Exp	Day and time	Branch	Number (M/F)	Mean Age (sd, range)
1	Monday, 18:00-19:00	Control-Financial	8 (5,3)	25.7 (5.73, 18.4-34.8)
2	Monday, 19:00-20:00	Control-Aggressor	11 (7,4)	24.6 (2.86, 20.9-30.0)
3	Tuesday, 19:00-20:00	Base – no prompt	5 (3,2)	28.7 (2.42, 25.6-32.0)
4	Tuesday, 20:00-21:00	Base – no prompt	8 (5,3)	25.8 (5.18, 19.3-34.3)
5	Wednesday, 18:00-19:00	Base – no prompt	8 (6,2)	25.1 (4.32, 20.2-32.0)
6	Wednesday, 19:00-20:00	Base – no prompt	10 (5,5)	23.5 (3.04, 19.6-27.8)
7	Thursday, 18:00-19:00	Base – prompt	8 (6,2)	26.5 (3.75, 22.3-31.8)
8	Thursday, 19:00-20:00	Base – prompt	5 (4,1)	24.7 (3.14, 20.0-27.6)
9	Friday, 18:00-19:00	Base – prompt	8 (6,2)	22.3 (2.64, 18.8-27.3)
10	Friday, 19:00-20:00	Base – prompt	9 (7,2)	24.9 (3.34, 21.1-31.5)

Table 5.1.: PR experimental timetable

Data type	Detail	Missing data	Comments
External	UWB	N/A	Issues with near-field effects from anchors. Manual and automated correction performed, before interpolation and smoothing.
Internal subjective	STAI-T	1	Out of a maximum of 80.
	STAI-S	1	Out of a maximum of 80.
	PANAS-P	4	Out of a maximum of 80.
	PANAS-N	4	Out of a maximum of 80.
	Social responses	10	Out of a maximum of 80.
	Demographic	3	Out of a maximum of 80. Obtained one week pre-experiment. Includes Big 5 personality questionnaire.
	Qualitative questions	0	
Internal objective	Cortisol	1	Insufficient saliva in one individual's sample. 4 samples per participant.
	ANS	See below	Limited number of sets.
Recall quality	Survey responses	57	Out of a maximum of 80.

Table 5.2.: PR datasets summary

Data type	Detail	Missing data	Comments
External	VR headset position and orientation	0	
Internal subjective	STAI-T	0	
	STAI-S	0	
	PANAS-P	0	
	PANAS-N	0	
	Social responses	0	
	Demographic	0	Obtained one week pre-experiment. Includes Big 5 personality questionnaire.
	Qualitative questions	0	
Internal objective	Cortisol	0	
	ANS	See below	Limited number of sets.
Recall quality	Survey responses	46	Out of a maximum of 55.

Table 5.3.: VR datasets summary

and PR experiments was a result of logistical requirements, as performing 55 experiments between 17:00-20:00 (as had been performed in the PR experiments) would have taken several months. This not only could have produced confounding effects, but also represented an unacceptable timeline for this PhD. Furthermore, these experiments were completed only weeks before the initial national COVID-19 lockdown in the UK. As a result of the difference in timings, there was a potential for a difference in participant cortisol responses as a result of each individual's circadian rhythm. This was mitigated as much as possible by taking baseline measurements, and will be discussed further in Section 7.2.

The total number of ANS datasets obtained is provided in Table 5.4.

VR social belief

One of the requirements for the participants of the VR experiment was that they believed the avatars they observed were real participants. Prior to the experiment, the participants were informed that the avatars they would see in the experiment were controlled by other participants that were also located

Paradigm	Branch	E4 Empatica	Equivital EQ02	
			Base	Base+GSR
PR	Financial control	0	0	0
	Aggressor control	2	1	1
	Stress: prompt	8	5	1
	Stress: no-prompt	8	4	4
VR	Stress: movement	26	2	1
	Stress: no movement	27	0	0

Table 5.4.: ANS datasets summary

in the building, and that they would meet each other after the experiment. Further to this, each participant was identifiable by a name above their avatar's head. After completing the post-experiment questionnaires, participants were informed that the NPCs were in fact all computer controlled. The participants were then asked to rate on a 7-point Likert scale the degree to which they had believed the NPCs were controlled by humans (1=Not at all, 7=Completely). The results obtained are shown graphically below in Figure 5.1.

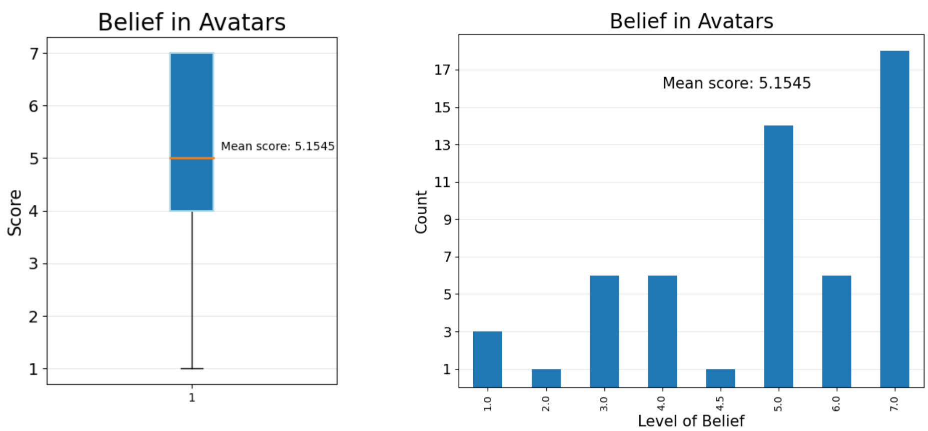


Figure 5.1.: VR participant social belief levels

As the above results indicate, the majority of the participants believed that the NPCs were human-controlled, with the most common occurrence indicating complete belief. It is unclear whether the participant responses that indicated a lack of belief were a result of a cognitive bias in desiring to appear to have understood the deception, or instead were a result of actually having understood the deception beforehand. Further to this, it is not known whether an individual's responses differ based on this belief level. As a consequence, the VR results will now be analysed identically to the PR results, assuming the social responses of participants are based on the same degree of belief in the humanity of the avatars as if they were in a PR experiment.

Future research might consider varying the level of information provided to participants prior to the experiment, to understand any changes in participant response based on their belief in the NPCs.

5.2. Data pre-processing

5.2.1. Screening

PR position data

The UWB system presented with near-field effects from the anchors and the tags, whereby if the tag was within approximately 2m of the anchor, then the localisation algorithm did not perform to specification. This caused the UWB system to estimate the tag as moving and accelerating at a high rate. These times were identified both manually (by tagging during a visualisation of the data, and comparing this data with the camera recordings of the experiment), and automatically (e.g. when the velocity of the participant was above 100m/s), and the incorrect data points removed.

After this process, the remaining positional data was analysed using the Python *traja* package [263]. This allowed the missing data to be interpolated and the resulting data points to be smoothed using a Savitzky-Golay filter, before resampling at 10Hz, providing a complete trajectory for each participant and the aggressor. The data points were combined into single snapshots of the complete experimental area every 0.1s, showing the position, orientation, velocity, and acceleration for all participants and the aggressor at that point in time. These snapshots will henceforth be referred to as 'frames'.

The data used to create these frames was truncated for each participant, starting the moment at which the aggressor entered the environment, and ending at the moment they were taken out of the experiment. If participants performed crypsis ('playing dead'), or moved out of the environmental area (e.g. behind the separating boards, or leaving the room), their positional data at these moments was removed from the analysis.

VR position data

There were no issues with the functionality of the VR position capturing, and consequently no resampling was required. The frames were compiled using the same method as the PR position data.

5.2.2. Rotation and centering

The compiled frames provided positional data, along with the heading, speed and acceleration (magnitude) vectors. The frames were then processed to represent the data from perspective of the aggressor. Each frame was centred on the position of the aggressor and rotated by the heading of the aggressor. This is shown below in Figure 5.2.

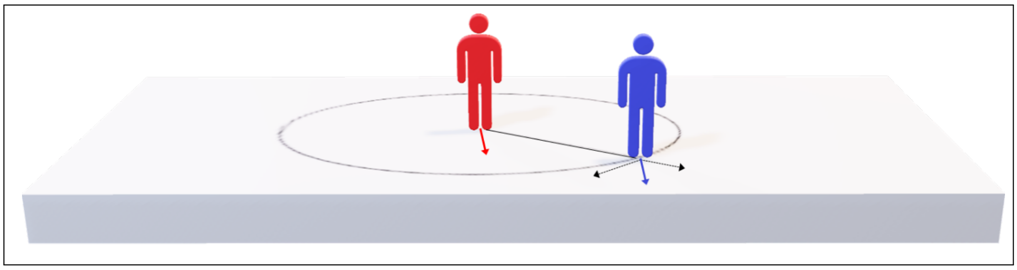


Figure 5.2.: Aggressor perspective analysis

5.3. Analysis types

5.3.1. Data identifiers

In order to compare between the separate branches of the experiment, several identifiers are used to isolate specific datasets. These are detailed in Table 5.5

Identifier	Data subset	Max number of whole datasets
Overall	A combination of all Physical Reality and Virtual Reality data points.	135
PR	Data from Physical Reality experiments only.	80
VR	Data from Virtual Reality experiments only.	55
Stress (also: Overall Stress)	Data from all stressful experiments, including both Physical Reality and Virtual Reality participants.	116
PR Stress	Data from the Physical Reality stressful experiments.	61
Control (also: Overall Control)	Data from all non-stressful experiments, including both Financial and Aggressor controls (note: these occurred in Physical Reality control experiments only).	19
Financial control	Data from the financial control Physical Reality experiment.	8
Aggressor control	Data from the aggressor control Physical Reality experiment.	11
PROMPT	Data from prompted experiments (note: these occurred in Physical Reality stressful experiments only).	31
NO-PROMPT	Data from non-prompted experiments (note: these occurred in Physical Reality stressful experiments only).	30
MOVE	Data from experiments were the avatars attempted to run from the aggressor (note: these occurred in Virtual Reality experiments only).	27

Identifier	Data subset	Max number of whole datasets
NO-MOVE	Data from experiments where the avatars did not move (note: these occurred in Virtual Reality experiments only).	28
Male	Data from self-identified male participants.	88
Female	Data from self-identified female participants.	47

Table 5.5.: Data identifiers

5.3.2. Statistical methods

Throughout this chapter significance levels will be reported using the following codes: *p<0.05, **p<0.01, ***p<0.001. Any result that is found at 1% significance or lower will be referred to as strongly significant, and any result that lies between 5-10% significance will be referred to as marginally significant, identified with the † symbol.

The reported significance codes are by default based on Welch's t-test, which is used to differentiate between samples of unequal variance and sample size, as a more robust substitute to the Student's t-test. Where the underlying data does not satisfy the conditions for Welch's t-test (for instance, normality), the significance codes are based on the Mann-Whitney U test, identified by the prefix "MW".

Welch's t-test

Welch's t-test defines the test statistic, t, between two groups with mean $\bar{X}_i$, sample standard deviation s_i, and sample size N_i using the following formula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{5.1}$$

Effect size	Classification
0-0.2	None
0.2-0.5	Small
0.5-0.8	Medium
0.8+	Large

Table 5.6.: Hodge's g* effect size

The number of degrees of freedom associated with this test is defined using the following approximation:

$$\nu \approx \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2\nu_1} + \frac{s_2^4}{N_2^2\nu_2}} \tag{5.2}$$

Where the individual degrees of freedom, $\nu_i = N_i - 1$. This definition accommodates different sample sizes and standard deviations which would otherwise skew the results of a Student's t-test. This test has an added benefit that, with equal sample sizes and standard deviations, it reduces into the Student's t-test.

Effect size

The effect size for Welch's t-test will be calculated using Hedge's g* effect size, defined as:

$$g = \frac{\bar{X_1} - \bar{X_2}}{s^*} \tag{5.3}$$

Where s^* is the pooled standard deviation estimator, defined as:

$$s^* = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}} \tag{5.4}$$

This effect size is very similar to Cohen's d, but with an added scaling factor to accommodate any small sample sizes. For large sample sizes, they become approximately equal. When interpreting the effect sizes for the Welch's t-test, this PhD will follow the established rules of thumb (see Table 5.6).

To avoid any reported data becoming confusing and/or unwieldy, any significant result that has either a medium or large effect size will have their p-value reported in **bold**.

Mann-Whitney U test

The Mann-Whitney U test is a non-parametric test that is used to determine the probability of the values of one group (X_1) being larger than the values of another, independent group (X_2). As a non-parametric test, it does not make any assumptions about the shape of the distribution of the groups, but rather checks the number of pairwise differences between the two groups, counting the number of times each value is higher than the values in the opposing group.

The U score of each group can be calculated as follows:

$$U_i = R_i - \frac{n_i\left(n_i + 1\right)}{2} \tag{5.5}$$

where R_i is the sum of the ranks of the values in the i^{th} group, and n_i is the sample size. The U_i score has a p-value associated with it, and for large sample sizes the U distribution can be approximated with a Normal distribution. This test is preferred to Welch's t-test for non-normally distributed and non-ordinal datasets, as it is not dependent on outliers and has a lower Type II error probability.

Effect size

The effect size for the Mann Whitney U test is calculated as follows:

$$r_i = \frac{Z_i}{\sqrt{n_1 + n_2}} \tag{5.6}$$

where Z_i is the normal approximation of U_i, the Mann-Whitney score, which is calculated as follows:

$$Z_i = \frac{U_i - m_U}{\sigma_U}$$

$$m_U = \frac{n_1 n_2}{2}$$

$$\sigma_U = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}$$

Effect size	Classification
0-0.1	None
0.1-0.3	Small
0.3-0.5	Medium
0.5+	Large

Table 5.7.: Mann Whitney effect size

When interpreting the effect size, the established rules of thumb (see Table 5.7) will be used.

As with the Welch's t-test, in order to avoid any reported data becoming confusing and unwieldy, any significant result that has either a medium or large effect size will have its p-value reported in **bold**.

Standardisation

Whenever datasets need to be compared across different metrics, they are often standardised, ensuring the mean value is 0, and the standard deviation is 1. This is done using the following standardisation formula:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{5.7}$$

5.3.3. Normality tests

For the next step in this analysis, it is first necessary to test the data for normality. As can be seen in Table 5.8, the majority of this data can be considered normal. The normality of the data was assessed using three separate tests: the Shapiro-Wilk, Kolmogorov-Smirnov and D'Agostino-Pearson Omnibus tests. If the dataset passes at least two tests (at 5% significance), then it will be considered normal. If a dataset only passes one of the three tests, further investigation with plots will be performed, and the normality will be assessed qualitatively. If the dataset does not pass any of the tests, it will be considered non-normal.

As stated earlier, Welch's t-test will be used for all normal datasets, while the Mann Whitney U test will be performed on any non-normal datasets.

Dataset		Sample size	Shapiro-Wilk score	Kolmogorov-Smirnov score	D'Agostino-Pearson score	Treated as normal?
STAI-T		134	0.0009***	0.0167**	0.0926†	N (see below)
Delta-STAI		134	0.0078**	0.0879†	0.4108	Y
PANAS-N		131	3.70e-05***	0.0865†	0.0019*	N (see below)
PANAS-P		131	0.6612	0.7760	0.6503	Y
Personality	A	132	0.05357	0.1400	0.1266	Y
	C	132	0.1134	0.6139	0.1510	Y
	E	132	0.0431*	0.6279	0.1429	Y
	N	132	0.1710	0.2187	0.6215	Y
	O	132	0.1020	0.4463	0.1580	Y
Social responses	Solidarity	134	0.0012**	0.0189*	0.0355*	N
	Centrality	134	5.81e-07 ***	0.0032**	0.0054**	N
	Satisfaction	133	0.0069**	0.0185*	0.0934†	N (see below)
	In-group Homogeneity	133	0.0005***	0.0007***	0.3806	Y (see below)
	Individual Self-stereotyping	133	0.0057**	0.0273*	0.1536	Y (see below)
	Shared Identity	133	0.1638	0.1827	0.5252	Y
Cortisol	Baseline	134	1.99e-20 ***	6.75e-10 ***	1.11e-43***	N
	Relative 20-mins	134	4.17e-09***	0.0042**	0.000165***	N
	Relative 40-mins	134	4.17e-12***	0.0027**	1.18e-08***	N
	Relative 60-mins	134	6.79e-14***	0.0008***	2.01e-22***	N
	Average Relative value	134	7.19e-11***	0.0072***	2.61e-14***	N

Dataset		Sample size	Shapiro-Wilk score	Kolmogorov-Smirnov score	D'Agostino-Pearson score	Treated as normal?
	Maximum Relative value	134	2.50e-17***	1.21e-10***	4.02e-24***	N
	Maximum-Minimum difference	134	2.03e-15***	1.62e-05***	5.21e-26***	N
Age		135	5.73e-07***	0.0447*	0.0009***	N
Exercise level (1-5)		135	8.01e-10***	6.99e-05***	0.0307*	N

Table 5.8.: Dataset normality tests

The following datasets failed two of the three tests: Overall Pre-STAI, Overall PANAS-N, Overall Satisfaction, Overall In-group Homogeneity, Overall Individual Self-stereotyping, and Overall Cortisone Relative 40, Relative 60 and Average Relative values. These datasets were then assessed individually based on their Q-Q plots. Examples shown below in Figure 5.3 and Figure 5.4, and the remaining plots are provided in the Appendix A.

As a result of the normality checks, the Overall In-group Homogeneity and Overall Individual Self-stereotyping datasets were classified as normal, and the remaining datasets were classified as non-normal.

For the next section, all normally distributed datasets were compared using Welch's t-test, whereas any non-normally distributed datasets were compared using the Mann-Whitney U test. Any comparisons between normally and non-normally distributed datasets were also analysed using the Mann-Whitney U test. All Mann-Whitney U test comparisons will be identified using the prefix "MW".

5.3.4. Movement analysis

Activity splits

The investigators qualitatively observed several different activity types during the PR and VR experiments. The most obvious activity was the decision by

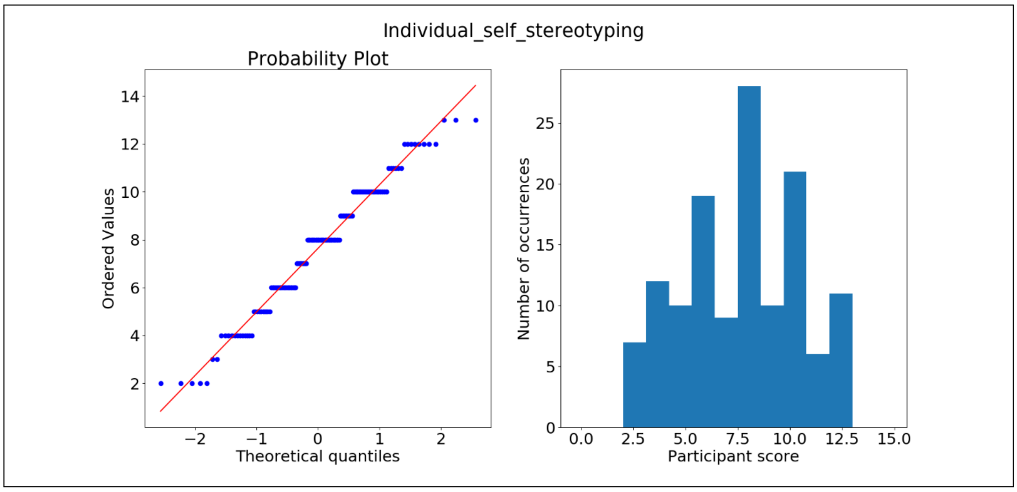


Figure 5.3.: Q-Q plot for Individual self-stereotyping

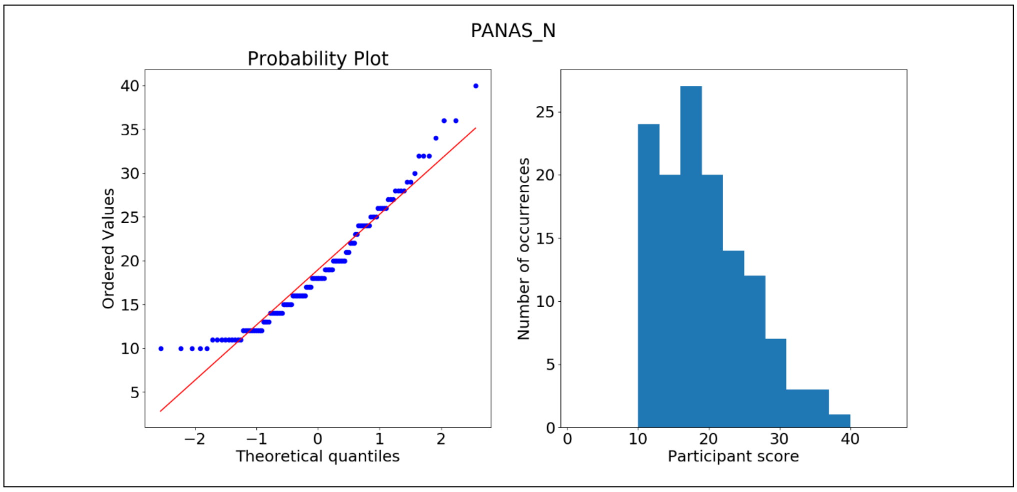


Figure 5.4.: Q-Q plot for PANAS-N

participants to move at all, which will be discussed in the discrete movement section.

Further activity choices included:

- Hiding behind the central column (majority of moving participants)
- Crypsis ('playing dead') (3 participants)
- Evacuating the room (5 participants)
- Moving behind the dividing wall (2 participants)

- Protecting other participants (1 participant after they had been taken out of the experiment)

Example of these activity choices are shown below in Figure 5.5, Figure 5.6, and Figure 5.7.

Figure 5.5.: Participant performing crypsis (grey top, next to central column)

The VR environment allowed for the majority of the above behaviours. However, it did not permit the following discrete behaviours:

- Crypsis ('playing dead')
- Protecting other participants

This is because these behaviours are particularly complex, and not readily implementable. They also provide further interactions between participants and avatars, which creates a further risk of breaking the immersion of the participant.

<u>Discrete movement</u>

The major qualitative observation from the experiments was the tendency for participants to react in response to the actions of the surrounding participants. This was most obviously present in the decision to move away or remain stationary as the aggressor approached them.

Figure 5.6.: Participant evacuating the room (right hand side)

In keeping with previous work [264], these discrete choices were analysed with a logistic regression model. The data for this model was obtained by analysing the frame each time a participant decided to move or was touched by the aggressor. For more information on how this data was extracted, see Section 8.2

Continuous movement

Multivariable linear regression was used to identify trends in the continuous data. After having centred and rotated the data to provide the perspective of the aggressor, it was possible to understand the impact of relative position on velocity and acceleration, before splitting up these terms into radial and tangential components. To understand any asymmetry, linear and non-linear combinations of the relative x and y position variables were used, while the models used a forward-backward stepwise selection process, selecting only the statistically significant variables.

5.3.5. Self-assessed data analysis

There were several types of analysis performed on the self-assessed measures. Initially, simple Welch's t-tests were performed, accommodating different sample sizes and variances. Further to this, linear regression models were created

Figure 5.7.: Participant protecting another

Value		PR (n=80)	VR (n=55)
Gender	M	54	37
	F	26	18

Table 5.9.: Gender split between paradigms

to understand the relationships between predictor variables and responses.

All analysis was performed within Python, using several open-source packages including *traja* [263], *statsmodels*[265], *pandas*[266], *sklearn*[267], *scipy*[268], and *numpy*[269].

5.4. Participant data

The summary of the participant demographics is shown below, separated by experimental paradigm. Any significant differences are indicated, with medium and large effect sizes in bold.

Table 5.10 shows that the major significant difference between the two experimental paradigms is in the age of the participants. This could be due to the time of year of the two experiments (the PR experiment took place in December 2018 during a pre-January revision period, while the VR experiment took place in January-March 2020, after university exams had ended), or due

Value		PR (n=80) Mean (SD, range)	VR (n=55) Mean (SD, range)	Test statistic (p-value)
Age		25.0 (3.88, 18.4-34.8)	21.3 (2.89, 21.3-34.3)	**MW 647.0** **(6.17e-09)*****
Personality	A	2.82 (0.63, 1.1-4)	2.91 (0.52, 1.4-4)	-0.8915 (0.374)
	C	2.61 (0.66, 1.1-4)	2.52 (0.62, 1.1-3.6)	0.8192 (0.414)
	E	2.13 (0.76, 0.8-4)	2.21 (0.72, 0.7-3.8)	-0.5429 (0.588)
	N	2.24 (0.75, 0.4-3.7)	2.42 (0.80, 0.8-4)	-1.300 (0.196)
	O	2.63 (0.53, 1.4-3.8)	2.73 (0.51, 0.8-3.8)	-1.097 (0.275)
Height		173.4 (9.16, 153-198)	174.6 (10.8, 151-200)	0.2501 (0.803)
Weight		67.4 (12.3, 38-96)	67.2 (10.3, 45.5-89)	0.5623 (0.575)
Exercise level (1-5)		3.66 (1.21, 1-5)	3.65 (1.14, 1-5)	MW 1676.0 (0.4977)
Pre experiment STAI-T		39.1 (8.33, 23.3-60.0)	36.91 (7.79, 23.3-56.7)	MW 1328.5 (0.035)*

Table 5.10.: Demographic split between paradigms

Value		Male (n=88) Mean (SD, range)	Female (n=47) Mean (SD, range)	Test statistic (p-value)
Age		23.7 (3.97, 18.5-34.3)	23.1 (3.97, 18.4-34.7)	MW 1875.5 (0.1649)
Personality	A	2.75 (0.566, 1.1-4)	3.08 (0.567, 1.7-4)	**-3.2720 (0.0015)****
	C	2.56 (0.643, 1.1-3.8)	2.62 (0.659, 1.1-4)	-0.5736 (0.5676)
	E	2.15 (0.782, 0.7-3.9)	2.20 (0.681, 1-4)	-0.38021 (0.7045)
	N	2.50 (0.682, 1.2-4)	1.98 (0.816, 0.4-3.7)	**3.6891 (0.0004)*****
	O	2.71 (0.519, 0.8-3.8)	2.60 (0.536, 1.4-3.6)	1.12544 (0.2633)
Height		178.7 (7.61, 165-200)	165.1 (6.90, 151-184)	**10.5689 (3.138e-18)*****
Weight		72.1 (10.4, 38-96)	58.7 (7.71, 45-76)	**8.4926 (5.874e-14)*****
Exercise level (1-5)		3.74 (1.19, 1-5)	3.49 (1.15, 1-5)	MW 1806.0 (0.0886)†
Pre experiment STAI-T		37.6 (8.24, 23.3-56.7)	39.7 (8.28, 23.3-60)	MW 1718.0 (0.0520)†

Table 5.11.: Demographic split between genders

to the paradigm itself, as VR may attract a younger demographic.

Further to this, there is a difference in the pre-experiment STAI-T between VR and PR participants, with PR participants reporting slightly higher levels of pre-experiment stress. It is possible that this difference is an effect of the presence of a group, or a result of the different experimental paradigms, or it could be an inherent difference in the participants themselves.

However, the overall differences between paradigms are not considered extreme, as the only differences were found in age (strongly significant and notable effect size) and pre-experiment STAI-T (significant but not a large effect size). It can therefore be concluded that, given the large overlap between the two experimental participant groups, there is a valid basis for future comparisons between the two groups.

As can be seen in Table 5.11, there are some significant differences between the male and female participants. Firstly in the results from the Big-5 questionnaire, females self-assess as more Agreeable and less Neurotic; secondly, the female participants were generally shorter and lighter; thirdly, marginal ($p<0.1$) differences indicated that the female participants exercised less and reported a higher level of anxiety in the pre-experiment STAI-T. While the differences in height and weight were expected as demographic differences, the differences between personality values, exercise level and STAI-T are unexpected, and could present an interesting source of data for sociological research on the staff and students of Imperial College London. For the purposes of this PhD, however, these differences will not be investigated further.

5.5. Chapter summary

This chapter has detailed the logistical elements of the experiments that were performed as part of this PhD, as well as the datasets obtained as a result. It continues by describing the different statistical measures that will be used in the remainder of the thesis, including Welch's t-test (for normal datasets) and the Mann-Whitney U test (for non-normal datasets). The comparison of the datasets showed that some subsets should be treated as non-normal, using a combination of Shapiro-Wilk, Kolmogorov-Smirnov, and D'Agostino Pearson tests, alongside Q-Q plots.

The VR data is underpinned by the levels of participant belief indicating that they treated the NPCs as though they were real humans. Analysis of these scores concluded that these scores suggested that the majority of participants fully treated the NPCs as real human participants. There was no distinction between the remaining scores where participants actually treated NPCs and computer-controlled, underestimated their levels of belief, or had a combination of low levels of belief but maintained their treatment of NPCs as though they were human regardless. Therefore the VR datasets could be compared with the PR experiment datasets as though the participants had the same social stimuli.

The final assessment showed that there were minimal inter-paradigm differences but some systematic intra-paradigm differences in the demographic of the participants, specifically between male and female participants. Any differences between these participant groups will therefore be assessed at all points in the following chapters.

The next four chapters will describe the quantitative analysis of these datasets, initially quantifying the participants' levels of emotional response from both a subjective and objective perspective, before examining their movement responses and finally their recall quality three months after their experiment.

6. Results: Internal Subjective

The previous chapters focused on the implementation of an experimental methodology to investigate human behavioural responses to an unexpected hostile aggressor, detailing the procedures and the datasets obtained. This chapter will describe the analysis undertaken on the resulting data, specifically focusing on the Internal Subjective data.

This chapter will first focus on understanding the differences in the self-assessed emotional states of the participants, specifically through the State-Trait Anxiety Inventory (STAI) and the Positive And Negative Affect Schedule (PANAS). This will be followed by an investigation into the social responses of the experiments. By analysing the results from the Group Identification and Shared Identity questionnaires that were provided to participants after each experiment, this chapter will assess any differences in social subjective responses as a result of a stressor, as well as any differences between the experimental paradigms.

Both investigations will compare differences between all experimental subgroups and both experimental paradigms (PR and VR), as well investigating any relationships between the results and participant demographics. The differences investigated here are presented relative to the control experiment levels, which are representative of day-to-day responses, and any differences identified are assumed to be caused by the experimental process.

Finally, this chapter will summarise and discuss the key results found within these datasets, including a discussion of any limitations of the results, and their implications. Throughout the chapter, the datasets will be split into subgroups according to several different elements. The subsequent datasets will have associated identifiers, such as 'PR' and 'Stress'. These identifiers are detailed in Subsection 5.3.1, Table 5.5.

6.1. Self-assessed measures- descriptive statistics

The State-Trait Anxiety Inventory (STAI) and the Positive and Negative Affect Schedule (PANAS) are two regularly used methods of assessing internal emotional states. They both have large amounts of validation work supporting their use in various contexts and have been used extensively over the past decades.

The STAI survey assesses negative emotional states (fear and stress) in either Trait (STAI-T) or State (STAI-S) form. The STAI-T measures negative emotional states over a longer time period, providing an assessment of long-term emotional states (e.g. depression). In contrast, the STAI-S measures negative emotional state at specific points in time. Both surveys come in long or short form versions, and the responses are used to calculate a single number for each participant.

The PANAS survey assesses both positive and negative emotional states, with participant responses used to calculate two separate numbers: one for Positive Affect (P), and one for Negative Affect (N). It also comes in long and short form versions.

Both PR and VR experimental paradigms used a short form STAI-T prior to the experiment, and a short form STAI-S short form and short form PANAS after the experiment. These surveys are provided in Appendix C. By performing both the STAI-T and the STAI-S, it was possible to assess the difference in the participants' scores measured by the STAI as a direct response to the experiment itself. This dataset, called the Delta-STAI, will be used alongside the PANAS-P and PANAS-N scores as the primary indicators of the psychological emotional state.

As the PANAS-N was assessed to be non-normal (see Subsection 5.3.3), it will be analysed throughout using the Mann-Whitney U-test, indicated by the prefix "MW".

6.1.1. PR data

The initial self-assessed emotional state measures in the experiments are the Delta-STAI (the difference between the STAI-T and the STAI-S), the PANAS-P, and the PANAS-N. To begin with, comparisons will be made between the control and experimental groups in the PR paradigm. As can be seen in Table 6.1, there are strongly significant differences observed between the levels of

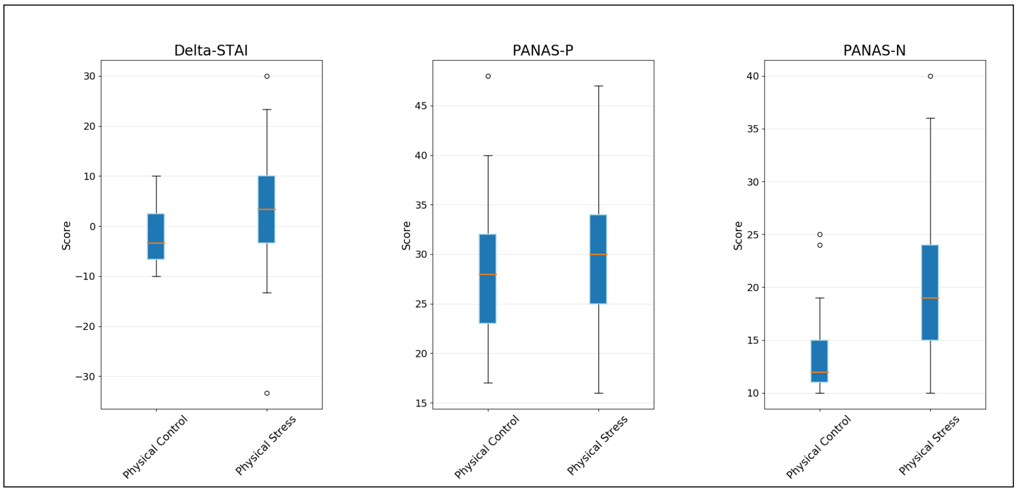


Figure 6.1.: Differences in emotional measures between PR experimental groups and control groups

Measure	PR Control (mean, SD, range)	PR Stress (mean, SD, range)	Test statistic (p-value)
Delta-STAI	-2.41 (6.35, -10-10)	4.389 (10.8, -33.3-30)	**3.329 (0.002)****
PANAS-P	28.47 (7.42, 17-48)	30.22 (6.68, 16-47)	0.9145 (0.3681)
PANAS-N	14.1 (4.43, 10-25)	20.2 (6.40, 10-40)	**MW 199.5 (8.61e-06)*****

Table 6.1.: Differences in emotional measures between PR experimental groups and control groups

emotional state measured by both the Delta-STAI ($p<0.01$) and the PANAS-N ($p<0.001$). However, there are no significant differences in the PANAS-P, indicating that PANAS-P and PANAS-N are not negatively correlated. Further evidence for this will be detailed below. The effect sizes of the differences in Delta-STAI and PANAS-N are larger than 0.5 and 0.3 respectively, indicated by the bold font for the test statistic, is classified as a notable result (see Section 5.3 for a definition).

The next analysis will investigate any differences seen as a result of the experimental interventions. The intervention within the PR paradigm involved the actor-participant (who was always the first participant to be approached) either remaining still as the aggressor approached them, or attempting to

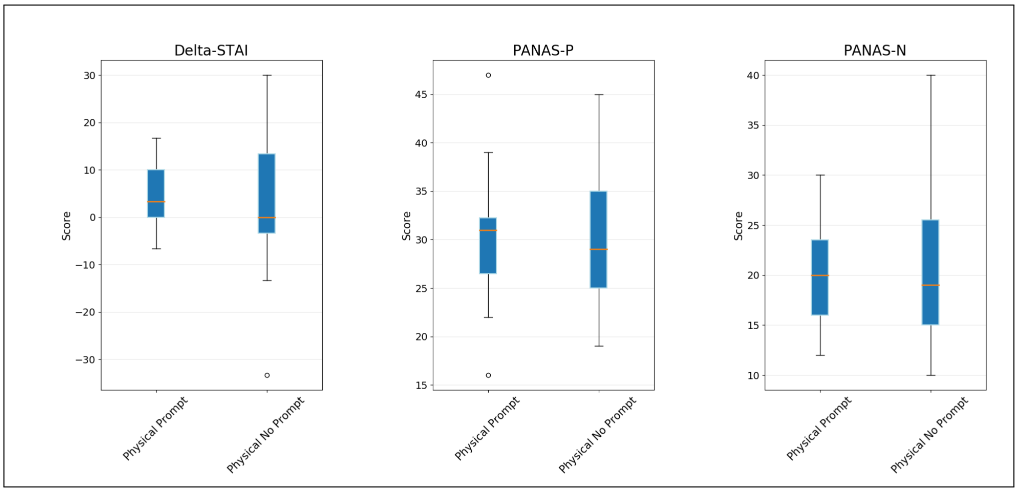


Figure 6.2.: Differences in emotional measures between PR PROMPT and NO-PROMPT groups

Measure	**PR PROMPT (mean, SD, range)**	**PR NO PROMPT (mean, SD, range)**	**Test statistic (p-value)**
Delta-STAI	5.17 (6.58, -6.67-16.7)	3.66 (13.6, -33.3-30.0)	0.5537 (0.5825)
PANAS-P	30.3 (6.00, 16-47)	30.2 (7.38, 19-45)	0.0242 (0.9807)
PANAS-N	19.8 (4.69, 12-30)	20.5 (7.69, 10-40)	MW 421.5 (0.2672)

Table 6.2.: Differences in emotional measures between PR PROMPT and NO-PROMPT groups

evade the aggressor. These different experiments were coded as PROMPT and NO-PROMPT. Table 6.2 shows that there are no significant differences in self-assessed emotional state as a result of this intervention.

6.1.2. VR data

Within the VR experimental paradigm, the two separate experimental groups were coded as MOVE and NO-MOVE. The differences between these groups lies within the movement of the NPC avatars in response to the aggressor, where the MOVE avatars followed pre-defined trajectories that were obtained in the PR experiment, and the NO-MOVE avatars remained still. Both of

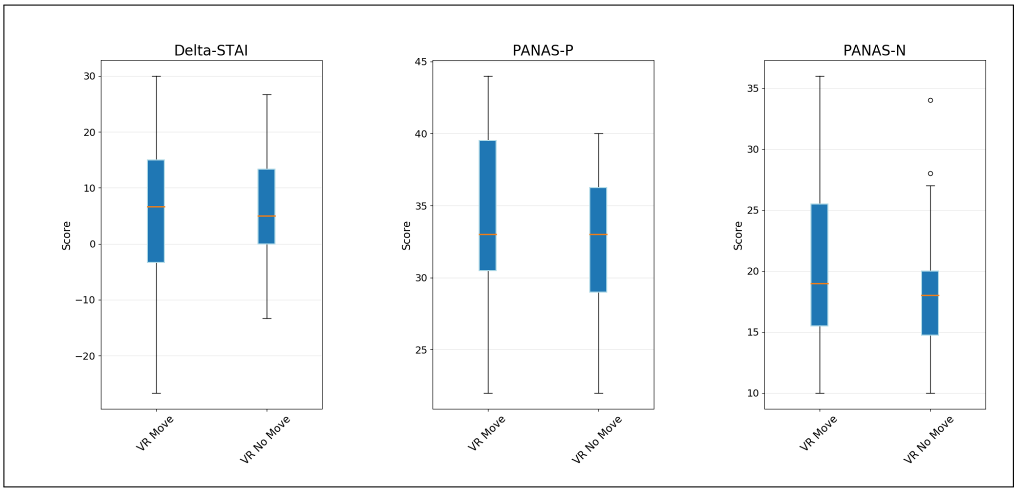


Figure 6.3.: Differences in emotional measures between VR MOVE and NO-MOVE groups

Measure	VR MOVE (mean, SD, range)	VR NO MOVE (mean, SD, range)	Test statistic (p-value)
Delta-STAI	7.284 (13.0, -26.7-30)	5.60 (10.2, -13.3-26.7)	0.5331 (0.2042)
PANAS-P	34.2 (6.36, 22-44)	32.4 (5.00, 22-40)	1.136 (0.2614)
PANAS-N	20.3 (7.13, 10-36)	18.6 (5.39, 10-34)	MW 328.5 (0.2042)

Table 6.3.: Differences in emotional measures between VR MOVE and NO-MOVE groups

these trajectory sets occurred in the PR experiments, providing internal validation to the VR experimental design.

Further analysis (detailed in Table 6.3) shows that there are no significant differences in the level of emotions felt during the VR MOVE and NO-MOVE conditions.

6.1.3. Combined data

The next comparison is between the VR experiments and the PR experiments. As there were no control experiments within the VR paradigm, the initial comparison will be between the VR data and the PR experimental group data (using the identifiers “PR” and “Stress”). As can be seen below in

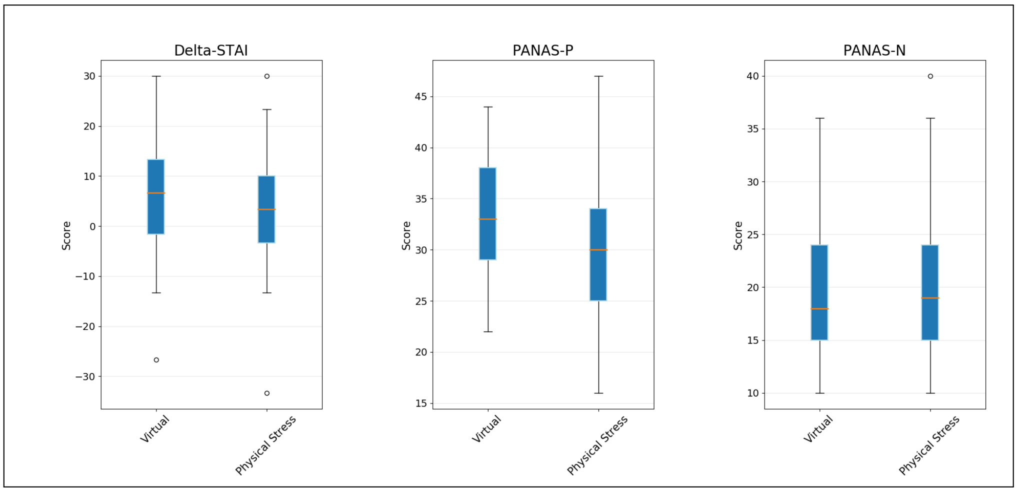


Figure 6.4.: Differences in emotional measures between PR experimental groups and VR participants

Measure	PR-Stress (mean, SD, range)	VR (mean, SD, range)	Test statistic (p-value)
Delta-STAI	4.389 (10.8, -33.3-30)	6.424 (11.6, -26.7-30)	-0.9721 (0.3331)
PANAS-P	30.22 (6.68, 16-47)	33.3 (5.72, 22-44)	-2.608 (0.0104)*
PANAS-N	20.2(6.40, 10-40)	19.4 (6.31, 10-36)	MW 1470.0 (0.1258)

Table 6.4.: Differences in emotional measures between PR experimental groups and VR participants

Table 6.4, the only significant difference between the two datasets is in the PANAS-P. This indicates that there were no significant differences in the level of negative emotional states, assessed by STAI and PANAS-N, in the PR and VR experimental paradigms. Furthermore, it indicates that the participants in VR experienced a higher degree of positive emotion than those in the PR experiments. This will be discussed further below.

In assessing the figures and tables above, it is possible to see that there are minimal differences as a result of the different experimental paradigms, and that consequently it is possible to combine datasets. This will allow the comparison of a larger dataset (i.e. every participant who was stressed in both PR and VR experiments) with the control experiments, providing a larger

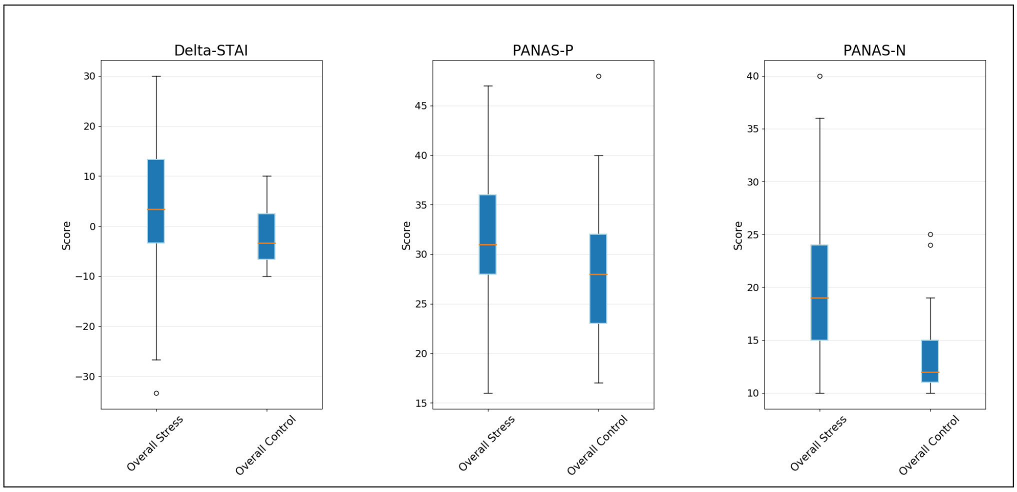


Figure 6.5.: Differences in emotional measures between Overall Stress and Control participants

Measure	Overall-Stress (mean, SD, range)	Overall Control (mean, SD, range)	Test statistic (p-value)
Delta-STAI	5.36 (11.2, -33 to 30)	-2.41 (6.35, -10 to 10)	**4.262 (0.00013)*****
PANAS-P	31.7 (6.39 16-47)	28.47 (7.42, 17-48)	1.804 (0.0844)
PANAS-N	19.8 (6.34, 10-40)	14.1 (4.43, 10-25)	**MW 443.5 (1.52e-05)*****

Table 6.5.: Differences in emotional measures between Overall Stress and Control participants

degree of certainty in the results. This analysis will therefore compare two datasets, Overall-Stress and Overall-Control, detailed in Table 6.5. It is worth noting that the Overall-Control dataset is identical to PR-Control dataset.

Table 6.5 shows that there are strongly significant differences in the self-reported levels of negative emotional state between the stressful and non-stressful conditions, as measured by the Delta-STAI and the PANAS-N. It can also be seen that there is only a marginally significant ($p<0.1$) difference between the two conditions in the PANAS-P, indicating that the levels of positive emotions experienced by the participants were only slightly mediated by the negative emotions experienced in the experimental conditions. However, this is potentially an artefact of the combination of the VR and PR experimental groups, as the VR PANAS-P scores were significantly higher than the PR experimental group scores.

6.2. Self-assessed measures- inferential statistics

The next section will report on any correlations between the measures described above. The initial correlation assessed will be the investigation into the relationship between two measures of negative emotional state, namely the Delta-STAI and the PANAS-N. Two outliers were observed, shown in Figure 6.6. One of these outlying data points was identified as coming from a participant who, after the experiment aims were revealed to them, explained that they had been present during the 2008 Mumbai attack, and therefore this experiment represented for them a much lower level of stressor than would have been necessary to provoke a high level of negative internal emotion. The exclusion criteria for the experiments could not filter out this participant without raising the suspicions of the participant, and these criteria were considered acceptable by the Imperial College Research Ethics Committee.

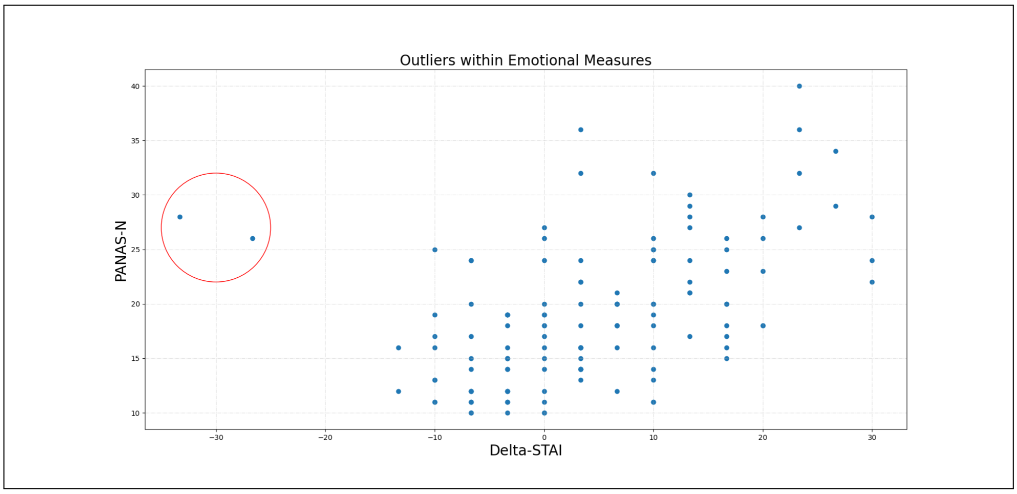


Figure 6.6.: Outliers in Delta-STAI and PANAS-N

6.2.1. Emotional Measures and demographic

The final emotional state analysis investigates any relationships with participant demographic. A multiple regression model was generated using all demographic predictor inputs, with the emotional state of the participants in

Overall PANAS-P				
Variable	**Coefficient**	**Std error**	**t-statistic**	**p-value**
Constant	32.3690	2.046	15.824	<0.000***
Overall PANAS-N	-0.0815	0.114	-0.718	0.433
Overall Delta-STAI	0.0559	0.071	0.786	0.474
Adjusted r^2			-0.010	
F-statistic (p-value)			0.3597 (0.699)	
Omnibus (p-value)			0.359 (0.836)	

Table 6.6.: Correlation between Overall PANAS-P and remaining emotional states

Overall PANAS-N				
Variable	**Coefficient**	**Std error**	**t-statistic**	**p-value**
Constant	18.648	2.269	8.219	<0.000***
Overall Delta-STAI	0.3671	0.046	8.009	<0.000***
Overall PANAS-P	-0.0512	0.071	-0.718	0.474
Adjusted r^2			0.333	
F-statistic (p-value)			32.15 (5.87e-12)***	
Omnibus (p-value)			13.963 (0.001)***	

Table 6.7.: Correlation between Overall PANAS-N and remaining emotional states

Overall Delta-STAI				
Variable	**Coefficient**	**Std error**	**t-statistic**	**p-value**
Constant	-15.3664	4.286	-3.585	<0.000***
Overall PANAS-N	0.9337	0.117	8.009	<0.000***
Overall PANAS-P	0.0893	0.114	0.786	0.433
Adjusted r^2			0.333	
F-statistic (p-value)			32.15 (5.58e-12)***	
Omnibus (p-value)			0.232 (0.890)	

Table 6.8.: Correlation between Overall Delta STAI and remaining emotional states

Variable	PANAS-N	PANAS-P	Delta-STAI
Constant	23.295***	19.605**	10.948
Personality A	-0.047	1.641	3.292
Personality C	1.280	0.714	-0.423
Personality E	1.435†	0.261	1.138
Personality N	-2.642**	2.146*	2.651†
Personality O	-1.233	1.579	-4.587*
Age	-0.117	-0.139	-0.473†
F	-0.903	-0.706	-2.903
Exercise level	0.726	-0.208	0.441
Adjusted r^2	0.0739	0.0713	0.055
F-statistic (p-value)	2.0878 (0.0437*)	2.0460 (0.04829*)	1.81376 (0.08261†)
Omnibus test (p-value)	13.133 (0.00140**)	0.41797 (0.81141)	3.4324 (0.179747)

Table 6.9.: Correlation between emotions and demographic

stressful experiments as the dependent variable. As is seen below in Table 6.9, significant relationships were observed between the following emotion sets and demographic types:

- Negative emotions:
  - PANAS-N: Neuroticism (strong negative), Extroversion (marginal positive)
  - Delta-STAI: Openness (negative), Neuroticism (marginal positive), Age (marginal negative)
- Positive emotions
  - PANAS-P: Neuroticism (positive)

No significant relationship was seen between emotional state and Agreeableness, Conscientiousness, Gender, or Exercise level.

6.3. Social responses- descriptive statistics

Both experimental paradigms used a combination of Group Identification and Shared Identity measures as part of the post-experiment questionnaires. The next set of results will detail any differences in this level of group identification and shared identity during the different experimental paradigms and conditions. Since the subsections “Solidarity”, “Centrality” and “Satisfaction” were deemed to be non-normal (see Section 5.3 for further details), all comparisons within these datapoints will be performed using the Mann-Whitney U-test and will identified by the prefix “MW”.

6.3.1. PR data

Table 6.10 shows that there are no significant differences between the PR experimental groups and PR Control groups in any of the measures of group identification.

Further to this, Table 6.11 shows that there are no significant differences introduced by the interventions within the PR experimental paradigm.

6.3.2. VR data

Table 6.12 shows a marginally significant ($p<0.1$) difference in the level of reported Centrality between participants within the VR MOVE and VR NO-MOVE conditions, with VR MOVE participants reporting slightly lower scores. The Centrality score is obtained by summing the responses (on a 1-7 Likert scale) to the following statements:

- I often think about the fact that I am in this group.
- The fact that I am in this group is an important part of my identity.
- Being in this group is an important part of how I see myself.

6.3.3. Combined data

In order to combine VR and PR data, it is first necessary to establish any pre-existing differences between the datasets. These differences are shown in Table 6.13, where significant differences are found in the levels of Solidarity and strongly significant differences in the levels of In-group homogeneity between the PR Stress and VR participants. There are marginally significant ($p<0.1$) differences between the levels of Satisfaction within these groups. The difference in In-group homogeneity has a large effect size, indicated by the bold font. Table 6.14 shows that, after combining the PR Stress and VR datasets

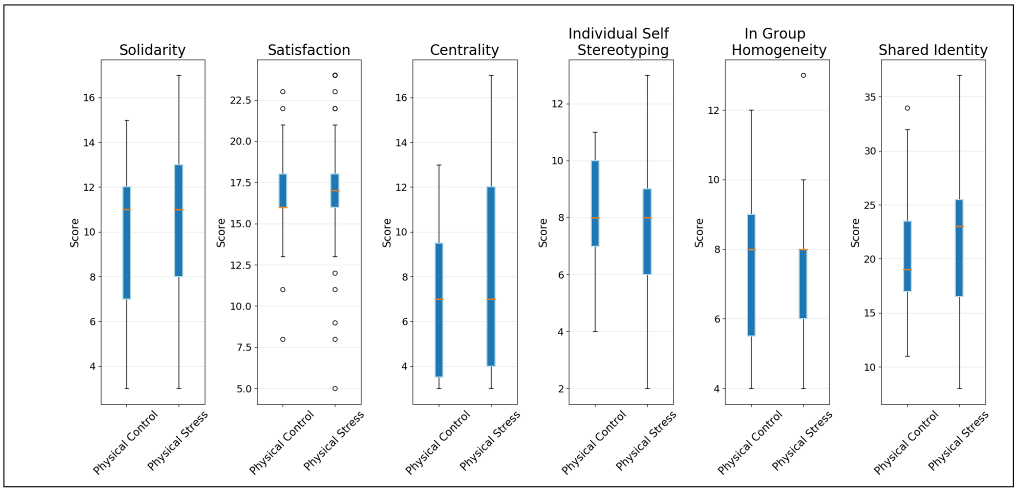


Figure 6.7.: Differences in social responses between PR Stress and Control participants

Measure	PR-Stress (mean, SD, range)	VR (mean, SD, range)	Test statistic (p-value)
Solidarity	10.3 (3.55, 3-17)	9.68 (3.64, 3-15)	MW 513.5 (0.2275)
Centrality	7.96 (4.10, 3-17)	6.89 (3.36, 3-13)	MW 492.5 (0.1621)
Satisfaction	16.8 (3.55, 5-24)	16.6 (3.76, 8-23)	MW 561.5 (0.4205)
Individual self-stereotyping	7.53 (2.67, 3-13)	8.05 (2.30, 4-11)	-0.8348 (0.4094)
In-group homogeneity	7.37 (1.86, 4-13)	7.57 (2.43, 4-12)	-0.3387 (0.7377)
Shared identity	21.5 (6.63, 8-37)	20.0 (6.08, 11-34)	0.9089 (0.3699)

Table 6.10.: Differences in social responses between PR experimental and control groups

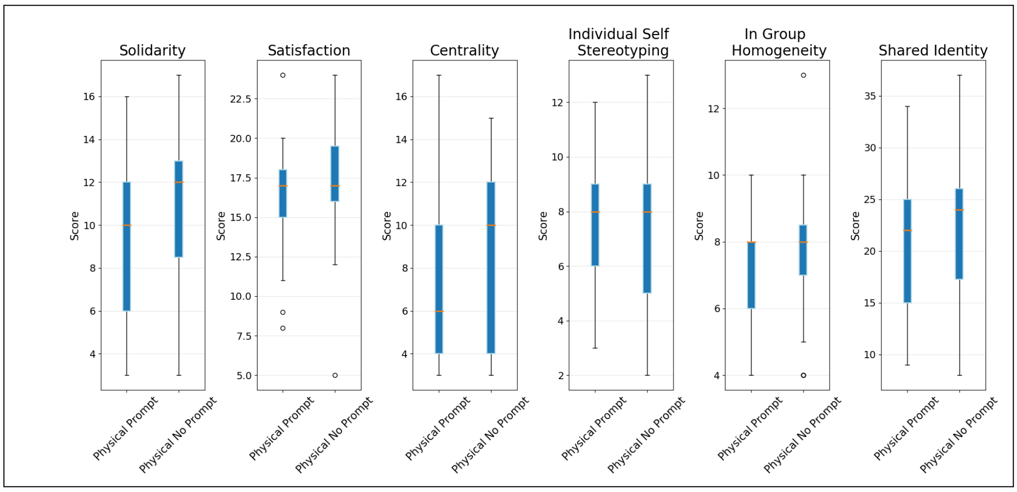


Figure 6.8.: Differences in social responses between PR PROMPT and NO-PROMPT groups

Measure	PR PROMPT (mean, SD, range)	PR NO PROMPT (mean, SD, range)	Test statistic (p-value)
Solidarity	9.69 (3.51, 3-16)	10.8 (3.56, 3-17)	MW 401.0 (0.1777)
Centrality	7.38 (4.25, 3-17)	8.50 (3.95, 3-15)	MW 411.5 (0.2205)
Satisfaction	16.2 (3.25, 8-24)	17.4 (3.77, 5-24)	MW 416.5 (0.2418)
Individual self-stereotyping	7.79 (2.30, 3-12)	7.27 (3.00, 2-13)	0.7565 (0.4526)
In-group homogeneity	7.00 (1.74, 4-10)	7.71 (1.92, 4-13)	-1.489 (0.1420)
Shared identity	21.0 (6.80, 9-34)	22.0 (6.54, 8-37)	-0.5953 (0.5540)

Table 6.11.: Differences in social response measures between PR PROMPT and NO-PROMPT groups

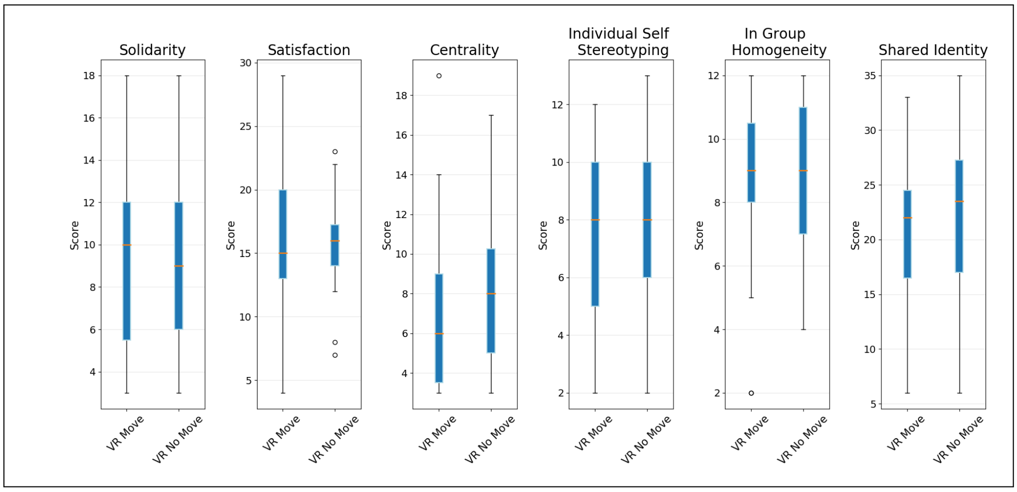


Figure 6.9.: Differences in social responses between VR MOVE and NO-MOVE participants

Measure	**VR MOVE (mean, SD, range)**	**VR NO MOVE (mean, SD, range)**	**Test statistic (p-value)**
Solidarity	9.15 (4.19, 3-18)	9.21 , (3.77, 3-18)	MW 375.5 (0.4865)
Centrality	7.07 (4.27, 3-19)	8.64 (4.31, 3-17)	MW 287.5 (0.0636)†
Satisfaction	16.3 (5.74, 4-29)	15.9 (3.67, 7-23)	MW 371.0 (0.4563)
Individual self-stereotyping	7.48 (2.97, 2-12)	7.75 (2.63, 2-12)	-0.3546 (0.7244)
In-group homogeneity	8.89 (2.62, 2-12)	8.89 (2.28, 4-12)	-0.0060 (0.9953)
Shared identity	20.2 (6.31, 6-33)	22.6 (7.39, 6-35)	-1.289 (0.2032)

Table 6.12.: Differences in social response measures between VR MOVE and NO-MOVE groups

and comparing them against the control groups, all previously significant differences have disappeared.

The In-group homogeneity score is comprised of the responses (on a 1-7 Likert Scale) to two statements:

- In this group, people have a lot in common with each other.
- In this group, people are very similar to each other.

This result may be a product of the VR environment, since the avatars observed by the participants fell into two separate categories, male and female, but were visually identical within these categories (as seen in Figure 4.9), potentially leading to an increase in perceived homogeneity.

In contrast, the Solidarity score is comprised of the responses to the following questions:

- I feel a bond with the group
- I feel solidarity with the group
- I feel committed to the group

It is considered that the significant difference in Solidarity scores may be a result of the different experimental paradigms, with physically present groups producing a higher level of Solidarity than VR groups. However, this effect is smaller than the difference in In-Group Homogeneity and could theoretically be an artefact of the analysis.

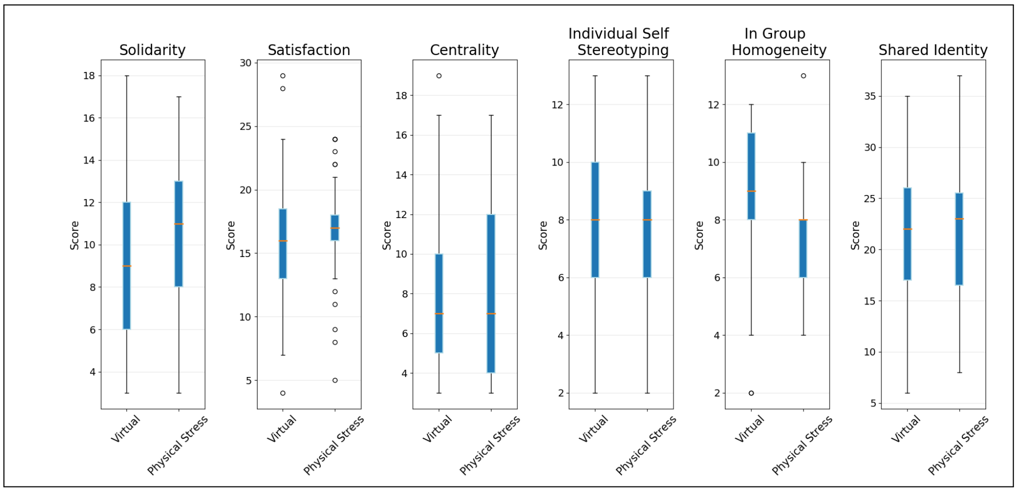


Figure 6.10.: Differences in social responses between PR Stress and VR participants

Measure	PR Stress (mean, SD, range)	VR (mean, SD, range)	Test statistic (p-value)
Solidarity	10.3 (3.55, 3-17)	9.18 (3.94, 3-18)	MW 1348.5 (0.0338)*
Centrality	7.96 (4.10, 3-17)	7.87 (4.33, 3-19)	MW 1614.0 (0.3630)
Satisfaction	16.8 (3.54, 5-24)	16.1 (4.76, 4-29)	MW 1386.5 (0.0538)†
Individual self-stereotyping	7.53 (2.67, 2-13)	7.62 (2.78, 2-13)	-0.1814 (0.8564)
In-group homogeneity	7.37 (1.86, 4-13)	8.89 (2.43, 2-12)	**-3.727 (0.0003)*** **
Shared identity	21.5 (6.63, 8-37)	21.4 (6.92, 6-35)	0.0434 (0.9655)

Table 6.13.: Differences in social response measures between PR and VR groups

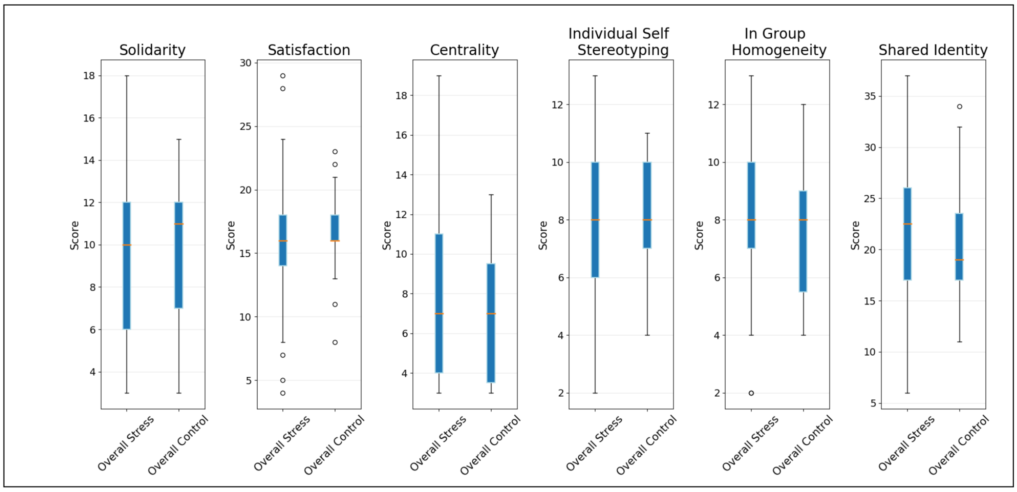


Figure 6.11.: Differences in social responses between Overall Stress and Control participants

Measure	**Overall Stress (mean, SD, range)**	**Overall Control (mean, SD, range)**	**Test statistic (p-value)**
Solidarity	9.76 (3.77, 3-18)	9.68 (3.64, 3-15)	MW 1088.0 (0.4658)
Centrality	7.92 (4.19, 3-19)	6.89 (3.36, 3-13)	MW 962.5 (0.1880)
Satisfaction	16.5 (4.16, 4-29)	16.6 (3.76, 8-23)	MW 1038.5 (0.3441)
Individual self-stereotyping	7.57 (2.71, 2-13)	8.05 (2.30, 4-11)	-0.8247 (0.4167)
In-group homogeneity	8.11 (2.27, 2-13)	7.58 (2.43, 4-12)	0.8806 (0.3874)
Shared identity	21.5 (6.74, 6-37)	20 (6.08, 11-34)	0.9564 (0.3477)

Table 6.14.: Differences in social response measures between Overall stress and control groups

Variable	**Shared identity**	**In-group homogeneity**	**Individual self-stereotyping**	**Centrality**	**Satisfaction**	**Solidarity**
Constant	19.358**	9.436***	3.972	7.366†	14.488**	2.170
Personality A	0.470	0.060	0.012	0.700	-1.218	-0.227
Personality C	0.380	-0.023	-0.483	-0.566	1.397*	1.545*
Personality E	1.270	0.193	0.098	-0.445	0.621	0.192
Personality N	0.910	0.427	0.572	-0.650	0.365	0.012
Personality O	-1.053	-0.524	-0.035	-0.774	0.308	-0.431
Age	0.013	-0.103†	0.161*	0.147	-0.062	0.192*
F	-1.325	0.038	-0.551	-0.422	0.926	0.340
Exercise level	-0.596	0.255	-0.045	0.345	-0.022	0.092
Adjusted r^2	-0.03033	0.022518	0.036249	-0.01337	-0.00521	0.0439
F-statistic (p-value)	0.591566 (0.782869)	1.31964 (0.242068)	1.52187 (0.158629)	0.815347 (0.59068)	0.92748 (0.497176)	1.6432 (0.12149)
Omnibus test (p-value)	2.657498 (0.26481)	2.1081 (0.34852)	0.76987 (0.68049)	6.31105 (0.04262)	5.05656 (0.0798)	0.6327 (0.72882)

Table 6.15.: Overall Stress social responses and demographic regression models

Variable	Shared identity	In-group homogeneity	Individual self-stereotyping	Centrality	Satisfaction	Solidarity
Constant	18.386***	8.190***	6.999***	5.2976*	17.510***	7.318**
Delta-STAI	0.019	0.010	-0.006	-0.0472	-0.039	-0.063†
PANAS-P	0.134	0.013	-0.006	0.0157	0.008	0.062
PANAS-N	-0.056	-0.023	0.043	0.1274†	-0.052	0.043
Adjusted r^2	-0.0101	-0.0231	-0.0191	0.0055	-0.0009	0.0170
F-statistics (p-value)	0.6419 (0.5897)	0.1811 (0.9090)	0.32631 (0.8063)	1.19916 (0.31378)	0.96847 (0.4105)	1.6289 (0.1871)
Omnibus test (p-value)	1.47759 (0.4777)	2.87955 (0.23698)	3.19674 (0.2022)	9.22297 (0.0099**)	3.96023 (0.13805)	3.85140 (0.14577)

Table 6.16.: Overall Stress social responses and self-assessed emotional regression models

6.4. Social responses- inferential statistics

As can be seen in Table 6.16, only marginal correlations were found between group identification or shared identity and emotional scores. This implies that, for the experiments performed, the level of group identification was independent of the level of self-assessed emotional state.

- Delta-STAI and Solidarity (marginal significance)
- PANAS-N and Centrality (marginal significance)

6.4.1. Social responses and demographic

Some correlations were observed between (shown in Table 6.15) social responses and demographic. These were:

- In-group homogeneity: Age (marginal negative)
- Individual Self-Stereotyping: Age (positive)
- Solidarity: Age (positive) and Conscientiousness (positive)
- Satisfaction: Conscientiousness (positive)

However, the level of significance of these results is low enough (i.e.: 5% to 10% significance) that there is doubt as to whether a real effect has been identified, or whether it is artefactual. This will be discussed further below.

6.5. Discussion and conclusion

Previous pedestrian dynamics approaches have claimed that it is not possible to examine hostile emergencies (see [3]), since it is necessary to provide appropriate stressors in order to perform experiments in an ecologically valid manner. Therefore, when designing the methodology for this experiment, it was necessary to both induce these emotional responses to the highest ethically-permitted degree, and then accurately measure them.

This chapter has pursued two separate approaches in investigating these responses, namely the participant self-assessed emotional states, and social responses. The key results from this aspect of the experimental approaches are detailed below, and tabulated in Table 6.17.

Finding	Impact
Participants were psychologically stressed.	It is possible to create psychologically stressful pedestrian dynamics experiments with an unannounced hostile actor.
No difference in psychological stress between VR and PR experiment.	Future VR experiments will be able to produce the same level of psychological stress as would occur in PR experiments, reducing the logistical and financial costs of such research, as well as any risks and ethical considerations.
Minimal correlations between emotions or group identification, and demographics.	The emotional and social responses were only slightly dependent on participant demographic.
Evidence for independent emotions model.	Provides insight into the different emotional states of participants, and presents some novel results for psychological research.

Table 6.17.: Key Internal subjective results

Participants were psychologically stressed

This chapter provides strong evidence for the internal validation of these experiments, showing that it is possible to perform ecologically valid experiments surrounding human behaviour in emergencies. The levels of reported self-assessed emotion indicate that the experimental procedure developed by this PhD, where a hostile aggressor represents a financial threat to the participants, increases the levels of negative emotion. This is highlighted in Table 6.5, where it can be seen that the Overall Stress condition is significantly different to the Overall Control condition for both measures of negative emotion, but has no discernible difference for the measure of positive emotion (PANAS-P). This difference in emotional state provides ecological validity for the experiment, suggesting that the responses shown are indicative of those that would occur during an actual hostile emergency.

This experiment methodology is ethically viable, having been assessed by the independent Imperial College Research Ethics Committee board for each experimental paradigm (ICREC, PR reference number 18IC4637, VR reference number 19IC5216). Previous studies (detailed in [3]) considered that examining human behaviour in stressful conditions such as terrorist attacks to be an intractable problem owing to the ethical implications, resulting in limited progress over the past decades.

The internal validity of these experiments will be assessed further in the next chapter (Results: Internal Objective data) which concludes that participants also had a physiological response to the stressor. When examining both chapters' findings, it is possible to conclude that these experiments produced the stress responses expected. As this methodology has now been shown to be viable and generalisable (see Chapter 3), it is possible to extend these experimental procedures to both obtain further data and examine different contexts. Future possible experimental approaches will be discussed further in Section 10.4.

Minimal reported differences between PR and VR experiments

When examining the Internal Subjective results of both the PR and VR participants, there were very few statistically significant differences identified in the psychological responses of participants between either paradigm. Differences in self-assessed emotional states were only detected for positive emotion, where VR participants self-assessed with a higher level of positive emotions (see Table 6.4). When considering the degree of social response, VR participants self-assessed with a higher level of In-group homogeneity between group members, as well as higher levels of Solidarity Table 6.13).

The results are considered a product of the environment itself, where:

- The impact on self-assessed positive emotion is considered a result of VR representing the cutting edge of gaming technology, and therefore an exciting experience for participants. This may change in future as VR technology becomes more widespread.
- The impact on In-group homogeneity is considered a result of the VR avatars being limited to two separate avatars (male and female, for examples see Section 4.5). This could change for future experiments with the development of a character selection and personalisation process.
- Finally, the difference in Solidarity is not considered a strong result, as it is only significant at the 5% level (p=0.0338). Given this relatively high p-value (with all other results recorded at 1% significance or lower) it is possible that this is an anomalous result.

Therefore, the secondary finding of this chapter is that there are minimal differences in emotional and group identification states between participants in either VR or PR experiments. This is an important result, as it provides some of the first data on the validity of using VR experiments to investigate human behaviour in emergencies.

However, this result is based on an experiment in which participants were informed that the observed avatars were controlled by other participants, whom they would meet later. This was performed to ensure that some degree of social consideration would be present within the environment and the author suggests that participant reactions would be significantly different without this effect. Future work could investigate any differences in behaviour between participants who are aware of the computer-controlled nature of avatars, and those who believe that other avatars are controlled by other participants.

Low but significant correlations between emotional state, social response and demographic

When examining the correlations between the emotional state, social response, and personality levels, several statistically significant relationships were identified. The relationships between self-assessed emotional state and demographic are tabulated in Table 6.9, and repeated below:

- Negative emotions:
    - PANAS-N: Neuroticism (strong negative), Extroversion (marginal positive)
    - Delta-STAI: Openness (negative), Neuroticism (marginal positive), Age (marginal negative)
- Positive emotions
    - PANAS-P: Neuroticism (positive)

These relationships indicate that Neuroticism is the strongest mediator of self-assessed emotional state. Somewhat confusingly, participants who scored highly in Neuroticism also scored higher on Delta-STAI, but lower on PANAS-N, both of which are the key indicators used to measure negative emotional states. These results seem to contradict each other, given the strongly significant positive correlation between PANAS-N and Delta-STAI (Table 6.8). These participants also self-assessed with higher PANAS-P scores. Further investigation is required to better understand these relationships.

Relationships between self-assessed social response and demographic are tabulated in Table 6.15, and repeated below:

- In-group homogeneity: Age (marginal negative)
- Individual Self-Stereotyping: Age (positive)
- Solidarity: Age (positive) and Conscientiousness (positive)
- Satisfaction: Conscientiousness (positive)

One implication of self-categorisation theory is that the degree of social response which a group experiences after having undergone a stressful scenario should be independent of personality types. However, as has been indicated in Table 6.10 and Table 6.14, this experiment did not generate any statistically significant differences in the levels of self-assessed group identification as a result of the experimental intervention. Therefore, any difference in group identification observed in these results is not expected to be a result of the experiment.

The regression models that were used to understand the impact of demographic on the emotional and social response scores showed a low level of explanatory power, shown in the adjusted r^2 values, which ranged from 5 to 7% for the emotional results (Table 6.9), and -2 to 4% for the social responses (Table 6.15).

Finally, since these experiments are the first to quantitatively measure pedestrian movement in response to a hostile aggressor, existing theories may not accommodate the observed responses. As a result, new theoretical constructs may be required to fully explain these datasets.

Evidence for the distinct emotional state model

As detailed in the Chapter 2, there are several different basic models of emotion. Two of the most well-known models are the discrete emotional model (DEM), and the two-dimensional valence-arousal model. The DEM suggests that there are defined, discrete emotions which are independent of each other, and can occur simultaneously, whereas the Valence-Arousal model suggests that all emotions can be characterised by a valence (positive or negative) and an arousal level (low or high).

The results obtained here for positive emotions (PANAS-P) and negative emotions (PANAS-N and Delta-STAI) are found to be independent, with high levels of self-assessed positive emotions occurring simultaneously with high levels of self-assessed negative emotions (Table 6.7). This provides evidence that the DEM is a more appropriate model than the valence-arousal model in characterising emotional state.

This result should be considered alongside the theory that emotional states are also characterised by behavioural and physiological information (see Section 2.3). While it is out of the scope of this thesis to fully investigate or conclude the research on emotions, the different physiological and behavioural responses of participants will be investigated further in the next chapters.

6.6. Chapter summary

This thesis reports results using standard significance levels (5% significance, or 10% significance for marginal results), while any results that are found at the 1% significance level or lower are referred to as 'strongly significant'. Using these levels, this chapter has analysed the subjective results obtained from participants in both the PR and VR experiments as part of this PhD. Specifically, it performed a quantification of the psychological stress that was a direct consequence of the introduction of an unannounced hostile actor, who represents the potential for financial loss, acting as a proxy for a knife-based terrorist. This quantification provides internal validation for the experimental procedures, ensuring that ecologically valid conditions were achieved.

The primary analysis in this chapter focused on the measures of emotional state, using previously validated surveys (the STAI and the PANAS). These measures showed a significant increase in negative emotion within the experimental groups, which was not seen in the control groups. There was no statistically significant difference between these emotional responses in PR and VR environments. The secondary analysis in this chapter showed the impact of the experiment on the degree of social responses within the participant groups. This analysis did not detect any significant impact on these levels, with the exception of increased perceived homogeneity within VR, which is considered to be a result of the avatars used in the environment.

The next chapter will continue this research by quantifying the physiological stress undergone by participants, showing further validation of these experimental procedures. Consequently, this PhD will show that these experiments generated realistic data on the movement dynamics of pedestrians in hostile emergencies, thus providing more complete and relevant datasets than earlier studies have achieved.

7. Results: Internal Objective

This chapter will discuss the objective measures of emotion that were detailed in Section 4.7, providing insight into the physiological responses of the participants to the hostile aggressor introduced during the experiments. These responses include endocrine measures (salivary cortisol) and autonomic nervous system (ANS) measures, such as heart rate, skin temperature and galvanic skin response (GSR).

The outline of this chapter is shown in Figure 7.1. This chapter will first discuss the analysis performed on the samples which measured levels of cortisol in the participant's saliva. This will include a comparison between PR and VR paradigms, as well as a further comparison with Control experiments. Next this chapter will detail the analysis performed on the ANS measures, including an assessment of the use of ANS data for classification of the stress state of participants using multiple different algorithms. Finally, this chapter will analyse the constituent frequencies of the ECG trace of individual participants wearing the chest-mounted EQ02 monitor, investigating the power spectral decomposition of heartbeats as a measure of stress.

A feature of the analysis described here is the wide variation in the physiological responses of individuals, which is a widely noted feature in previous research. Therefore, while trends are observed within grouped data, there is a large range in the measured values. This will be discussed at the end of the chapter.

Since most of the data analysed in this chapter was found to be non-normally distributed (see Section 5.3), all hypothesis tests and significance codes will be based on the Mann-Whitney U test, instead of Welch's t-test. As in previous chapters, the significance levels will be reported with the prefix "MW".

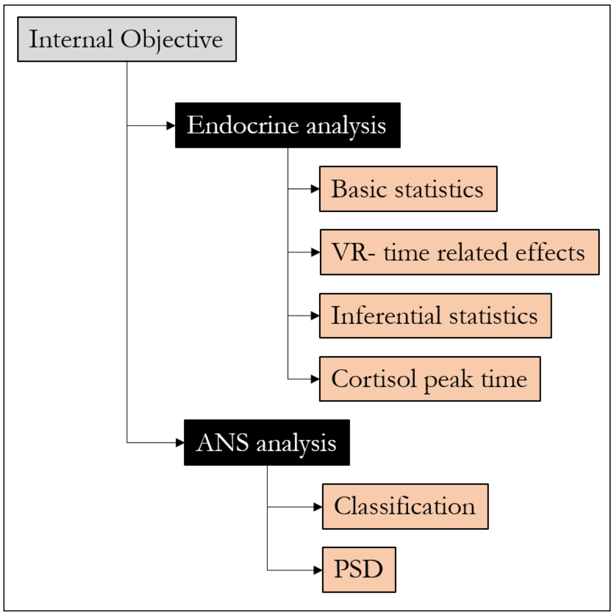


Figure 7.1.: Internal Objective analysis types

7.1. Data gathering

There were several issues regarding the data gathering, resulting in a limit to the number of datasets available for the following analysis. The primary limitation was the cost of the equipment, restricting the experiments to two chest-mounted Equivital EQ02 monitors and two wrist-based e4 Empatica monitors. There were secondary limitations including equipment failure and logistical issues, reducing the number of datasets that could be obtained. The full number of datasets is detailed in Chapter 5.

7.2. Endocrine analysis

Chapter 4 thesis explains how four separate saliva samples were obtained from all participants: a baseline value (pre-experiment), and samples taken 20-, 40- and 60-minutes after the hostile aggressor's arrival. These samples provided quantification of cortisol in the participants' saliva, which is commonly used as a metric for stress levels (see Section 2.3 for further information).

The measurements from these samples were then interpolated using a 1D cubic spline interpolation, providing a continuous measure of both hormones over time for each participant, as detailed above. This time series data was normalised relative to the initial (baseline) measure of cortisol, and four measures were then extracted for each participant. These measurements, shown graphically in Figure 7.2, include:

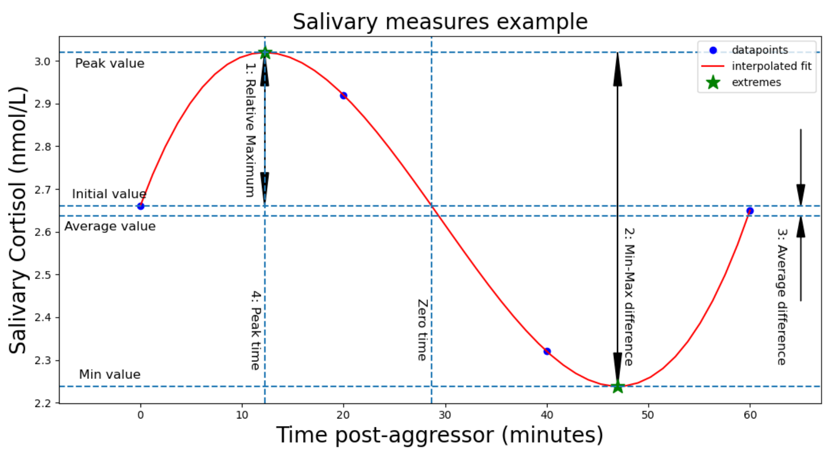


Figure 7.2.: Different salivary cortisol measures

No.	Measure	Description
1	Relative maximum	The value of the global maximum of the time series, relative to the baseline level.
2	Minimum-maximum difference	The difference between the global maximum and global minimum of the time series.
3	Average difference	The average difference between the baseline and the three subsequent measures.
4	Peak time	The time at which the global maximum occurs after the onset of the aggressor.

Table 7.1.: Different salivary cortisol measures

The measures obtained from these experiments are shown below graphically in Figure 7.3, and tabulated in Table 7.2. They indicate that the cortisol levels of participants within the PR paradigm have increased over the first 40-minutes post-aggressor entrance, while the Control participants showed a cortisol peak approximately 20-minutes post-experiment start (measured after experiment start as no aggressor was introduced). There are statistically significant differences between the PR-Stress and Control cortisol values at 40 minutes post-aggressor entrance, as well as the average relative cortisol value across all measurements, and further marginally significant differences

Mean values (nmol/L)	PR PROMPT	PR NO-PROMPT	MW p-value
20 minutes	1.288	0.816	0.2071
40 minutes	0.809	0.118	0.1062
60 minutes	-0.430	-0.310	0.3399
Average difference	0.416	0.156	0.1469
Peak relative value	2.000	1.101	0.0680†
Min-max difference	2.831	1.724	0.2477

Table 7.4.: Differences in cortisol production in PR branches

Mean values (nmol/L)	VR MOVE	VR NO-MOVE	MW p-value
20 minutes	0.152	-0.340	0.365
40 minutes	-1.132	-1.906	0.325
60 minutes	-1.949	-2.949	0.316
Average difference	-0.732	-1.299	0.427
Peak relative value	1.295	1.030	0.110
Min-max difference	3.541	4.238	0.188

Table 7.5.: Differences in cortisol production in VR branches

($p<0.1$) at the 20-minutes post-aggressor entrance and peak value metrics. This indicates that the addition of an aggressor had a significant effect of increasing cortisol production relative to the Control paradigm, as expected.

In contrast to this, the VR participants cortisol levels remained approximately constant for the first 20 minute period, before reducing steadily across for the remaining samples, with all measurements being on average lower than both PR and Control experiments. The VR results are significantly different to those of the PR experiment, with the exception of the relative maximum value. This is investigated in Figure 7.4.

The next analysis was to understand any potential differences between participants based on experimental branch. As can be see in Table 7.4 and Table 7.5, there are no differences seen between PROMPT and NO PROMPT branches, nor in MOVE and NO MOVE branches.

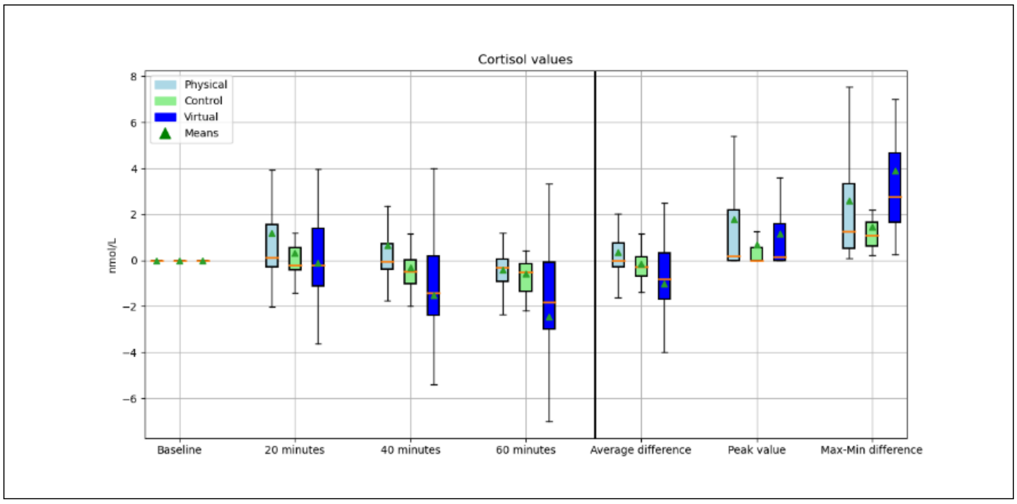


Figure 7.3.: Differences in cortisol production between paradigms

Mean values (nmol/L)	**PR Stress**	**Control**	**VR**
20 minutes	1.186	0.336	-0.098
40 minutes	0.659	-0.311	-1.526
60 minutes	-0.404	-0.566	-2.458
Average difference	0.360	-0.135	-1.021
Peak relative value	1.806	0.683	1.160
Min-Max difference	2.591	1.448	3.896

Table 7.2.: Differences in cortisol production between paradigms

MW p-values	**PR - Control**	**VR- Control**	**PR-VR**
20 minutes	0.0535†	0.2763	0.0168*
40 minutes	0.0252*	0.0313*	**4.338e-05*** **
60 minutes	0.1430	0.0091**	**6.200e-05*** **
Average difference	0.0300*	0.0505†	**0.0001*** **
Peak relative value	0.0672†	0.2192	0.1628
Min-Max difference	0.3049	**2.744e-05*** **	**8.507e-05*** **

Table 7.3.: Mann-Whitney U test p-values for cortisol differences between paradigms

7.2.1. VR results

It took approximately 1.5 hours to perform a single experiment and consequently performing all of the experiments at the same time as the PR experiments (i.e. 17:00 or later) could have exceeded the time available during the PhD, representing an unacceptable logistical limitation. Instead, the VR experiments were performed at the following specific times throughout the day: 09:15, 11:15, 13:15, 15:15, and 17:15, in January and February 2020, narrowly finishing before the first UK COVID-19 lockdown in March 2020.

It was expected that individual participants would produce differing absolute levels of salivary cortisol depending on several factors, including the time of day. It was noted in Section 2.3 that cortisol production is strongly dependent on an individual's circadian rhythm, starting high in the morning before reducing throughout the day. These experiments aimed to mitigate this effect by observing relative measures, rather than absolute measures.

However, it can be seen in Table 7.6 that the measured relative cortisol values are dependent on whether the experiment took place in the morning or the afternoon.

Further analysis that could be performed on this dataset might involve the inclusion of a model of a participant's individual circadian rhythm, in order to produce a relative cortisol measurement that is based on predicted, rather than absolute levels. However, this is beyond the scope of this thesis.

7.2.2. Relationships to other measures

The following tables show the results of regression models for the relative salivary measures, using several predictors. These predictors are based on the subjective emotional measures detailed in the previous chapter, and on the individual participant's demographic, measured by gender, exercise level and age. Participant personality was not used as a demographic predictor for these variables.

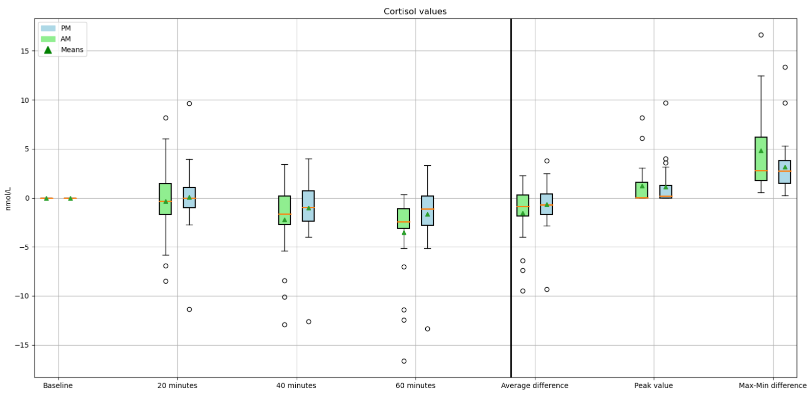


Figure 7.4.: Differences in cortisol production in VR by time of day

Mean values (nmol/L)	**VR-AM**	**VR-PM**	**MW p-value**
20 minutes	-0.333	0.084	0.323
40 minutes	-2.218	-0.991	0.152
60 minutes	-3.499	-1.652	0.044*
Average difference	-1.513	-0.640	0.203
Peak relative value	1.206	1.124	0.398
Min-Max difference	4.827	3.175	0.086†

Table 7.6.: Differences in cortisol production in VR, by time of day

PR Stress: Cortisol	**20 minutes**	**40 minutes**	**60 minutes**	**Average difference**	**Peak value**	**Min-Max difference**
Constant	6.2413	6.8856	-0.7904	3.0841	9.7942*	12.9869*
Delta-STAI	0.0234	-0.0122	0.0277	0.0097	-0.0127	-0.0321
PANAS-N	0.0001	0.0204	0.0095	0.0075	-0.0037	-0.0115
PANAS-P	-0.0432	-0.0433	-0.0310	-0.0294	-0.0428	-0.0297
Female	-1.8317*	-0.4728	-0.9432	-0.8119*	-0.9940	-0.6026
Exercise level	-0.1064	-0.4075	0.0514	-0.1156	-0.2947	-0.3840
Age	-0.1338	-0.1689	0.0305	-0.0680	-0.2236	-0.3104
Adjusted r^2	0.0378	-0.0511	-0.0208	-0.0034	-0.0328	-0.0303
F-Statistic (p-value)	1.3469 (0.2558)	0.5709 (0.7514)	0.8200 (0.5601)	0.9702 (0.4557)	0.7194 (0.6359)	0.7399 (0.6202)
Omnibus test (p-value)	17.8439 (0.0001***)	63.1405 (1.9463e-14***)	30.4538 (2.4380e-07***)	21.1019 (2.6168e-05***)	53.2403 (2.7480e-12***)	77.3365 (1.6091e-17***)

Table 7.7.: Relationships between PR Stress cortisol values and emotions and demographic

VR: Cortisol	20 minutes	40 minutes	60 minutes	Average difference	Peak value	Min-Max difference
Constant	-2.7883	-3.6243	-3.9424	-2.5888	0.4137	4.8557
Delta-STAI	0.1056*	0.0785	0.0490	0.0583	0.0686**	0.0266
PANAS-N	0.1602*	0.0887	0.0323	0.0703	0.0987*	0.0666
PANAS-P	-0.1137	-0.1007	-0.0764	-0.0727	-0.0579	-0.0002
Female	-0.0164	-0.5511	-1.0538	-0.4053	1.1657*	2.7501*
Exercise level	-0.7054	-0.4117	-0.1416	-0.3147	-0.4972	-0.3498
Age	0.0690	0.1072	0.1148	0.0727	-0.0200	-0.1397
Adjusted r^2	0.1158	0.0023	-0.0530	0.0058	0.2099	0.0949
F-Statistic (p-value)	2.1785 (0.0614†)	1.0206 (0.4235)	0.5472 (0.7696)	1.0527 (0.4039)	3.3904 (0.0072**)	1.9440 (0.0927†)
Omnibus test (p-value)	5.4736 (0.0648†)	16.0217 (0.0003***)	24.8587 (3.9994e-06***)	18.3488 (0.0001***)	15.3077 (0.0005***)	17.9209 (0.0001***)

Table 7.8.: Relationships between VR cortisol values and emotions and demographic

It is possible to see from Table 7.7 and Table 7.8 that the salivary levels of cortisol are dependent on many factors. These included emotional state, where higher levels of negative emotions (measured by the Delta-STAI and PANAS-N) resulted in higher differences in cortisol levels in the VR experiments, and gender, where female participants had lower differences in cortisol levels in the PR-Stress experiments and larger peak values and Min-Max differences in the VR experiments. This gender-based difference has also been seen in previous research [270].

However, it is also noted that there is wide variation in the datasets; as a result the adjusted r^2 values are relatively low, meaning these models have a low level of explanatory power. Further to this, the omnibus levels are large enough that the residuals of the dataset can be considered non-normal. This suggests that linear regression may be an inefficient modelling method, and instead non-linear regression should be used.

7.2.3. Cortisol peak

As a final metric, this PhD observed the average peak time for salivary cortisol. Previous data [116, 271] suggest that this peak should be anywhere between 10 and 30 minutes after the arrival of the aggressor. The data obtained during this PhD (see Table 7.9) suggests that cortisol levels arising from the PR experiments peaked at approximately 15 minutes post-stressor arrival, although there is wide variation in this value.

7.3. ANS analysis

The next set of physiological measures were Autonomic Nervous System responses, examining heart rate (HR), galvanic skin response (GSR), skin temperature (ST), respiration rate (RR). Further to this, for some participants a 2-channel ECG was obtained, providing heart rate variability (HRV) measures.

There were two separate devices for measuring these responses: the EQ02, and the e4 Empatica. In each PR experiment, the participants who were providing ANS measures were equipped with one of each device. However, logistical and technical limitations restricted the number of datasets that were obtained. This is fully detailed in Section 5.1.

These were analysed using the open-source Python package, BioSPPy [272],

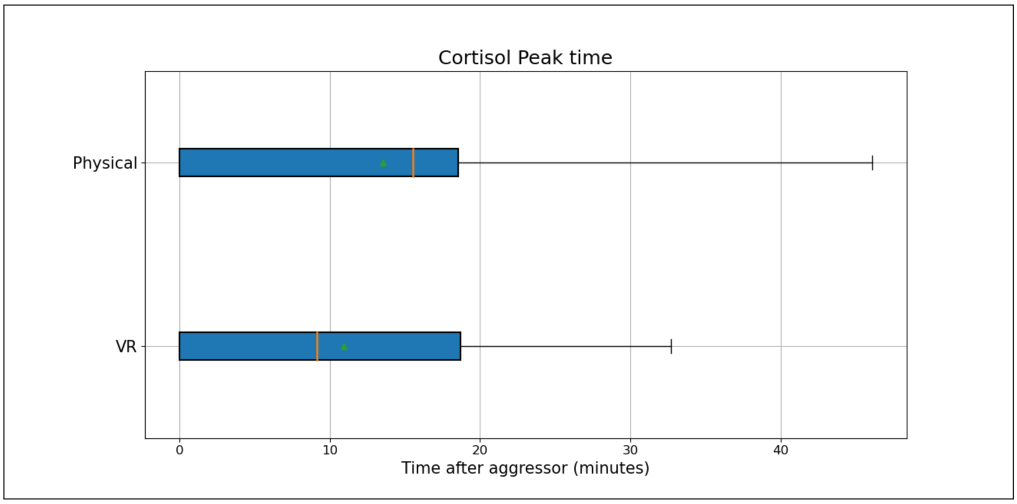


Figure 7.5.: Peak salivary cortisol time after aggressor

Paradigm	**Mean peak time (minutes)**	**Median peak time (minutes)**
PR	13.552506	15.543091
VR	10.931399	9.140259

Table 7.9.: Peak salivary cortisol time after aggressor

providing automated peak detection and standard analysis methods such as power spectral decomposition. Further analysis was performed using the scikit-learn package[267].

The data for each individual dataset was separated into stress and non-stress periods. These were defined by the start of the experiment, the aggressor's entrance to the experiment, and the end of experiment, when the participant was touched by the aggressor.

Several example comparisons of the data obtained by this equipment are shown below. In all figures, the vertical red lines indicate the point at which the aggressor enters the environment, and the respective measurements are colour coded: acceleration in green, HR in blue, GSR in black, ST in purple and RR in orange.

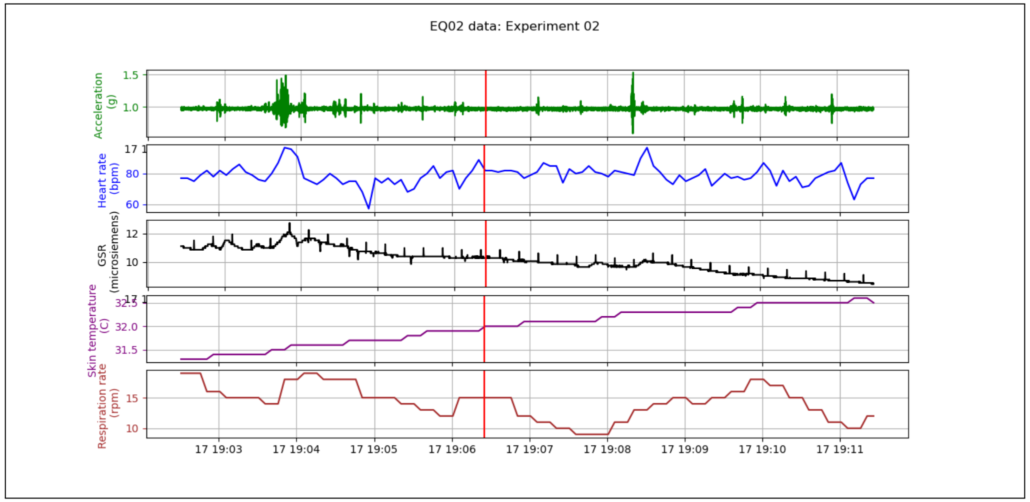


Figure 7.6.: Example EQ02 dataset from Aggressor Control experiment

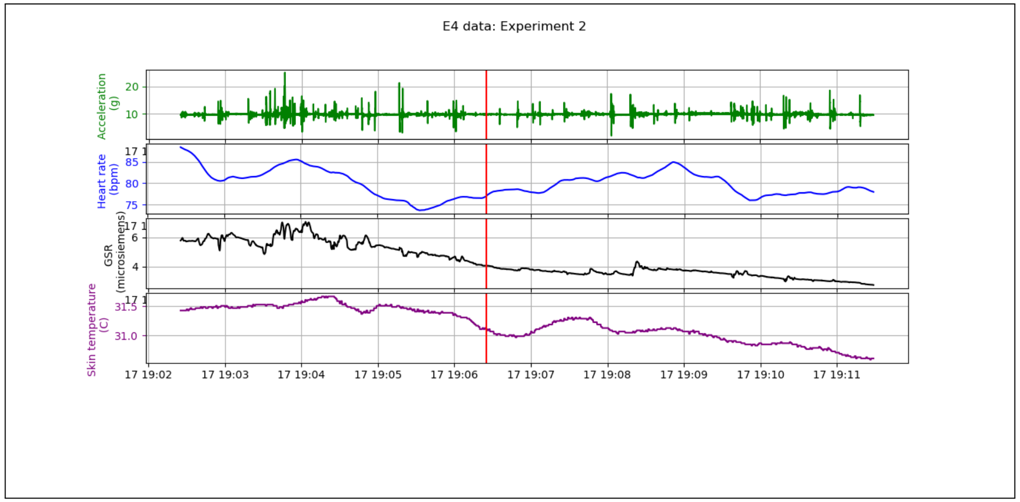


Figure 7.7.: Example e4 dataset from Aggressor Control experiment

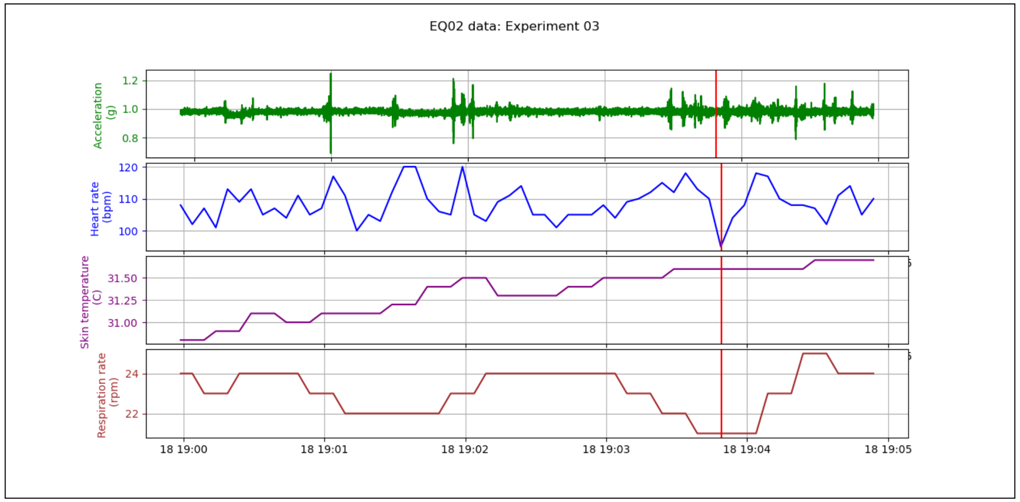


Figure 7.8.: Example EQ02 dataset from PR NO PROMPT experiment

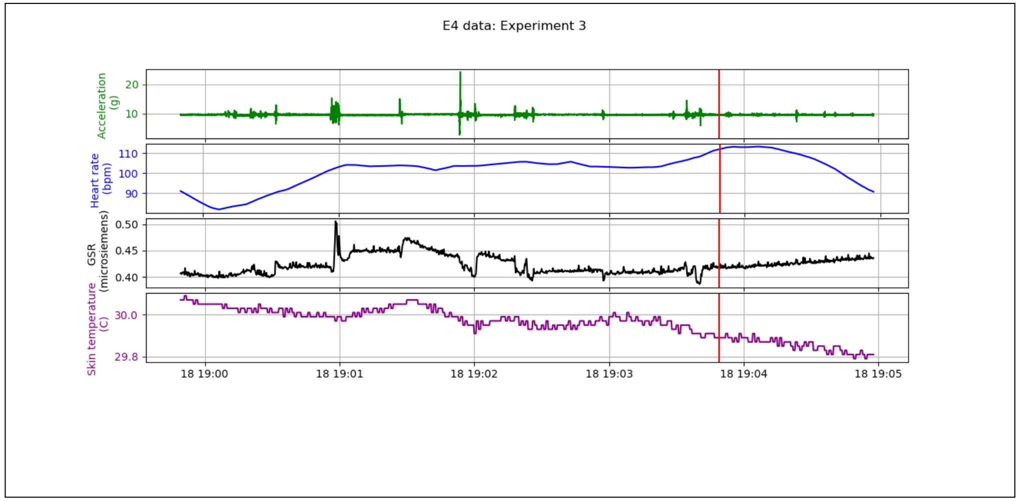


Figure 7.9.: Example e4 dataset from PR NO PROMPT experiment

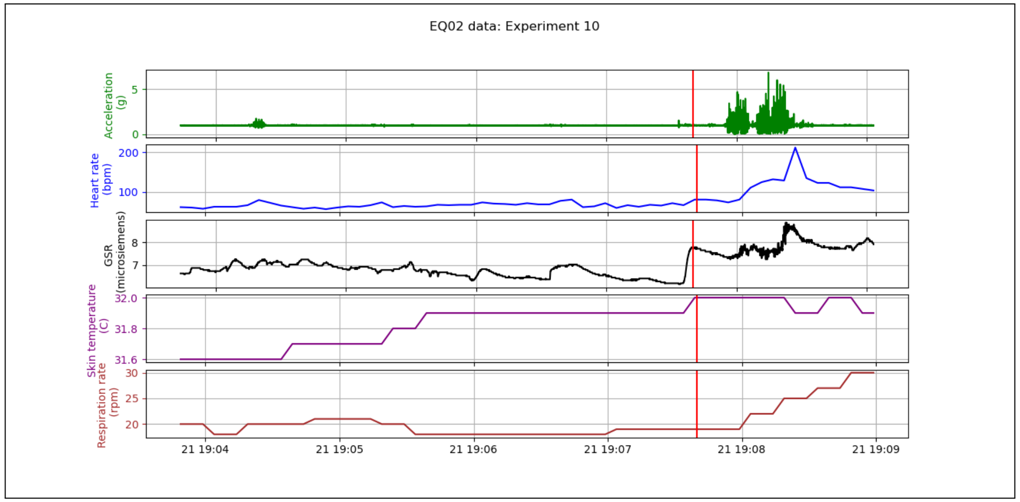


Figure 7.10.: Example EQ02 dataset from PR PROMPT experiment

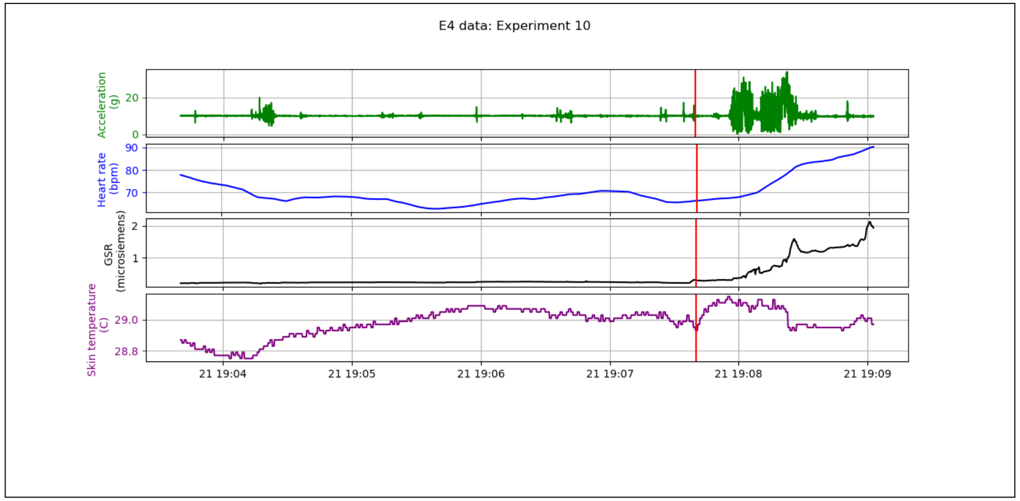


Figure 7.11.: Example e4 dataset from PR PROMPT experiment

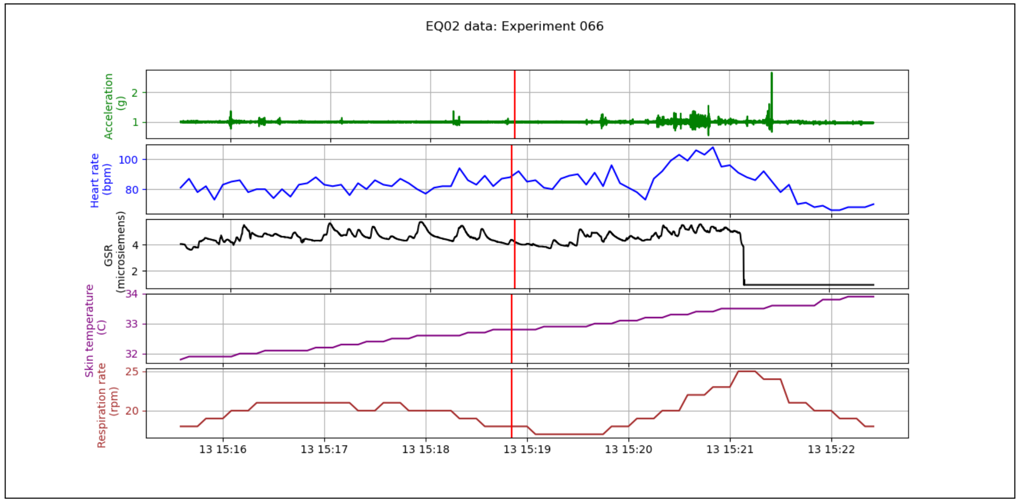


Figure 7.12.: Example EQ02 dataset from VR experiment

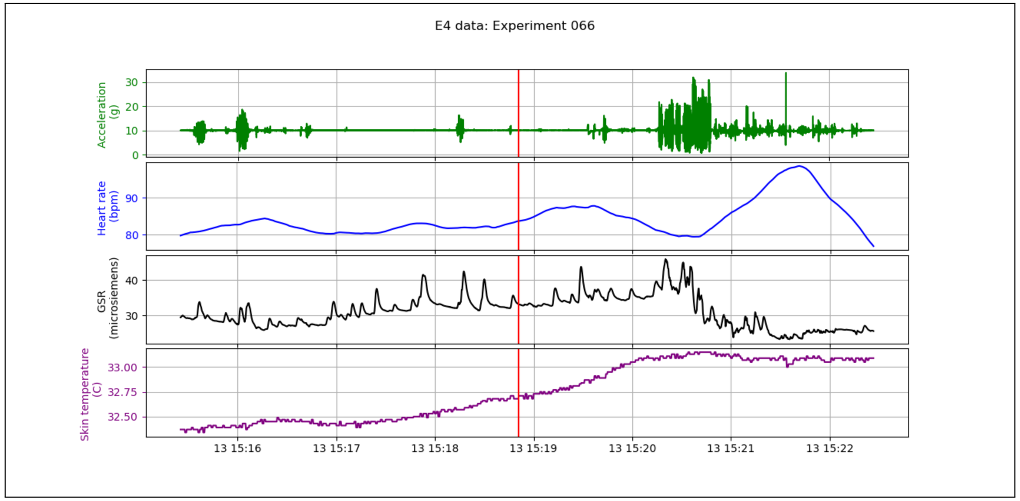


Figure 7.13.: Example e4 dataset from VR experiment

The initial comparison (Figure 7.6 and Figure 7.7) shows the ANS responses of one participant in the Aggressor Control experiment, in which the aggressor entered the environment but did not act in a hostile manner. The measures from the EQ02 and e4 showed similar results for the participant's heart rate (which stayed approximately constant at 80bpm), and for the participant's GSR (which reduced steadily throughout the experiment). Further parallels can be seen in the acceleration measure, where spikes occur at similar times, but the EQ02 shows a lower level of acceleration than the e4. This is expected since participants moved their arms more than their torsos.

In contrast to this, there is a difference over the skin temperature levels, which are shown to be rising for the EQ02, and reducing for the e4. The primary explanation is considered to be the different placement of the equipment around the body. The EQ02 is a chest-mounted monitor, with a GSR add-on which is placed on the fingers, while the e4 Empatica is a wrist-based monitor. This difference in placement can explain the discrepancy in skin temperature, as this is dependent on location, while heart rate is independent of placement, and the GSR measurements were approximately in the same location.

The next two figures (Figure 7.8 and Figure 7.9) show physiological responses from participants in Experiment 3, a NO-PROMPT PR experiment. The EQ02 GSR attachment failed in both cases, leading to fewer points of comparison, but it can still be seen that there is significant overlap in the different measurements. Neither participant moved post-aggressor entrance (indicated by the lack of an acceleration signal after the red line), and the e4 measurements show that there is a sustained increase in GSR around the time of the aggressor entry, while heart rate and respiration rate remain approximately constant. Again, it is possible to see a reversal in skin temperature responses, with the EQ02 indicating an increase in torso skin temperature across the experiment, while the e4 indicates a reduction in wrist skin temperature.

The final PR comparison (shown in Figure 7.10 and Figure 7.11) details the measurements taken in Experiment 10, a PR-PROMPT experiment, in which the participant did move after the aggressor entered. In this instance sustained increases in heart rate, GSR and respiration rate are observed, while again there is a difference between skin temperature measurements. There is an increase in e4 skin temperature, followed by a decrease, while the EQ02 torso skin temperature increases and remains high.

The next two figures (Figure 7.12 and Figure 7.13) show a comparison between the EQ02 and e4 measurements from a VR participant. The EQ02 traces show a much reduced level of acceleration, while the e4 shows a level that is comparable to the PR experiments. Increases are seen in heart rate and skin temperature across both systems, while the EQ02 shows respiration rate peaking after the aggressor enters the environment. There is also overlap in the GSR measurements between the two systems, with drastic drops in conductivity after the aggressor enters, but the discontinuous nature of the EQ02 measurements suggest equipment failure is the cause of its measurements, while the e4 appears to have functioned correctly.

7.3.1. Classification analysis

It is widely stated in the literature surrounding physiological assessment of emotional states that numerous measures of different physiological states are required before an acceptable level of accuracy is reached (see Section 2.3 for further information). Therefore, this analysis used all possible ANS responses as predictors.

This PhD assessed the relative merits of physiological signals by checking whether they be used to determine between participant states pre- and post-aggressor entrance. To do this, the signal datapoints that occurred before the entrance of the aggressor were labelled “0”, and those that occurred after the entrance of the aggressor were labelled “1”. This resulted in a binary classification system and does not provide any information on the relative level of different emotions or the type of emotional response, but rather whether the model predicts that the participant is currently undergoing a stressful response.

As this approach does not measure specific emotions, it can more accurately be described as measuring the arousal level of participants, and as such, this approach ascertains the benefits of these algorithms for assessing stressful conditions. However, the previous chapter (see Section 6.5) saw evidence in favour of a discrete emotional model. Therefore, any future research aiming to characterise emotional models should consider using a training period where different emotional states are induced and measured outside of the experimental procedure. These different emotional state measurements can then be used to train a model for assessing the emotional states induced by the experiment. This approach will increase the logistical requirements for an experiment, and it is suggested VR is a more appropriate paradigm, given the potential overlap between physiological responses to physical movement and emotional states.

The datasets obtained from individual experiments were combined, producing a single dataset for the EQ02 and another for the e4 Empatica. In order to combine these datasets, the results from individual experiments were subjected to the same standardisation procedure detailed in Equation 5.7, ensuring a zero-mean value and unit standard deviation. The resulting datasets were then analysed using the following machine-learning classification algorithms:

- Radial SVM
- Linear SVM
- K-Nearest Neighbours
- Random Forest
- Decision Tree
- Naïve Bayes
- Logistic Regression

These are commonly used algorithms, and were implemented using scikit-learn, a Python library that is dedicated to data analysis and which includes easily implementable machine-learning classification and regression algorithms. More information on these algorithms can be found in [273], and further justification for the use of machine learning techniques can be found in [136].

The models aimed to differentiate between pre- and post-aggressor states, as a proxy for emotional state classification. The results of these models were assessed using k-folds cross validation, with five folds. The results of this classification are shown below in Figure 7.14 and Figure 7.15 and are tabulated in Table 7.10.

To summarise this analysis, Table 7.10 shows that data from the PR experiments produced higher accuracy classification models than those generated by VR data. The PR experiments produced models with a peak of 86.4% accuracy, while the models trained on VR data produced a peak of 83.2%.

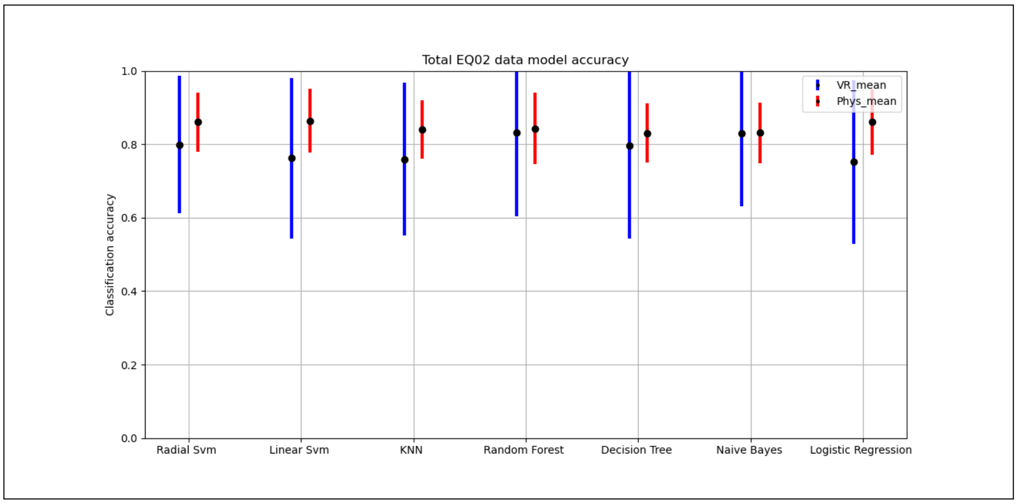


Figure 7.14.: Classification accuracy scores from EQ02 datasets

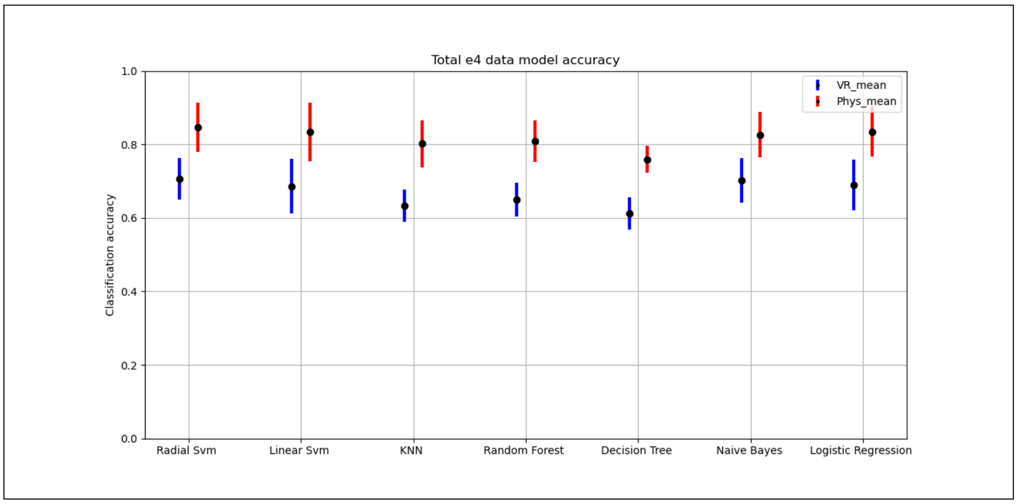


Figure 7.15.: Classification accuracy scores from e4 dataset

When split by measurement device the PR data produced models that performed between 5.7% points (using the EQ02) and 14.8% points (using the e4 Empatica) better than those generated by VR data. This was expected, as the PR experiments included physical movement, and consequently the physiological responses within these datasets were exaggerated. In contrast to this, the VR experiments only utilised VR movement, and consequently the physiological responses are assumed to be primarily part of the emotional responses. The accuracy of these classification algorithms are of a similar level to those detailed in Section 2.3.

Model	PR Stress		VR	
	EQ02	**e4**	**EQ02**	**e4**
Radial SVM	0.860	**0.847**	0.799	**0.706**
Linear SVM	**0.864**	0.834	0.762	0.685
KNN	0.840	0.802	0.759	0.633
Random Forest	0.843	0.809	**0.832**	0.649
Decision Tree	0.830	0.759	0.796	0.612
Naïve Bayes	0.831	0.826	0.829	0.702
Logistic regression	0.860	0.835	0.751	0.690

Table 7.10.: Classification accuracy scores for physiological data

It can also be concluded that data from the chest-mounted EQ02 produced higher accuracy classification models than those generated by the e4 Empatica data, with an average of 3.1% points higher accuracy in the PR models, and 12.2% points higher accuracy in the VR models. Across all algorithms and paradigms, the EQ02 outperformed the e4 by an average of 7.6% points. This is an indication of the benefits of chest-mounted measurement devices. However it should also be noted that these devices incur higher logistical requirements, including the time and private space required to put on the devices.

Finally, these results indicate that SVM-class algorithms, both the linear and radial models, performed better than the other algorithms in this classification. The radial SVM worked best for the e4 Empatica trained models (84.7% for the PR paradigm, and 70.6% for the VR paradigm), and the linear SVM worked best for the PR EQ02 trained model (86.4% accuracy). However, while best practice was followed in the implementation of these models, no optimisation procedure was performed (such as varying the number of folds in the k-folds cross validation process), and therefore it is possible that other models could perform better.

7.4. ECG data and Heart rate variability

The next analysis investigated the benefit of ECG datasets in contrast to other measures of heart rate (for example, the BVP measures provided by the e4 Empatica). ECG signals produce far higher more data than other sources of heart rate (e.g.: the blood volume pulsimetry used in the e4 Empatica), providing the ability to perform more detailed analyses, including a full power spectral decomposition (PSD). As detailed in [274], analysing heart rates in the frequency domain can provide useful insight into the emotional state of the participant, independently of the physiological response to physical movement.

Example analysis methods are shown graphically below, created using the BioSPPy and pyHRV packages. The left-hand side of Figure 7.16 shows the steps from raw ECG data to the filtered data and the peak detection method, before showing the resulting calculated heart rate. The right-hand side of this figure ('Templates') shows the individual heart beats, isolated by the peak detection method.

Using these templates, a PSD can be performed, isolating the relative importance of different frequencies within the ECG signal. Figure 7.17, below, shows the spectral decomposition for the entirety of the data shown in Figure 7.16. The PSD splits the frequencies into three separate categories, using standard values for Very Low Frequency (VLF, 0.00-0.04Hz), Low Frequency (LF, 0.04-0.15Hz) and High Frequency (HF, 0.15-0.40Hz) categories.

Finally, Figure 7.18 shows a waterfall plot of graphs showing the PSD of entire datasets from a single experiment. This figure shows the PSD of the dataset split into 30-second periods and highlight the entrance of the aggressor using a red horizontal line along the frequency axis.

While the waterfall plots shows the spectral decomposition for 30-second periods during the experiment, [275] states that in order to perform spectral decomposition of ECG data for short experiments, at least two-minute periods should be used. Therefore, as with the ANS measurements, the ECG datasets from all experiments were separated into pre- and post-aggressor entrance in an identical method to the ANS analysis above, and a power spectral decomposition (PSD) performed on the total data in both periods. This represents the maximal amount of data available for each separate section of the experiment.

In order to analyse this data the "LF norm" was used, which is defined as:

$$\frac{LF}{HF} = \frac{P_{LF}}{P_{HF}} \tag{7.1}$$

Here P_{HF} is the absolute power in the high frequency band of the PSD and P_{LF} is the absolute power in the low frequency band. The LF norm is a dimensionless measure that has been proposed as a strong candidate value for assessing the degree of physiological stress of a participant [274]. This

Paradigm	Mean delta value
PR Stress	-0.097
VR	3.148
Control	-0.188

Table 7.11.: Mean PSD LF norm pre- and post-aggressor differences

Compared paradigms	PR - Control	VR- Control	PR-VR
MW p-values	0.331	0.270	0.330

Table 7.12.: Mann-Whitney U test p-values for PSD LF norm pre- and post-aggressor differences

measurement quantifies the proportion of total power in the LF region of the PSD, and is correlated with emotional stress [275] where a higher proportion is associated with higher stress states.

The LF norm was quantified for each section of data, and the difference between these values was then compared. The average differences are shown below, tabulated for Control, PR and VR experiments.

As can be seen, these results are inconclusive regarding the impact of the aggressor on the PSD of the ECG data, and no significant differences are observed between any of the paradigms. Consequently, it is uncertain how powerful this measurement can be when determining the emotional state of the participants. Several explanations of this are proposed below:

1) Participants were stressed at the beginning of the experiment.

It is possible that the participants began the experiments in a state of stress equivalent to that experienced as a result of the aggressor, and that the addition of this aggressor thus had no impact on their emotional state. This is unlikely, as all other metrics indicate that the addition of the aggressor was a stressor during these experiments.

2) LF norm is not affected, or is complexly affected, by emotional state.

It is also possible that either the LF norm metric is not affected by emotional state, or it is affected by emotional state in a way that is not fully understood. In this instance, it is difficult to predict which responses would be expected as the result of a stressor. This explanation is supported by other research:

> *"The expanding literature on ultra-short-term, short-term, and 24 h HRV norms requires careful interpretation. Due to the lack of standardization [...] clinicians should not use ultra-short-term interchangeably with 5 min and 24 h values." [275] Page 13.*

3) More data is required.

As there were limitations to the number of datasets obtained, and given the variability seen in previous physiological measures (e.g. cortisol), this is the most likely explanation to the inconclusive results seen above. Any further research looking to investigate HRV could perform experiments that obtain data using the methodology detailed in this thesis.

7.5. Discussion and conclusion

This chapter has pursued three separate approaches in quantifying the physiological stress undergone by the participants, namely: salivary measures, ANS measures, and heart rate variability measures. These will be summarised and discussed below.

<u>Salivary measures</u>

The initial section of this chapter investigated the cortisol produced by participants across all experimental paradigms. The analysis of these results showed that the introduction of the aggressor correlated with higher cortisol production in the PR experiment, and that the peak of this cortisol production occurred 10-20 minutes after the stressor, which coincides with previous research in this area.

However, during the investigation of the VR participants' saliva samples, it was seen that their results do not mirror those of the PR stress conditions. Investigation into this effect showed that the measured changes in cortisol level are significantly dependent on the time of day at which the experiment took place. While this was expected, owing to logistical constraints it was not possible to perform all VR experiments at the same time that the PR experiments had been performed. Furthermore, the results indicate that the afternoon VR participants' cortisol level peaked, similar to the PR experiments but with a smaller effect size.

Overall, it is concluded that salivary cortisol is a sufficient differentiator between PR Stress and Control conditions, but that the effect size is not particularly large. It is also noted that the results from the VR participants are inconclusive. Therefore, it is suggested that future work consider alternative methods to assess the level of reaction to a stress-inducing stimulus.

ANS measurements

In contrast to the salivary measures, the ANS measurements represent a far denser dataset, with multiple high-frequency measures across various different established metrics, including HR, ST, GSR and RR. Using these datasets and state-of-the-art classification algorithms , this PhD established that there are distinct differences between the stressful and non-stressful periods within the experiments.

Table 7.10 shows that that the models based on PR data provide a better accuracy level than those based on VR data. This was expected, as the physical movement of participants in response to the aggressor in the PR experiments led to increases in ANS responses (such as heart rate and skin temperature) that would not have been present in the VR experiments. Therefore the PR experiments consequently provided data that was easier to differentiate.

It was also observed that data from the chest-mounted EQ02 monitor produced higher accuracy models than the wrist-based e4 Empatica. This could be due to several factors, including:

1) EQ02 uses a full two-channel ECG trace to measure heart rate, whereas the e4 Empatica uses BVP, which can be less accurate.

2) EQ02 measures respiration rate, providing another predictor variable.

3) EQ02 takes higher frequency measurements, providing larger datasets for classification.

However, it should also be noted that the EQ02 represents a higher logistical requirement for experiments, and the e4 monitors still produced high accuracy results. Therefore for the type of experiments considered in this thesis, the wrist-mounted e4 monitor is more appropriate.

Finally, it is possible to infer that the SVM-based algorithms produce the best classification models; however, further work should investigate the benefits of more detailed analysis, including optimisation of the classification algorithms, to fully assess the respective benefits of other algorithms.

Heart rate variability

The final set of analysis was to investigate the PSD of individual ECG traces. This quantified the change in heartbeat shape from pre-to post-aggressor entrance, utilising the LF-norm metric. This analysis suffered from a lack of data and produced inconclusive results.

7.5.1. Conclusion

This chapter has found three main conclusions on the use of physiological measures when investigating behavioural responses to hostile aggressors. The primary conclusions are:

- There are easily distinguishable stress and non-stress states makers based on ANS physiological responses.

Previous work has examined the use of ANS signals to understand emotional responses. This PhD utilises these approaches to create models that distinguish between the pre- and post- aggressor states. These models perform well, with up to 86.4% accuracy, and function well across both PR and VR paradigms.

These models do not quantify individual emotional responses, but rather predict binary states of stress and non-stress. Future work aiming to include emotional inference should incorporate training sessions where specific emotional responses are elicited (e.g. via video stimuli).

- Cortisol can be used as a measure of stress, and shows correlation with subjective results, but should be used with caution.

The PR experiments showed a significant peak in salivary cortisol approximately 10-20 minutes after the aggressor entered the environment, relative to the Control experiments. The VR experiments did not show this peak, however, these measurements were inconclusive to an extent, as logistical limitations meant that the VR experiments had to be performed throughout the day, while the PR experiments were only performed after 17:00. This difference in time caused significant differences in the cortisol responses of individual participants, leading to confounding effects within the analysis.

- Multiple emotional measures should be used

After examining the physiological data in conjunction with the subjective responses (detailed in Chapter 6 and Chapter 7), it is concluded that these experimental processes have produced the emotional reactions that are expected in hostile emergencies (e.g. fear, stress, etc.). However, given the questions raised about the nature of these reactions, including the theoretical constructs surrounding them (i.e. valence-arousal vs discrete emotions models), multiple measures should be used in future work. These should include both subjective and objective measures, with this chapter showing ANS measurements to be a better objective differentiator of emotion than salivary cortisol measures. Therefore, if there are financial or logistical limitations for any future work, this PhD recommends using psychological survey measures alongside ANS measurements, before then extending to include endocrine measurements and then potentially neurological measurements.

This result satisfies the Objectives 2 and 3, as stated in Chapter 1. This represents some of the first controlled experiments that produce these emotional responses while simultaneously gathering accurate data relevant to pedestrian dynamics.

7.6. Chapter summary

This chapter has investigated the physiological signals produced by the participants of the experiments. These signals objectively quantify the physiological stress undergone by these participants providing a secondary measure to the psychological stress detailed in the previous chapter. This chapter shows that the participants showed signs of physiological stress in numerous ways.

The initial analysis on salivary measures showed that PR participants experienced an increase in cortisol that was not experienced by participants in control experiments. The VR participants showed reductions in cortisol, beyond that seen in the Control participants, but this was concluded to be a product of the logistical limitations, since individual experiments were performed throughout the day, rather than at 17:00 or later.

The next analysis investigated the data obtained by the ANS monitors. Utilising the multiple sources of data from each monitor, the primary section of this analysis focused on identifying stressful periods, achieving an accuracy of up to 86%. The secondary section of this analysis investigated the benefit of using ECG traces to identify stressful periods, over simpler methods (such as BVP). This proved inconclusive.

There were several lessons learned from the implementation of these experiments, including the fact that, for group experiments, unless ECG traces are required, ease of use should be a primary deciding factor in the choice of ANS measurement system. For example, while the EQ02 performed better than the e4 Empatica, this came at an extra logistical cost. Wrist-mounted systems are easier to apply, and do not require participants to undress at any point, and can still result in beneficial data.

Building on the conclusions from this chapter and the previous chapter, it is possible to conclude that the participants were both psychologically and physiologically stressed. This stressful state provides an ecologically valid situation in which to assess their movement responses to hostile aggressors. The positional data showing these responses will be detailed in the next chapter, using both discrete and continuous movement analysis.

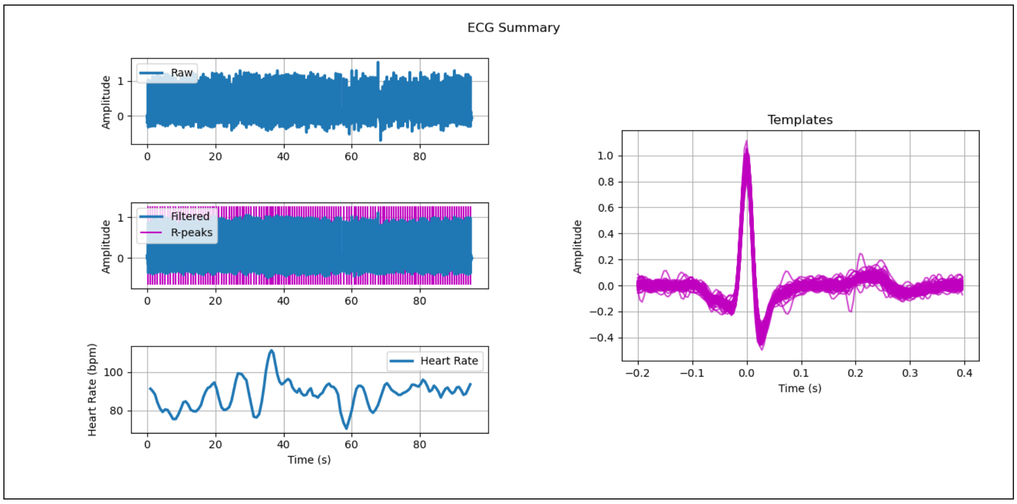


Figure 7.16.: Example ECG traces

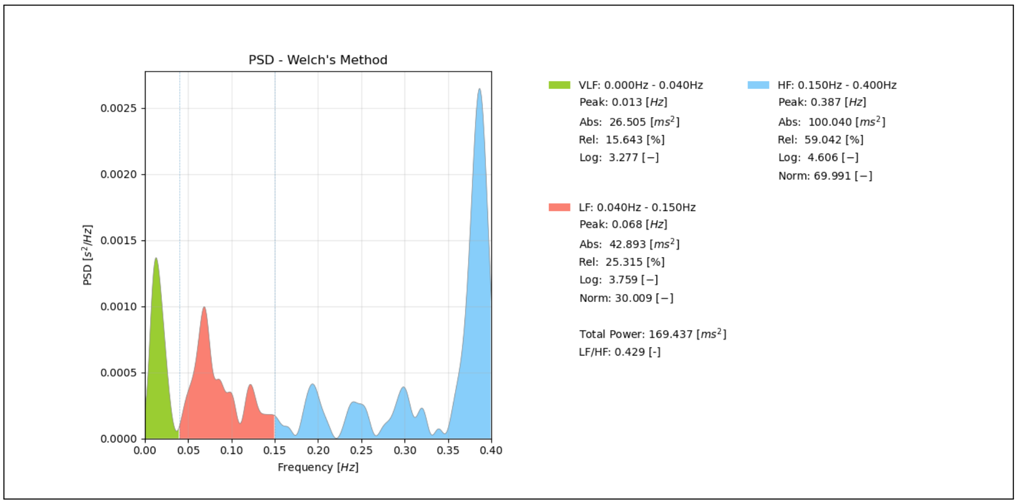


Figure 7.17.: Example PSD of above data

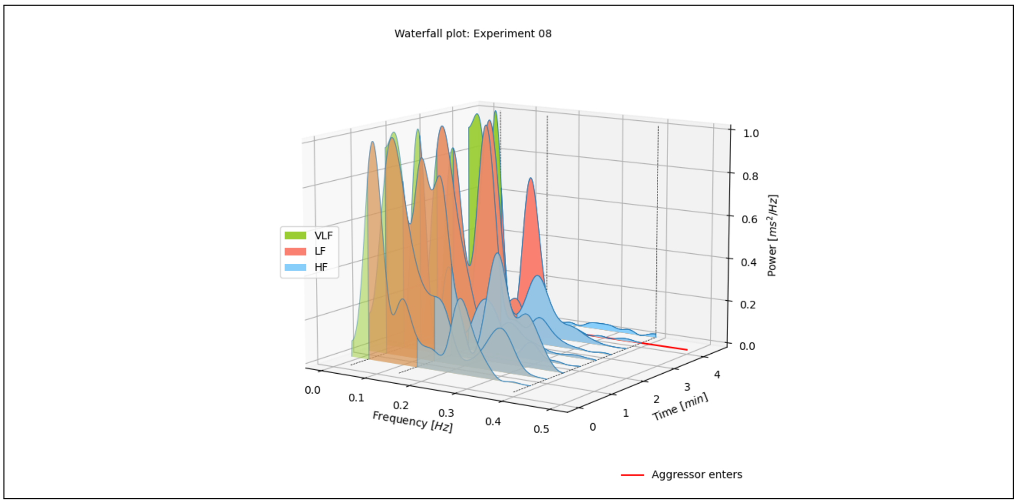


Figure 7.18.: Example waterfall plot of PSD for full experiment

8. Results: External

Following Chapter 4, which focused on the implementation of an experimental methodology to investigate human behaviour and movement in response to an unexpected hostile aggressor, this chapter will detail the analysis undertaken on the resulting data. Specifically, this analysis will describe and compare the results from the external measurements for the PR and VR experimental paradigms.

The initial analysis will describe any discrete movement behaviour, labelling the different observed activity choices. This discrete analysis then creates logistic models to predict whether a participant will move or not, depending on social observations, positional and demographic data. The logistic regression is performed on both PR and VR datasets, before a further regression model examines any observable differences within a combined dataset.

This analysis will be followed by an investigation into the continuous movement responses of participants confronted by a hostile aggressor. Specifically, this is a multiple regression, performed on the velocity and acceleration components as a function of positional and demographic data, reoriented to the perspective of the aggressor. The continuous analysis will then investigate differences between PR and VR experimental paradigms by combining the datasets.

Finally, this chapter will summarise and discuss the key results obtained within these datasets, and how they relate to the surrounding literature. This discussion will also consider the limitations of the results, and potential future work that can improve upon these results.

8.1. Comparison between PR and VR experimental paradigms

This chapter performs a quantitative comparison of movement responses between the experimental paradigms. In order to determine the differences be-

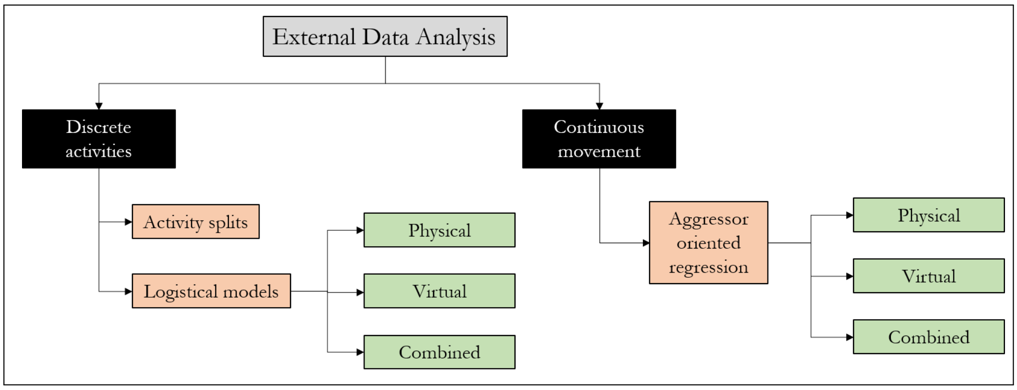


Figure 8.1.: Outline of the chapter

tween these paradigms, the responses are modelled using combined datasets, where the data from VR environments is also marked with a binary flag indicating its source. The comparator models use a list of predictor variables that are common to both datasets, and then a list of equivalent predictor variables that are only non-zero in the VR dataset.

The results of these models are then investigated, with any significance codes noted. Table 8.1 shows a colour-coded chart for assessing VR as a method of generating data on human behaviour in emergencies. If the response is the same in both VR and PR experiments, then the model results will show either no significance across both base and VR-linked variables, or only significance in the base variable (i.e. the VR-linked variables explain no more of the variance). In these instances, the colour classification is green.

However, if significance is only found for the VR-linked variable, then the effect is only significant in VR environments, and the classification is orange.

If significance is present in both base and VR-linked predictor variables, then the resulting classification depends on two factors: the sign and magnitude of the coefficient. If the coefficient sign is the same (i.e. both positive coefficients, or both negative), then the VR environment exaggerates the dependence on this variable, and the classification is yellow. If the sign is different between base and VR-linked variable, then further analysis is required on the standalone models investigating the individual datasets. Here there are three possible options:

i) If the significance disappears within the VR standalone model, then the effect is only present in the PR paradigm. This is classified as orange.

Base variable	VR-linked variable	Classification
Not Significant	Not Significant	Neither
Significant	Not Significant	Both
Not Significant	Significant	VR only
Significant	Significant (opposite sign, higher magnitude, significance disappearing in standalone model)	PR only
Significant	Significant (same sign)	Both VR exagg,
Significant	Significant (opposite sign, lower magnitude, significance remaining in standalone model)	Both VR reduced
Significant	Significant (opposite sign, higher magnitude, significance remaining in standalone model)	Both VR reversed

Table 8.1.: PR-VR comparison classification

If significance is found within the standalone model, there are two remaining avenues:

ii) The sign remains the same. In this instance the effect is present in both paradigms, but the VR environment reduces the dependence on this variable, and the consequent classification is yellow.

iii) The sign is reversed. Here the effect is present in both paradigms, but with a reversed sign in VR when compared with PR environments. This is classified as red.

8.2. Discrete analysis

The next section of this chapter will discuss the investigation of movement responses by the participants. Initially it will focus on the observed discrete activity choices, identifying and quantifying key responses observed by the author. This section will then detail a logistic model of the primary discrete activity choice: whether the participants moved or remained still in response to the aggressor.

Traditionally this analysis type has been performed on datasets that are specifically collected for this purpose, such as surveys or hypothetical choice experiments. These represent a forced-choice data gathering procedure, which limits the possible outcomes. This chapter, conversely, infers the choices made by participants by observing their actions (their revealed choices), representing a far more valid source of data.

8.2.1. Activity splits

During the experiments, the author observed several unique activity choices by the participants. Of these activities, there were several available to all participants. The primary option among these was the choice to remain still or move.

The choice between moving and remaining still can easily be framed as analogous to the fight-flight-freeze response detailed in Section 2.3. A participant was only labelled as remaining still if they did not move more than 1m in total between the moment of the hostile aggressor's arrival and when the participant themselves were taken out of the experiment by a physical touch from the hostile aggressor. Once a participant had decided to move, they were then classified as moving for the remainder of the experiment, regardless of instantaneous speed.

If the participant decided to move, there were several further discrete movement choices available to them. These were:

- Move around central column
- Move behind dividing wall
- Leave room

As these choices were labelled independently, and as participants only needed to perform these activities once to be labelled as having made that decision, it was possible for a single participant to perform all three discrete activities within a single experiment. However, if the participants chose to leave the room, then they were approached by one of the experimental team, who asked to remain there until the end of the experiment. This precludes the participant from being able to then move around the central column or the dividing wall. These options were shown in Section 5.3.

Owing to the limitations of the VR environment there were several discrete choices that were only available to participants within the PR experiment. These were:

Observed action	PR Stress	PR Financial control	VR
1) Remain still	21	0	10
2) Move	40	8	45
2.1) Central column	36	8	36
2.2) Move behind dividing wall	2	0	14
2.3) Leave room	5	0	4
3) Crypsis	3	0	N/A
4) Altruistic intervention	1	0	N/A

Table 8.2.: Activity splits between paradigms

- Crypsis ('playing dead')
- Intervention on behalf of other participants (altruism)

These activities were only performed by minority of participants (4 out of 80 PR participants), and therefore does not represent a significant issue. The breakdown of individual responses within each experimental paradigm is detailed in Table 8.2. The experimental paradigms considered are the PR Stress experiments, the PR Financial control experiment and the VR experiments. The PR Aggressor control was excluded from this analysis, since the participants' activity choices were not dependent on an external individual, and therefore are not comparable.

There is a notable overlap in the movement type responses between PR Stress and VR datasets, while there are observable differences between these datasets and the PR Financial control dataset. Specifically, the choice of any participants to remain still is conspicuously absent from the PR Financial control dataset, as is the choice of any participants to move behind the wall, leave the room, perform crypsis or intervene on each other's behalf.

8.2.2. Logistic models

This activity choice split provides some validation for the separate experimental paradigms and will be investigated further below. The only discrete choice with enough data to fully analyse is the choice to move or remain still. Therefore, it was decided to model this choice using a variety of inputs, including social observations. This analysis will detail the models obtained for both PR and VR paradigms, before combining these to reveal any differences.

In order to find the relevant factors in making this choice, a logistic regression was performed, aiming to predict whether a participant was moving or standing still, given a variety of inputs. These models aim to predict the probability of a discrete outcome, using a range of predictor variables (which can be either continuous or discrete). This predictive ability is obtained by fitting a sigmoid (of the form shown in Figure 8.2) to the observed data.

The sigmoid function, provided below, calculates the probability that a state occurs, given a set of inputs X_i. The sigmoid therefore has a range between 0 and 1 and is valid for all X_i.

$$P(Y=1) = \frac{e^{\alpha+\Sigma\beta_i X_i}}{1+e^{\alpha+\Sigma\beta_i X_i}} = \frac{1}{1+e^{-(\alpha+\Sigma\beta_i X_i)}} \tag{8.1}$$

By fitting the sigmoid function to the observed data, it is possible to interpret the coefficients of each predictor variable (β_i) according to whether they cause the resulting state to be more or less likely. This is known as the odds ratio.

While logistic models are extendable to multiple discrete choices, as multinomial logit models, the analysis performed here will be specifically a binary-logistic model, aiming to understand the important predictor variables for a participant's binary decision to move or remain still. Simply put, a higher coefficient yields a higher probability of a participant's choice to move.

This approach will also produce a confusion matrix, a grid of results that shows the ability of the model to classify positive and negative results. An example confusion matrix is shown in Table 8.3. This shows the relationships between the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) classifications, both in absolute numbers and proportions. A perfect classifier would have unit sensitivity and specificity.

In addition to fitting sigmoid functions and producing confusion matrices, the discrete choice models can output a receiver operating characteristic (ROC) curve, a graphical plot that illustrates the classification ability of the logistic models above by plotting the ranked logistic outputs based on their respective calculated probabilities. The plot compares the true positives and false positives for all datapoints, with a perfect classifier represented by a vertical line.

For more information on discrete choice modelling, there are numerous reference sources available (e.g.: [276]).

Example confusion matrix			Actual classification	
			Condition Positive	Condition Negative
Model prediction	Absolute	Positive	TP	FP
		Negative	FN	TN
	Proportion	Positive	$\frac{TP}{TP+FP}$	$\frac{FP}{TP+FP}$
		Negative	$\frac{FN}{TN+FN}$	$\frac{TN}{TN+FP}$
Sensitivity	$\frac{TP}{TP+FN}$			
Specificity	$\frac{TN}{TN+FP}$			

Table 8.3.: Example confusion matrix

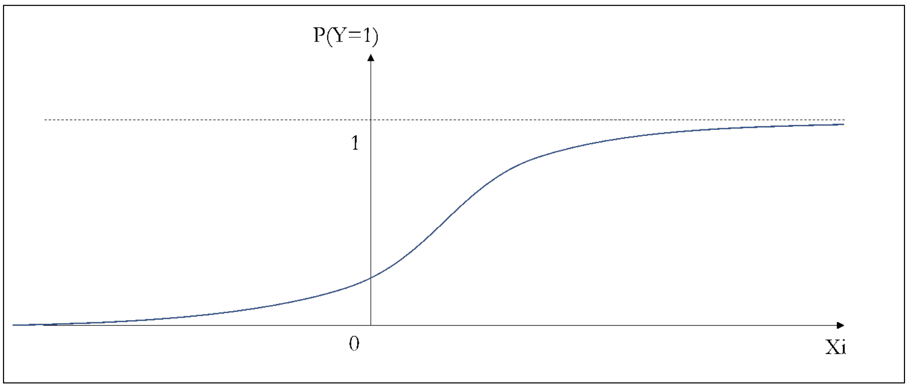


Figure 8.2.: Sigmoid form

8.2.3. Data collection

Data for this model was obtained by taking snapshots of the experiments at each discrete event, defined as either 1) the aggressor touching a participant, or 2) a participant's decision to move. Each snapshot was then analysed, providing the dataset for the logistic regression, similar to the method used in [264].

The information collated from every snapshot included:

- Each participant ID, and related demographic information
- The number of participants who were already moving

- The number of participants who had already been caught
- The distance between the aggressor and each participant
- Whether the participants are within the aggressor's field of view, defined as within ±60° of the direction of travel.

This is shown (not to scale) graphically below:

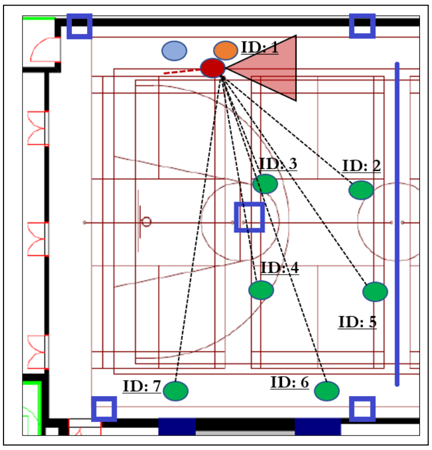


Figure 8.3.: Discrete data (i)

Discrete event: Participant 1 caught

- Number moving: 0
- Number caught: 1
- Distances: {1:NA, 2: 5.2m, 3:7.2m, ...}
- Within FOV: {1: NA, 2: False, 3: False...}
- Moving: {1: NA, 2: False, 3: False, ...}

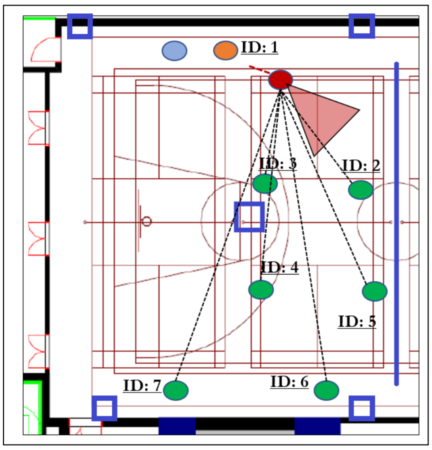


Figure 8.4.: Discrete data (ii)

Discrete event: None

No data collected

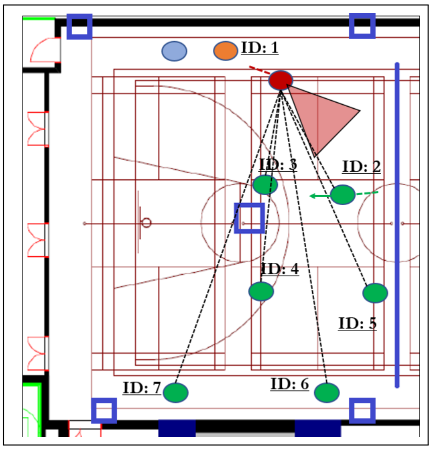


Figure 8.5.: Discrete data (iii)

Discrete event: Participant 2 initiates movement

- Number moving: 0
- Number caught: 1
- Distances: {1:NA, 2: 5.4m, 3:5.1m, . . . }
- Within FOV: {1: NA, 2: True, 3: False. . . }
- Moving: {1: NA, 2: True, 3: False, . . . }

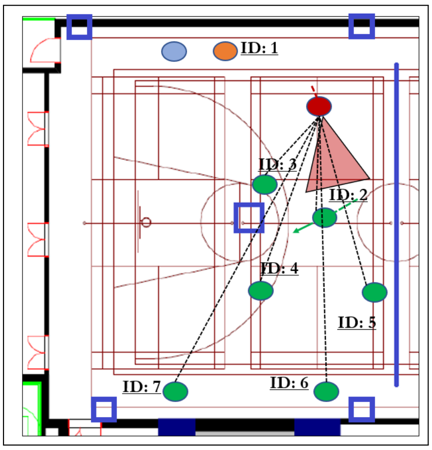


Figure 8.6.: Discrete data (iv)

Discrete event: None

No data collected

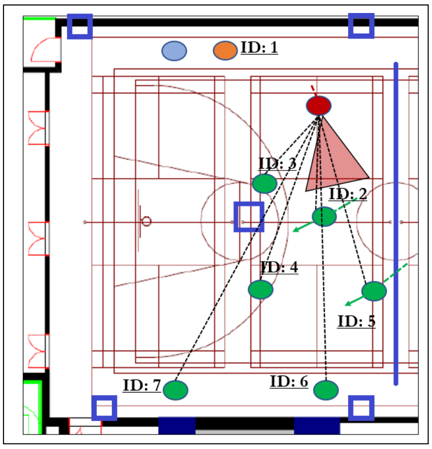


Figure 8.7.: Discrete data (v)

Discrete event: Participant 5 initiates movement

- Number moving: 1
- Number caught: 1
- Distances: {1:NA, 2: 3.8m, 3:2.9m, . . . }
- Within FOV: {1: NA, 2: True, 3: False. . . }
- Moving: {1: NA, 2: True, 3: False, . . . }

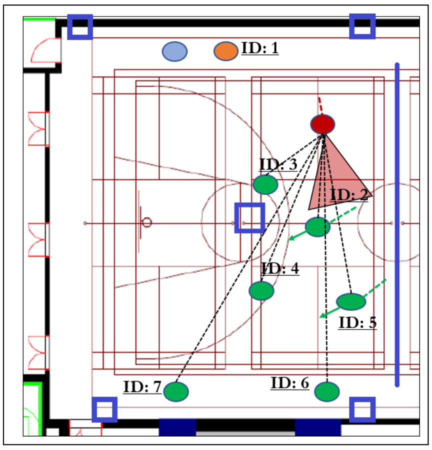


Figure 8.8.: Discrete data (vi)

Discrete event: None

No data collected

The predictor variables used within these models are provided below in Table 8.4.

8.2.4. PR Stress data

The results of the logistic regression for the PR Stress experiments are shown below in Table 8.5, with a positive coefficient indicating a positive relationship with the probability of a person moving.

The PR Stress dataset logistic model indicates that the decision to move is initially strongly unlikely, as well as strongly influenced by the number of ob-

<table>
<tr><th>Subset</th><th>Predictor variable</th></tr>
<tr><td rowspan="4">External information</td><td>Number of moving participants observed</td></tr>
<tr><td>Number of attacked participants observed</td></tr>
<tr><td>Distance to aggressor</td></tr>
<tr><td>Within aggressor's FOV</td></tr>
<tr><td rowspan="3">Demographic information</td><td>Gender</td></tr>
<tr><td>Exercise level</td></tr>
<tr><td>Age</td></tr>
</table>

Table 8.4.: Predictor variables used for Logistic Regression

served moving people, the number of people that have previously been touched by the aggressor, and whether the aggressor is currently moving towards the participant. The decision on whether to move or not is also dependent on the gender and exercise level of the participant.

However, the model also suggests that this decision is independent of the distance between the participant and the aggressor, as well as of the age of the participant.

8.2.5. VR data

The VR data logistic model suggests that, within a VR environment, the decision to move is initially strongly unlikely, as well as strongly based on the number of observed moving avatars, and the number of avatars that have been touched by the virtual aggressor. Furthermore, the model suggests that the demographics of the participant have an effect, with the decision to move being more likely for female participants.

However, it also suggests that the decision to move is independent of the distance to the aggressor, whether the virtual aggressor is heading towards the participant, and both the age and the exercise level of the participant.

8.2.6. Combined data

The next step in this analysis is to combine these datasets and analyse the resulting model to identify any differences in participant decision-making between the PR and VR experimental paradigms.

This is performed by combining the datasets that were used to generate

PR Stress data					
Variable	**Coefficient**	**Std error**	**Odds-ratio**	**Z-value**	**p-value**
Constant	-3.616	0.993	0.027	-3.643	0.0003***
Number reacted	0.614	0.049	1.848	12.486	8.92e-36***
Number caught	0.289	0.061	1.335	4.736	2.18e-06***
Distance to aggressor	-0.018	0.022	0.982	-0.818	0.413
Within FOV	0.430	0.231	1.538	1.867	0.062
F	-0.779	0.220	0.459	-3.544	0.0004***
Exercise level	0.306	0.090	1.358	3.390	0.0007***
Age	-0.020	0.030	0.980	-0.683	0.495
Mcfadden's pseudo r^2			0.382		
Log-likelihood (LLR p-value)			-319.84 (2.284e-81***)		

Table 8.5.: Logistic regression results for PR Stress discrete movement analysis

PR Stress data			Actual classification	
			Move	Remain still
Model classification	Absolute numbers	Move	411	57
		Remain still	81	223
	Proportions	Move	0.878	0.122
		Remain still	0.266	0.734
Sensitivity	0.836			
Specificity	0.796			

Table 8.6.: Confusion matrix for PR Stress logistic model

VR data					
Variable	**Coefficient**	**Std error**	**Odds-ratio**	**Z-value**	**p-value**
Constant	-6.063	1.593	0.002	-3.806	0.0001***
Number reacted	0.842	0.093	2.321	9.085	1.04e-19***
Number caught	0.471	0.138	1.601	3.400	0.0007***
Distance to aggressor	0.057	0.063	1.059	0.905	0.366
Within FOV	-0.341	0.358	0.711	-0.951	0.342
F	0.820	0.326	2.270	2.518	0.0118*
Exercise level	-0.044	0.137	0.957	-0.321	0.748
Age	0.060	0.049	1.061	1.215	0.224
Mcfadden's pseudo r^2			0.473		
Log-likelihood (LLR p-value)			-139.67 (1.563e-50***)		

Table 8.7.: Logistic regression results for VR discrete movement analysis

VR data			Actual classification	
			Move	Remain still
Model classification	Absolute numbers	Move	254	32
		Remain still	43	93
	Proportions	Move	0.888	0.112
		Remain still	0.316	0.684
Sensitivity	0.855			
Specificity	0.744			

Table 8.8.: Confusion matrix for VR logistic model

Combined data						
Variable		**Coefficient**	**Std error**	**Odds-ratio**	**Z-value**	**p-value**
PR data	Constant	-3.616	0.993	0.027	-3.643	0.0003***
	Number reacted	0.614	0.049	1.848	12.486	8.92e-36***
	Number caught	0.289	0.061	1.335	4.736	2.18e-06***
	Distance to aggressor	-0.018	0.022	0.982	-0.818	0.413
	Within FOV	0.430	0.231	1.538	1.867	0.062
	F	-0.779	0.220	0.459	-3.544	0.0004***
	Exercise level	0.306	0.090	1.358	3.390	0.0007***
	Age	-0.020	0.030	0.980	-0.683	0.495
VR differences	VR Constant	-2.447	1.877	0.087	-1.304	0.192
	VR Number reacted	0.228	0.105	1.256	2.169	0.0301*
	VR Number caught	0.181	0.151	1.199	1.199	0.230
	VR Distance to aggressor	0.075	0.067	1.078	1.122	0.262
	VR Within FOV	-0.771	0.426	0.463	-1.810	0.070
	VR F	1.599	0.393	4.948	4.070	4.71e-05***
	VR Exercise level	-0.350	0.164	0.705	-2.136	0.0327*
	VR Age	0.080	0.057	1.083	1.393	0.164
Mcfadden's pseudo r^2				0.415		
Log-likelihood (LLR p-value)				-459.52 (2.266e-129***)		

Table 8.9.: Logistic regression results for Combined discrete movement analysis

models for the PR Stress and VR paradigms (Table 8.7 and Table 8.5), while differentiating between them using VR-linked predictor variables. These are identical to the previous predictor variables, except that they are only non-zero in datasets from VR experiments. After fitting the model, any significant differences between the experimental paradigms are identified within these predictor variables.

8.2.7. Comparison of experimental paradigms

The combined model shows that some predictor variables are present in both PR and VR paradigms, while some are either only present in the PR paradigm, and some are present only in the VR paradigm. This is shown in a colour-coded format below in Table 8.11, using the coding taxonomy defined in Table 8.1.

Combined data			Actual classification Move	Remain still
Model classification	Absolute numbers	Move	667	87
		Remain still	126	314
	Proportions	Move	0.885	0.115
		Remain still	0.286	0.714
Sensitivity	0.841			
Specificity	0.783			

Table 8.10.: Confusion matrix for Combined logistic model

Subset	**Predictor**	**Paradigms**
Constant	Constant	Both
Situational	Number reacted	Both VR exaggerated
	Number caught	Both
	Distance to aggressor	Neither
	Within FOV	Neither
Demographic	F	Both VR reversed
	Exercise level	PR only
	Age	Neither

Table 8.11.: Comparison of discrete events between paradigms

Similarities

As can be seen in Table 8.11, there are several similarities between PR-Stress and VR datasets. The similarities are identified by the p-value of the VR-linked predictor variables, where non-significant variables ($p>0.05$) are considered similar across paradigms.

- Constant

Both PR and VR models indicate that the initial decision to move is highly unlikely.

- Number caught

The PR and VR model both indicate that a participant is more likely to move as the aggressor approaches multiple other participants.

- Distance to aggressor

The PR model indicates that a participant is less likely to move as the aggressor approaches them, while the VR model suggests the opposite. However, neither of these effects are significant.

- In aggressor FOV

The PR model indicates that a participant is more likely to move as the aggressor is moving towards them while the VR model suggests the opposite; however, both of these effects are only marginally significant ($p<0.1$).

- Age

Neither of the PR or VR models suggest significant effects based on age. However, given the limitations on recruitment (i.e. participants aged between 18-35), it is possible that a broader demographic might have a different outcome. This also contrasts with the results from the continuous analysis, below.

Differences

The differences are detailed below, split into categories corresponding to the comparisons classification detailed above (Table 8.1). Specifically, these differences are where a paradigm:

- Exaggerates a response, with the same coefficient sign.

- Produces a unique response that is not present in the alternate paradigm.
- Reverses the response, with the opposite coefficient sign.

Exaggerated responses

- Number reacted

In both models the observed number of participants that decided to move, ("Number reacted"), yielded a positive relationship with the decision to move. However, this effect was significantly stronger in the VR dataset (additional coefficient=0.2276, p=0.030).

Unique responses

- Exercise

In the PR experiment, participants who had self-assessed with higher exercise levels were also significantly more likely to move (total coefficient =0.306, $p<0.001$), whereas this effect was not seen in the VR environment.

Reversed responses

- Gender

In the PR-Stress dataset, female participants were less likely than male participants to decide to move (total coefficient=-0.7791, $p<0.001$), whereas in the VR dataset this effect was reversed, and the female participants were more likely to decide to move than male participants (total coefficient=0.8199, $p<0.05$). This is the only reversed response detected between paradigms.

8.2.8. ROC Curves

The receiver operating characteristic (ROC) curve for the above models is shown below (Figure 8.9). The curves below show that both models perform similarly, with the VR model performing slightly better than the PR Stress model.

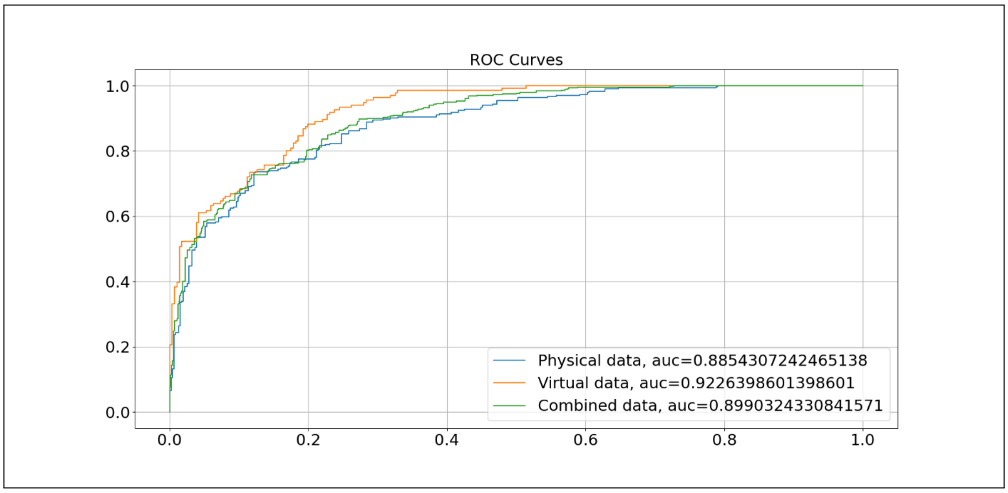


Figure 8.9.: ROC curves for PR Stress, VR, and Overall Stress logistic models

8.2.9. Discrete analysis summary

This analysis shows a breakdown of the discrete activities observed during the stressful experiments, as well as the non-stressful financial-control experiment. Part of this breakdown provides logistic models that identify the important factors within a participant's choice to move or remain still.

The data used to fit these models was collected during discrete events, including whenever a participant moved for the first time after the aggressor entered the room, and whenever the aggressor removed a participant from the experiment.

The analysis shows significant relationships between the predictor variables (both external and demographic) and a participant's decision to move. These datasets are then combined, and within the overall model there is a large overlap in the VR and PR experimental paradigms.

There are three significantly different effects, detailed above, with gender as the only factor where the sign of the predictor variable is reversed. This indicates that female participants are less likely to move within a physical environment, but more likely to move within a virtual environment. Further analysis on this effect is considered out of scope for this PhD, but could represent a fruitful area of sociological research.

The combined model also shows that there is an exaggerated social effect within VR, where participants are far more likely to move when they observe other participants moving than in PR experiments. Finally, this combined model also highlights a unique feature of the different experimental paradigms: within the PR experiment participants who self-assessed as exercising more were more likely to move, but this effect does not appear in the VR experiments.

8.3. Continuous analysis

The next section will focus on the empirical analysis performed on the continuous data obtained from the experiments. This analysis will first focus on highlighting the primary qualitative investigations, before detailing the multiple regression models that were used to characterise the different movement types. The predictor variables used within these models are based on two subsets: positional and demographic. The positional variables encompass all possible symmetric and asymmetric effects, alongside the impact of participant demographic. The models themselves use a forward-backward stepwise variable selection procedure to identify only the significant variables.

The locomotion method used in this experiment meant that the movement speed of participants in the virtual environment was converted from the movement speed of their controllers, using a conversion factor. The use of this conversion factor means that continuous responses in the VR environment may be a scaled version of real-world responses. As a result, the PR and VR models are initially presented parallel to each other, and to quantify any differences between the PR and VR continuous responses, this PhD standardised and combined both datasets, allowing the identification of any significant differences in the continuous responses. This standardisation is described in Equation 5.7.

8.3.1. Aggressor perspective analysis

As has been detailed in Section 5.2, the continuous responses will be analysed in reference to the perspective of the aggressor, shown graphically in Figure 8.10. This provides an analysis that is analagous to the social force model (SFM) approach, discussed in Section 2.1, and also provides an avenue for investigating any potential asymmetrical results. All analysis detailed below will pertain only to participants who had decided to move, to avoid any data-skewing from stationary participants.

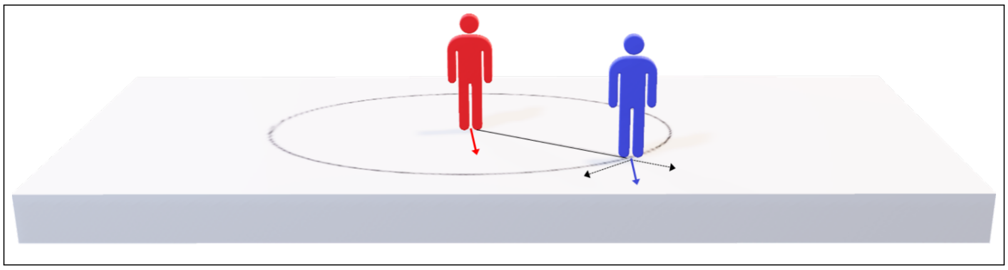


Figure 8.10.: Aggressor (in red) perspective orientation

8.3.2. Walking and running

The primary avenue of investigation was to ascertain whether there were any obvious correlations between the movement of participants and their position relative to the aggressor. An initial investigation, assessing any differences between walking (defined as instantaneous moving speed <3.0m/s) and running (>3.0m/s) is shown graphically in Figure 8.11 (PR experiments) and Figure 8.12 (VR experiments). These graphs show the relative positions of all participant datapoints, coloured red if the individual's instantaneous movement is classified as running, and blue if they are classified as walking.

In both experimental paradigms, participants tended to move faster when they were nearer the aggressor. There is some discernible front-back asymmetry in both datasets, with more running datapoints when participants are in front of the aggressor rather than when they are behind the aggressor. This indicates that the absolute movement speed is not only related to the absolute distance to the aggressor, but also to the direction in which the aggressor is moving (relative to the participant). There is also a potential left-right asymmetry in both datasets, with more running points appearing to the forward-right of the aggressor (quadrant-4 in all figures), although this is more marked in the VR dataset.

These continuous responses will now be analysed further to understand any correlations between relative position and movement speed and direction.

8.3.3. Independent Variables

The X and Y axis form the basis of all positional independent variables. The full correct term used for these variables is 'Heading-corrected relative X' and 'Heading-corrected relative Y'; however, these terms are simplified to X and

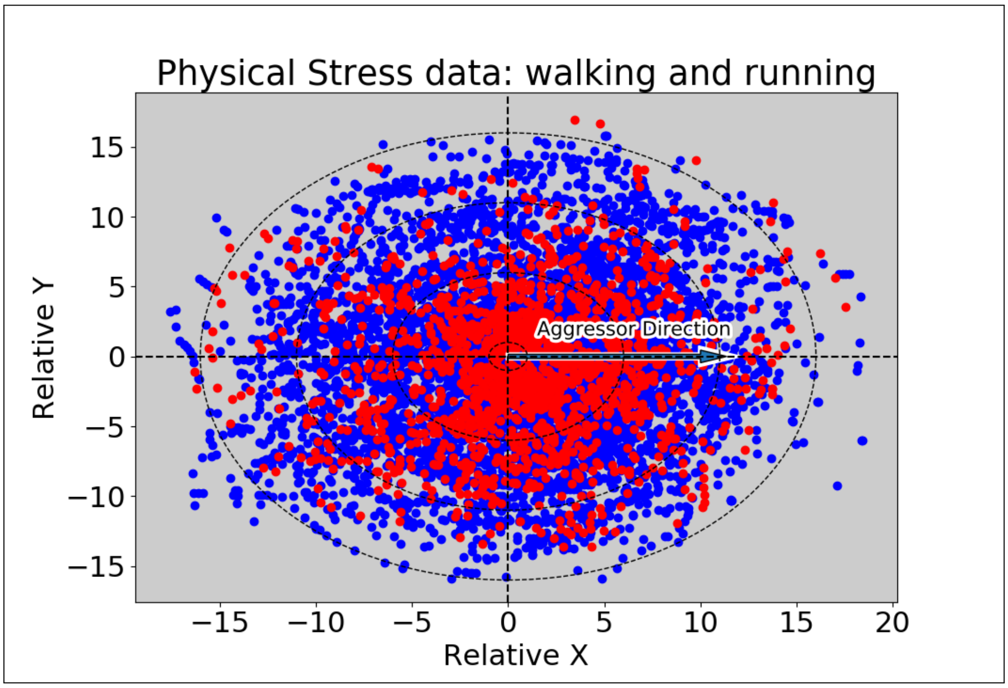


Figure 8.11.: PR Stress running (in red, >3m/s) and walking(in blue, <3m/s) split

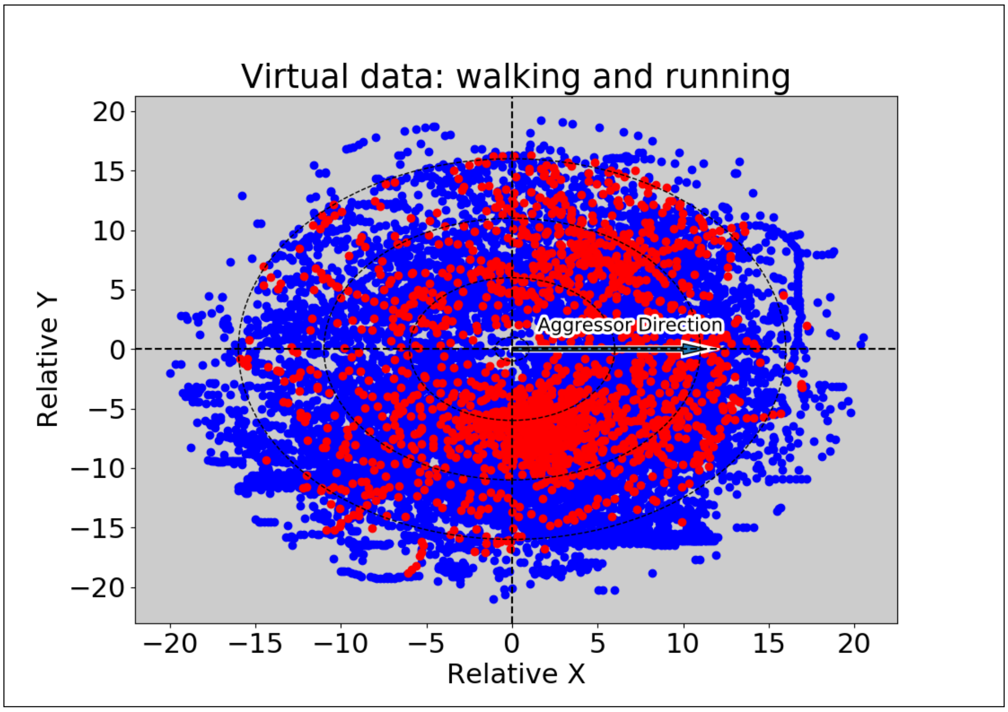


Figure 8.12.: VR running (in red, >3m/s) and walking(in blue, <3m/s) split

Subset	Variable	Linearity	Symmetry
Positional	Constant	Constant	N/A
	X,Y	Linear	Asymmetry
	X^2, Y^2	Non-linear	Mirror-symmetry
	X^3, Y^3	Non-linear	Asymmetry
	Abs(X), Abs(Y)	Linear	Mirror-symmetry
	Distance	Linear	Rotational symmetry
	Distance^2	Non-linear	Rotational symmetry
Demographic	Gender	N/A	N/A
	Age	N/A	N/A
	Exercise level	N/A	N/A

Table 8.12.: Variables used for continuous response analysis

Y for the remainder of the chapter.

The independent variables used in order to capture linear and non-linear effects in two dimensions, as well as accounting for any symmetric (rotational or mirror) and asymmetric effects, are detailed below in Table 8.12.

All of these fourteen variables were used in the regression analyses, which utilised a forward/backward stepwise variable selection process to identify the most important features. Therefore, the reported regression models only include the statistically significant variables, while all other variables were found to have no significant effect on the data.

8.3.4. Dependent Variables

In order to understand the movement behaviour shown above, several different movement types will be analysed. These are split into acceleration and velocity, from a total, tangential and radial perspective, detailed in Table 8.13.

Analysis of both total and tangential movement responses will only be analysed from an absolute perspective. This is due to the potential dilution of a model's discriminatory power to understand correlations between position and

Variable	Directional or Absolute	Details
Total acceleration	Absolute	Useful for pedestrian dynamics models
Total velocity	Absolute	Useful for pedestrian dynamics models
Tangential velocity	Absolute	Signed directions not distinctly different
Tangential acceleration	Absolute	Signed directions not distinctly different
Radial velocity	Both	Clear difference in signed directions
Radial acceleration	Both	Clear difference in signed directions

Table 8.13.: Continuous movement responses

response if direction is included. For example, if two participants are in the same position and moving at the same speed, but in opposite directions (e.g. one clockwise, the other anticlockwise), a simple linear regression model will not be able to discern a correlation between the position and the movement response, as the combined value would be a stationary participant. However, it is considered that clockwise and anticlockwise movement, while strongly context dependent, are similar enough with respect to the aggressor. Therefore, only strictly-positive movement velocity and acceleration values will be used for total and tangential responses.

In contrast to this, radial movement responses will be analysed in both directional (i.e. away from and towards the aggressor) and absolute terms. This is because the different radial movement directions with respect to the aggressor represent a far more relevant decision by the participant.

As all of the proposed dependent variables form a time-series, each datapoint is dependent on the previous state, and therefore the requirement for independent datapoints to produce a linear regression model is not satisfied. Performing this analysis on the total experimental datasets for each paradigm (61 participants for PR Stress, and 55 for VR) will mitigate this, providing some insight into the participants' movement responses. However, further modelling procedures will be discussed in Section 10.4, including time-series analysis through machine-learning approaches.

8.3.5. Log transformed datasets

While the significance levels of the variables used in these models are high, the explanatory power of the models themselves tends to be low (a maximum adjusted r^2 value of 0.228). Further to this, the results from the Omnibus test showed significantly non-normal residuals in all models, indicating that a different modelling approach may be more appropriate, such as non-linear regression. The response to this was to perform analysis on the log-transform of the dependent data. This analysis can only be performed for the absolute responses, and was calculated using the following transformation:

$$L_{data} = \log(1 + x_{data}) \tag{8.2}$$

This transformation results in higher explanatory power (a maximum r^2 value of 0.268). This chapter will show and analyse the results for the untransformed datasets, and the log-transformed results tables will be provided in Appendix E. However this transform does not represent the optimal modelling solution, as the Omnibus test results of these models continued to show significantly non-normal residuals. Further possible approaches will be discussed in Section 10.4.

8.3.6. Combined datasets

As the VR locomotion technique used a coded interface to interpret movement, quantitatively combining the raw positional data is likely to be inaccurate. Therefore, in order to combine the two datasets, it is first necessary to ensure that the two datasets are similar across macro-scale measures, by utilising the standardisation process defined in Equation 5.7.

This transformation was applied to the strictly positive datasets:

- Total acceleration
- Total velocity
- Tangential acceleration
- Tangential velocity
- Radial acceleration (absolute)
- Radial velocity (absolute)

The standardised datasets were then modelled using the same stepwise variable selection procedure as above, but including further VR-linked predictor variables, which are only non-zero for the VR datasets, similar to the combined discrete analysis.

8.3.7. Multiple regression models

The following sections will discuss the different continuous analysis approaches, highlighting any common features of all models (PR Stress and VR), and commenting on any unique model elements.

As most of the datapoints used within these models are within 15m of the aggressor, these models are only valid for this region. This is reflected in the fact that these models often produce unbounded behaviour for large positive or negative positional values. Therefore, this section will only comment on the behaviour of the models within these values.

The tables over the next section will report all coefficients, standard errors, t-statistics and significance levels of the continuous analysis. Further to this, the F-statistic (and significance level) is reported, alongside a test for residual normality (the omnibus test, providing a χ^2 value and significance level) and the adjusted r^2 value. Owing to the forward-backward stepwise selection process, all reported variables are significant, however, the adjusted r^2 value for these models remains low. As mentioned above, this value is improved by taking a log-transform of the datasets. In addition to this, the VR models have a higher explanatory power than the PR models, and there are generally more statistically significant variables in each model that remain after the selection process. This is at least partially due to measurement noise within the PR experimental paradigm; however, it is also possible that this is due to differences in behavioural responses within the VR environment. This will be investigated below.

The graphs accompanying the tables show the result of the models over a range of positional values. The position and direction of the aggressor (which is also the origin in these graphs) is shown by an arrow. A dashed circle is also shown in each graph, highlighting the 15m circle of validity within which most of the data lies.

Total acceleration and total velocity

The following tables show the models generated for total acceleration (Table 8.14) and total velocity (Table 8.15). Within these tables the models for PR Stress and VR data are presented side by side, for qualitative assessment.

Total acceleration

All models of total acceleration have a local maximum approximately located at the origin, indicating that the acceleration of participants increases as the aggressor approaches them. All acceleration models appear to have approximate front-back symmetry and left-right asymmetry. However, the PR models indicate a rising level of acceleration on the aggressor's right-hand side, whereas the VR model indicate a rising rate of acceleration on the aggressor's left-hand side. Demographically, all models indicate that older participants accelerate at a lower rate.

The VR model suggests that acceleration is based on further demographic details, with participants who self-rate as exercising more producing higher levels of acceleration. While the VR model suggests that female participants accelerate at a lower rate, the PR model indicates that they accelerate at a higher rate than male participants.

Total velocity

All models have a local maximum approximately located at a small offset from the origin, indicating that participant velocity decreases with distance from the aggressor. Similarly, all models indicate that velocity increases with self-reported exercise levels and decreases with age.

The VR model suggests that female participants move more slowly, however, this effect is not reflected in the PR model. The VR model has front-back and left-right asymmetry, whereas the PR model has approximate front-back and left-right symmetry.

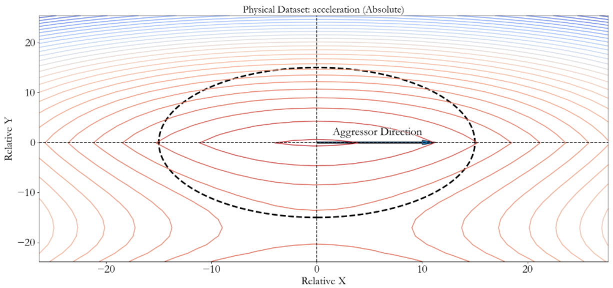
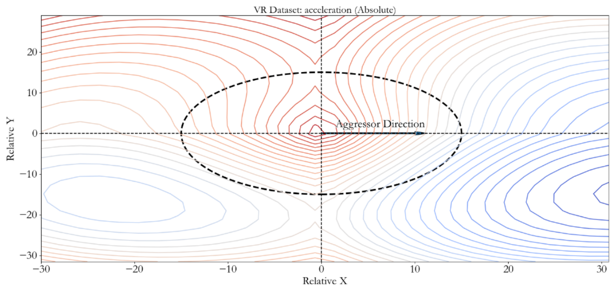
PR Stress: Acceleration (Absolute, ms^{-2})						**VR: Acceleration (Absolute, ms^{-2})**				
										
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	7.2912	0.6742	10.8139	4.53e-27***	Constant	8.8784	0.2790	31.8227	1.81e-218***
Positional	Distance					Distance	-3.27e-01	0.0471	-6.9555	3.59e-12***
	Distance^2	-0.0052	0.0018	-2.8240	0.0048**	Distance^2	1.08e-02	0.0011	9.2324	2.82e-20***
	X					X	-0.0840	0.0040	-20.9633	7.44e-97***
	X^2					X^2	0.0032	0.0014	2.3141	0.0207*
	X^3					X^3				
	Abs(X)					Abs(X)	-0.2391	0.0285	-8.3770	5.67e-17***
	Y					Y	0.1504	0.0085	17.7514	3.96e-70***
	Y^2					Y^2	0.0076	0.0014	5.4635	4.71e-08***
	Y^3	-0.0005	0.0001	-3.6361	0.0003***	Y^3	-0.0002	5.59e-05	-3.1850	0.0014**
	Abs(Y)	-0.0826	0.0335	-2.4683	0.0136*	Abs(Y)	-0.1238	0.0345	-3.5894	0.0003***
Demog.	Age	-0.0965	0.0266	-3.6233	0.0003***	Age	-0.0314	0.0091	-3.4486	0.0006***
	Exercise					Exercise	0.0679	0.0241	2.8124	0.0049**
	F	0.5112	0.2244	2.2786	0.0227*	F	-0.2668	0.0620	-4.3019	1.70e-05***
Adjusted r^2				0.009		**Adjusted r^2**			0.135	
F-statistic (p-value)				15.7166 (1.9789e-15***)		**F-statistic (p-value)**			413.663 (0.000***)	
Omnibus test (p-value)				8438.704 (0.000***)		**Omnibus test (p-value)**			7235.928 (0.000***)	

Table 8.14.: PR Stress and VR acceleration models

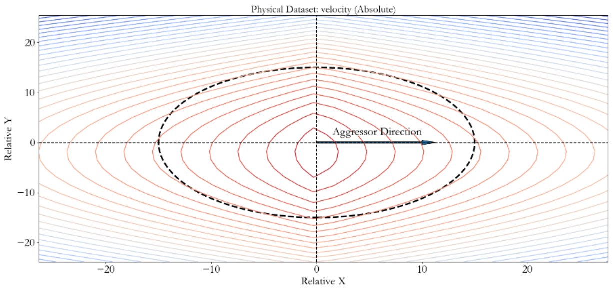
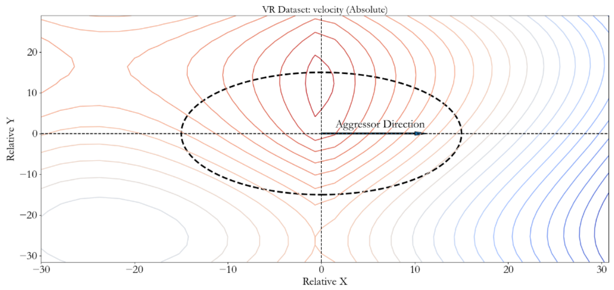
PR Stress: Velocity (Absolute, ms^{-1})						VR: Velocity (Absolute, ms^{-1})				
										
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	3.56545	0.2534	14.0700	1.93e-44***	Constant	2.3875	0.0545	43.8230	0.000***
Positional	Distance					Distance	-0.0332	0.0027	-12.4910	1.03e-35***
	Distance^2					Distance^2				
	X					X	-0.0092	0.0018	-5.0610	4.20e-07***
	X^2					X^2	0.0017	0.0004	4.4190	9.95e-06***
	X^3					X^3	-7.37e-05	1.24e-05	-5.9400	2.88e-09***
	Abs(X)	-0.0555	0.0065	-8.5178	1.92e-17***	Abs(X)	-0.0906	0.0057	-15.8612	2.03e-56***
	Y	-0.0159	0.0041	-3.8378	0.0001***	Y	0.0522	0.0017	31.4596	1.22e-213***
	Y^2	-0.0059	0.0005	-11.1080	1.84e-28***	Y^2				
	Y^3					Y^3	-9.17e-05	1.03e-05	-8.9407	4.09e-19***
	Abs(Y)					Abs(Y)				
Demog.	Age	-0.0620	0.0079	-7.8693	4.03e-15***	Age	-0.0255	0.0020	-12.553	4.76e-36***
	Exercise	0.0931	0.0243	3.8271	0.0001***	Exercise	0.0416	0.0054	7.7208	1.19e-14***
	F					F	-0.1038	0.0138	-7.4983	6.65e-14***
				Adjusted r^2	0.038				**Adjusted r^2**	0.223
				F-statistic (p-value)	‘64.852 (1.352e-66***)				**F-statistic (p-value)**	833.474 (0.000***)
				Omnibus test (p-value)	6021.877 (0.000***)				**Omnibus test (p-value)**	5609.697 (0.000***)

Table 8.15.: PR Stress and VR velocity models

Tangential acceleration and tangential velocity

The full results are shown below for the tangential acceleration (Table 8.16) and tangential velocity (Table 8.17).

Tangential acceleration

All models indicate a local maximum at the origin, suggesting that tangential acceleration increases as the aggressor approaches the participants. Furthermore, all models appear to have approximate front-back symmetry and left-right asymmetry. However, the PR model suggests that tangential acceleration increases to the right-hand side of the aggressor, while the VR model indicates an increase in tangential acceleration to the left-hand side of the aggressor. All models suggest that tangential acceleration decreases with age.

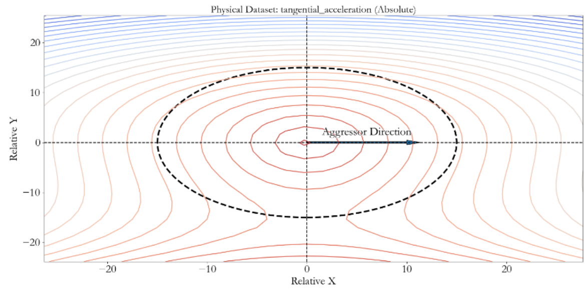
The VR model suggests a positive relationship between self-assessed exercise level and tangential acceleration. The models also show that female participants accelerate at a higher rate than male participants within PR paradigms, while female participants accelerate at a lower rate in VR models.

Tangential velocity

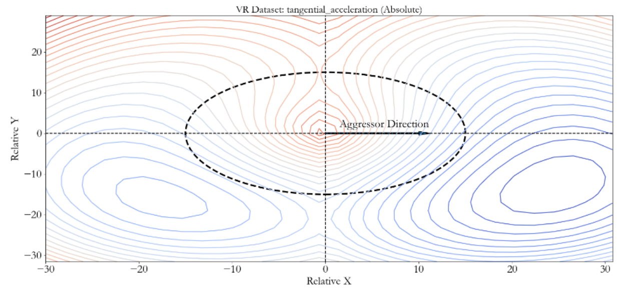
All models have a local maximum, located approximately at the origin, indicating that participant tangential velocity decreases with distance from the aggressor. Further to this, all models suggest an approximate front-back symmetry, and left-right asymmetry. All models also indicate that tangential velocity decreases with age and increases with self-assessed exercise level.

The VR model suggests that female participants moved with a lower average tangential velocity than male participants. Similar to the tangential acceleration models, the PR model suggests an increase in tangential velocity to the right-hand side of the aggressor, whereas the VR model suggests an increase to the left-hand side.

PR Stress: Tangential acceleration (Absolute, ms^{-2})



VR: Tangential acceleration (Absolute, ms^{-2})



Sub.	Variable	Coeff.	Std. error	t-value	p-value	Variable	Coeff.	Std. error	t-value	p-value
C	Constant	4.9950	0.4963	10.0654	1.08e-23***	Constant	6.8641	0.2252	30.4801	7.18e-201***
Positional	Distance	-0.0909	0.0158	-5.7687	8.28e-09***	Distance	-0.3889	0.0331	-11.7593	7.48e-32***
	Distance^2					Distance^2	0.0114	0.0010	11.9456	8.18e-33***
	X					X	-0.0559	0.0031	-17.7757	2.58e-70***
	X^2					X^2	0.0072	0.0009	7.7949	6.66e-15***
	X^3					X^3				
	Abs(X)					Abs(X)	-0.2096	0.0219	-9.5826	1.02e-21***
	Y					Y	0.1166	0.0064	18.3408	1.04e-74***
	Y^2					Y^2	0.0042	0.0009	4.7932	1.65e-06***
	Y^3	-0.0003	0.0001	-2.8764	0.0040**	Y^3	-0.0001	4.17e-05	-2.8590	0.0043**
	Abs(Y)					Abs(Y)				
Demog.	Age	-0.0632	0.0194	-3.2670	0.0011**	Age	-0.0239	0.0072	-3.3212	0.0009***
	Exercise					Exercise	0.0851	0.0191	4.4629	8.12e-06
	F	0.5917	0.1628	3.6336	0.0003***	F	-0.1526	0.0490	-3.1159	0.0018**
				Adjusted r^2	0.008				**Adjusted r^2**	0.129
				F-statistic (p-value)	18.027 (9.463e-15***)				**F-statistic (p-value)**	431.115 (0.000***)
				Omnibus test (p-value)	9194.157 (0.000***)				**Omnibus test (p-value)**	10930.646 (0.000***)

Table 8.16.: PR Stress and VR tangential acceleration models

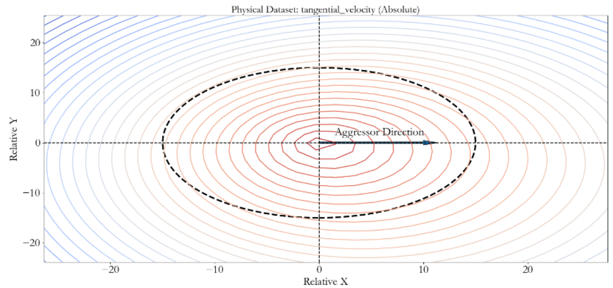
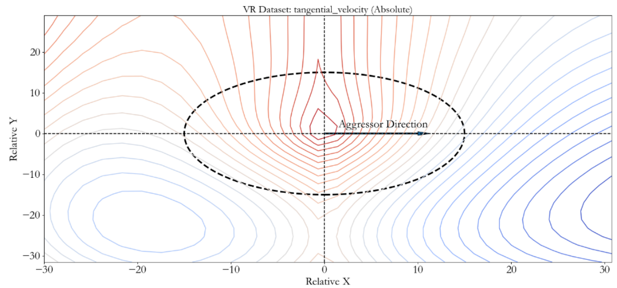
PR Stress: Tangential velocity (Absolute, ms^{-1})						VR: Tangential velocity (Absolute, ms^{-1})				
										
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	2.3455	0.1813	12.9362	6.67e-38***	Constant	1.9839	0.0518	38.2841	0.000***
Positional	Distance	-0.0511	0.0042	-12.1201	1.61e-33***	Distance	-0.0672	0.0076	-8.8222	1.18e-18***
	Distance^2					Distance^2	0.0020	0.0004	5.4335	5.57e-08***
	X	0.0096	0.0029	3.3093	0.0009***	X	-0.0077	0.0015	-5.2367	1.65e-07***
	X^2					X^2	0.0022	0.0004	6.2018	5.66e-10***
	X^3					X^3	-3.58e-05	1.01e-05	-3.5502	0.0004***
	Abs(X)					Abs(X)	-0.0865	0.0050	-17.2021	5.45e-66***
	Y	-0.0081	0.0030	-2.7231	0.0065**	Y	0.0333	0.0015	22.8016	4.48e-114***
	Y^2					Y^2				
	Y^3					Y^3	-3.94e-05	9.58e-06	-4.1162	3.86e-05***
	Abs(Y)					Abs(Y)				
Demog.	Age	-0.0420	0.0056	-7.4693	8.90e-14***	Age	-0.0187	0.0017	-11.3517	8.41e-30***
	Exercise	0.0922	0.0174	5.3039	1.16e-07***	Exercise	0.0429	0.0043	9.7974	1.25e-22***
	F					F	-0.0643	0.0112	-5.7219	1.06e-08***
Adjusted r^2				0.035		**Adjusted r^2**			0.206	
F-statistic (p-value)				59.435 (5.4993e-61***)		**F-statistic (p-value)**			682.399 (0.000***)	
Omnibus test (p-value)				5972.2111 (0.000***)		**Omnibus test (p-value)**			8010.850 (0.000***)	

Table 8.17.: PR Stress and VR tangential velocity models

Radial acceleration and radial velocity (absolute)

The full results are shown below for the radial acceleration (Table 8.18), and radial velocity (Table 8.19).

Radial acceleration (absolute)

There are significant similarities between the PR and VR models. All models indicate a local maximum approximately located at the origin, suggesting that radial acceleration increases as the aggressor approaches. Further to this, both models indicate a saddle point to the right-hand side of the aggressor. The models also suggest a decreasing relationship with age.

The VR model suggests that female participants accelerate radially at a lower rate than male participants.

Radial velocity (absolute)

Both models show decreasing radial velocity with increasing distance, and both models indicate decreasing radial velocity with age and gender.

The PR model shows front-back and left-right symmetry, while the VR model do not appear to show any symmetry. While the PR model indicates a local maximum at the origin, the VR model shows a local maximum to the left-hand side of the aggressor. Additionally, the VR model suggests that participants a higher self-assessed exercise level move faster.

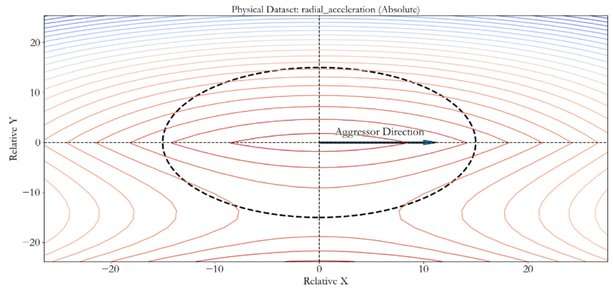
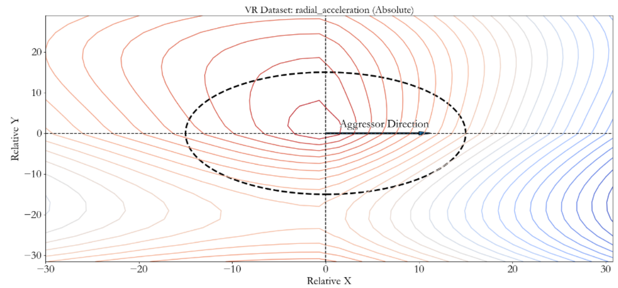
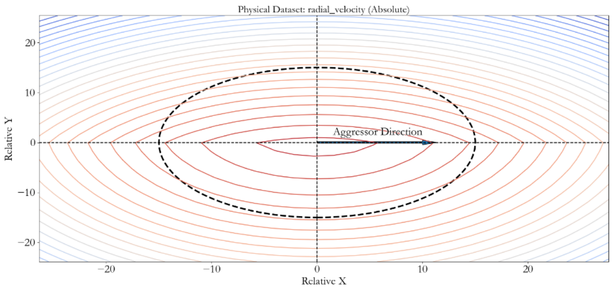
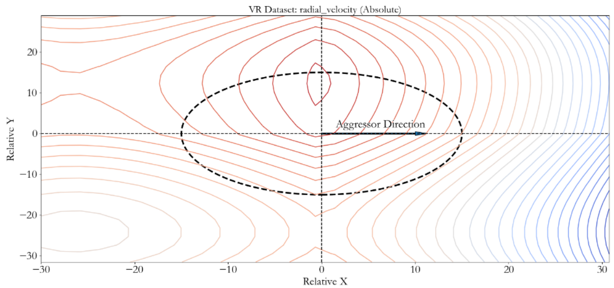
PR Stress: Radial acceleration (Absolute, ms^{-2})						VR: Radial acceleration (Absolute, ms^{-2})				
										
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	4.7197	0.4668	10.1103	6.89e-24***	Constant	4.4283	0.1673	26.4697	1.42e-152***
Positional	Distance					Distance	-0.0965	0.0263	-3.6765	0.0002***
	Distance^2					Distance^2				
	X					X	-0.0540	0.0027	-19.9600	4.80e-88***
	X^2	-0.0035	0.0014	-2.5886	0.0097**	X^2				
	X^3					X^3				
	Abs(X)					Abs(X)	-0.0504	0.0183	-2.7482	0.0060**
	Y					Y	0.0871	0.0059	14.7482	4.76e-49***
	Y^2					Y^2	0.0046	0.0014	3.1413	0.0017**
	Y^3	-0.0004	0.0001	-3.5720	0.0004***	Y^3	-0.0002	0.0000	-4.9722	6.66e-07***
	Abs(Y)	-0.0995	0.0179	-5.5586	2.81e-08***	Abs(Y)	-0.0519	0.0230	-2.2549	0.0241*
Demog.	Age	-0.0642	0.0187	-3.4391	0.0006***	Age	-0.0200	0.0064	-3.1428	0.0017**
	Exercise					Exercise				
	F					F	-0.1990	0.0408	-4.8793	1.07e-06***
				Adjusted r^2	0.007				**Adjusted r^2**	0.073
				F-statistic (p-value)	15.002 (3.2050e-12***)				**F-statistic (p-value)**	255.730 (0.000***)
				Omnibus test (p-value)	9533.032 (0.000***)				**Omnibus test (p-value)**	12625.929 (0.000***)

Table 8.18.: PR Stress and VR Radial acceleration (absolute) models





PR Stress: Radial velocity (Absolute, ms^{-1})						VR: Radial velocity (Absolute, ms^{-1})				
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	2.5909	0.1493	17.3596	2.62e-66***	Constant	1.3857	0.0389	35.6647	1.12e-72***
Positional	Distance					Distance	-0.0188	0.0058	-3.2629	0.0011**
	Distance^2	-0.0020	0.0004	-5.0367	4.84e-07***	Distance^2				
	X					X	-0.0043	0.0012	-3.4853	0.0005***
	X^2					X^2				
	X^3					X^3	-0.0001	0.0000	-6.3759	1.85e-10***
	Abs(X)					Abs(X)	-0.0316	0.0040	-7.8589	4.01e-15***
	Y	-0.0153	0.0034	-4.4661	8.07e-06***	Y	0.0301	0.0013	23.2353	2.43e-118***
	Y^2					Y^2	0.0007	0.0003	2.1530	0.0313*
	Y^3					Y^3	-0.0001	0.0000	-6.5598	5.48e-11***
	Abs(Y)	-0.0281	0.0074	-3.7911	0.0002***	Abs(Y)	-0.0146	0.0050	-2.8951	0.0038**
Demog.	Age	-0.0463	0.0059	-7.8553	4.50e-15***	Age	-0.0156	0.0014	-11.2084	4.26e-29***
	Exercise					Exercise	0.0148	0.0037	4.0076	6.15e-05***
	F	-0.1100	0.0497	-2.2145	0.0268*	F	-0.0658	0.0095	-6.9357	4.13e-12***
Adjusted r^2					0.022	**Adjusted r^2**				0.151
F-statistic (p-value)					37.812 (1.7784e-38***)	**F-statistic (p-value)**				470.096 (0.000***)
Omnibus test (p-value)					7468.050 (0.000***)	**Omnibus test (p-value)**				12243.562 (0.000***)

Table 8.19.: PR Stress and VR Radial velocity (absolute) models

Radial acceleration and radial velocity (directional)

The next section will discuss the relationships observed when modelling the radial acceleration and velocity from the perspective of the aggressor. The full results are shown below for the radial acceleration (Table 8.20) and radial velocity (Table 8.21). Within these tables the models for PR Stress and VR data are presented side-by-side, for qualitative assessment.

There are no log models available in Appendix E for the directional dataset, as the absolute value of the datapoints must be analysed, and therefore the resulting analysis is identical to that of the log-radial (absolute) datasets.

Radial acceleration (directional)

Within this analysis, it is seen that the PR data produces a null model, suggesting that radial acceleration is independent of position and demographic. This is the only occurrence of complete independence within the continuous analysis section, which consequently means that there are no common features to report.

The VR model suggests that radial acceleration is also independent of demographic information, and also no significant constant identified. The only other model where this occurs is the VR radial velocity- directional model (see below), and there is a consequent local minimum located at the origin. This suggests that radial acceleration increases with distance from the aggressor.

Radial velocity (directional)

Both models appear to have front-back symmetry, and some degree of left-right symmetry. This is especially the case within the 15m radius of validity.

There are numerous differences in the directional radial velocity suggested by the different models. Initially the VR model has a local minimum at the origin, suggesting that participants move faster as they get further from the aggressor, while the PR model has a local maximum approximately located at the origin, indicating the opposite.

The PR model suggests that there is a negative relationship between radial velocity and age, while the VR model instead suggests that there is a negative relationship between radial velocity and self-assessed exercise level, while also indicating that female participants moved radially slower than male participants.

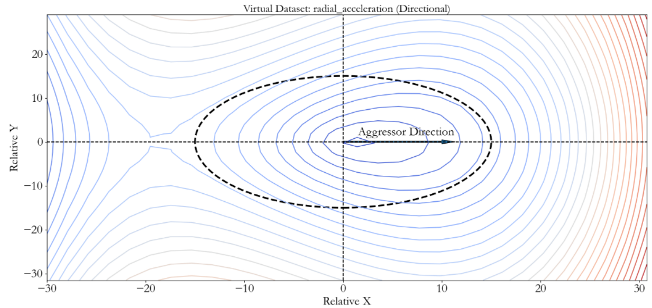
PR Stress: Radial acceleration (Directional, ms^{-2})						VR: Radial acceleration (Directional, ms^{-2})				
										
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	-0.5432	0.0687	-7.9010	2.91e-15***	Constant				
Positional	Distance					Distance	0.0352	0.0065	5.4013	6.67e-08***
	Distance^2					Distance^2				
	X					X	-0.0255	0.0065	-3.9000	9.64e-05***
	X^2					X^2				
	X^3					X^3	0.0001	4.62e-05	2.8306	0.004649**
	Abs(X)					Abs(X)				
	Y					Y				
	Y^2					Y^2				
	Y^3					Y^3				
	Abs(Y)					Abs(Y)				
Demog.	Age					Age				
	Exercise					Exercise				
	F					F				
				Adjusted r^2	0				**Adjusted r^2**	0.001
				F-statistic (p-value)	NA				**F-statistic (p-value)**	14.623 (1.6354e-09***)
				Omnibus test (p-value)	6420.399 (0.000***)				**Omnibus test (p-value)**	3695.218(0.000***)

Table 8.20.: PR Stress and VR Radial acceleration (directional) models

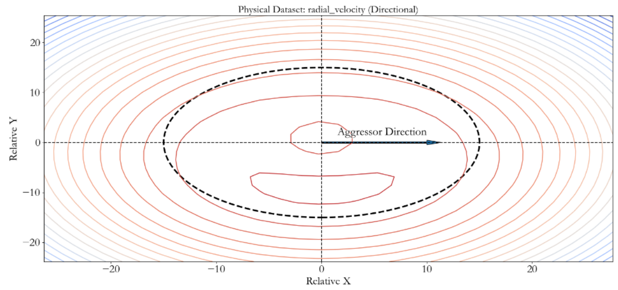
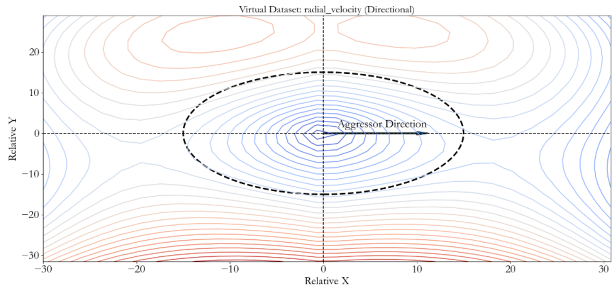
PR Stress: Radial velocity (Directional, ms^{-1})						VR: Radial velocity (Directional, ms^{-1})				
										
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	1.2203	0.1809	6.7500	1.63e-11***	Constant	-0.4462	0.0411	-10.8476	2.31e-27***
Positional	Distance	0.0823	0.0216	3.8090	0.0001***	Distance	0.1157	0.0091	12.7511	3.87e-37***
	Distance^2	-0.0076	0.0013	-5.6307	1.85e-08***	Distance^2	-0.0025	0.0002	-10.8331	2.71e-27***
	X					X	-0.0065	0.0008	-8.6692	4.58e-18***
	X^2					X^2	-0.0032	0.0003	-12.2774	1.46e-34***
	X^3					X^3				
	Abs(X)					Abs(X)	0.0308	0.0054	5.7403	9.55e-09***
	Y	-0.0140	0.0041	-3.4338	0.0006***	Y	0.0198	0.0016	12.4161	2.63e-35***
	Y^2					Y^2	0.0007	0.0003	2.8168	0.0049**
	Y^3					Y^3	-0.0001	1.05e-05	-8.0943	5.98e-16***
	Abs(Y)					Abs(Y)	-0.0444	0.0065	-6.8488	7.59e-12***
Demog.	Age	-0.0339	0.0067	-5.0164	5.38e-07***	Age				
	Exercise					Exercise	-0.0183	0.0045	-4.0467	5.21e-05***
	F					F	-0.0467	0.0116	-4.0121	6.03e-05***
Adjusted r^2				0.012		**Adjusted r^2**			0.028	
F-statistic (p-value)				26.047 (1.735e-21***)		**F-statistic (p-value)**			84.892 (2.488e-173***)	
Omnibus test (p-value)				4303.792 (0.000***)		**Omnibus test (p-value)**			3595.282 (0.000***)	

Table 8.21.: PR Stress and VR Radial velocity (directional) models

Combined analysis

In the results detailed here, both PR and VR datasets are standardised using Equation 5.7, and then combined for a single multiple regression. Similar to the discrete analysis performed earlier in this chapter, further VR-linked predictor variables were included that were only non-zero within the VR dataset.

The same forward-backward stepwise variable selection procedure was used, meaning that only significant predictor variables are displayed. Therefore, any non-VR-linked variable shown below is significant for both datasets. However, any VR-linked variables that are present indicate a significant difference between the datasets. This is crucial to understanding the differences between paradigms.

The final part of this combined analysis will provide a colour-coded table that assesses the overlap in continuous movement responses between paradigms. In order to populate this table, it is necessary to distinguish between VR-linked predictor variables that reduce the dependence on this variable, and those that reverse this dependence. To do this, the individual standardised datasets were modelled, and the resulting predictor variables noted. If the predictor variable coefficient within the PR model had the same sign as that of the VR model variable, then any significant differences (identified within the combined models) were simply an exaggeration or a reduction. However, if the sign of the PR model variable was opposite to that of the VR model variable, then any identified differences were concluded to be a reversal. The individual models are provided in Appendix E.

Acceleration

Table 8.22 shows that there is a positive constant value that is present in both datasets, with an additional positive VR constant variable, implying people accelerate more within VR. Self-assessed exercise level has a positive relationship with acceleration, while age has a negative relationship with acceleration, across both VR and PR paradigms.

There are several differences in the dependence of acceleration on position-based predictor variables. Specifically, the model suggests some unique predictor variables in VR, including in X, X^2, Y, Y^2, distance and distance^2. Further to this, there is an exaggeration in VR of the dependency on abs(X). Female participants are seen to accelerate more than male participants in PR experiments, and less than male participants in VR experiments.

Combined data: Standardised acceleration						
Subset		**Variable**	**Coeff.**	**Std error**	**t-value**	**p-value**
Const.	Base	Constant	0.2380	0.0505	4.7090	2.50e-06***
	VR	VR constant	0.9825	0.0454	21.6202	5.03e-103***
Positional	Base	Abs(X)	-0.0102	0.0030	-3.4392	0.0006***
		Y^3	-4.92e-05	9.58e-06	-5.1336	2.86e-07***
		Abs(Y)	-0.0195	0.0028	-6.9902	2.79e-12***
	VR	VR distance	-0.0723	0.0088	-8.2135	2.22e-16***
		VR distance^2	0.0022	0.0002	8.9711	3.07e-19***
		VR X	-0.0178	0.0009	-20.6772	1.90e-94***
		VR X^2	0.0008	0.0003	3.2665	0.0011**
		VR abs(X)	-0.0399	0.0067	-5.9308	3.04e-09***
		VR Y	0.0334	0.0016	21.3961	5.93e-101***
		VR Y^2	0.0014	0.0002	5.7299	1.01e-08***
Demog.	Base	Age	-0.0076	0.0017	-4.5136	6.39e-06***
		Exercise	0.0148	0.0046	3.2089	0.0013**
		F	0.0742	0.0269	2.7581	0.0058**
	VR	VR F	-0.1313	0.0302	-4.3432	1.41e-05***
Adjusted r^2				0.108		
F-statistic (p-value)				321.203 (0.000***)		
Omnibus test (p-value)				20660.306 (0.000***)		

Table 8.22.: Combined standardised acceleration model

Combined data: Standardised velocity						
Subset		**Variable**	**Coeff.**	**Std error**	**t-value**	**p-value**
Const.	Base	Constant	0.6383	0.0469	13.6056	4.68e-42***
	VR	VR constant	0.8850	0.0425	20.8177	1.06e-95***
Positional	Base	Abs(X)	-0.0262	0.0028	-9.2063	3.55e-20***
		Y^2	-0.0028	0.0002	-11.9158	1.12e-32***
		Y 3	-7.31e-05	9.14e-06	-8.0017	1.26e-15***
	VR	VR distance	-0.0347	0.0082	-4.2302	2.34e-05***
		VR distance^2	0.0017	0.0004	3.9478	7.90e-05***
		VR X	-0.0084	0.0017	-5.0260	5.30e-07***
		VR X^3	-6.53e-05	1.15e-05	-5.6693	1.44e-08***
		VR Abs(X)	-0.0543	0.0064	-8.4748	2.44e-17***
		VR Y	0.0459	0.0015	30.7581	3.63e-205***
		VR Y^2	0.0014	0.0005	3.0108	0.0026**
Demog.	Base	Age	-0.0250	0.0016	-15.6047	1.01e-54***
		Exercise	0.0406	0.0044	9.2124	3.35e-20***
	VR	VR F	-0.0924	0.0127	-7.2770	3.48e-13***
Adjusted r^2				0.182		
F-statistic (p-value)				592.084 (0.000***)		
Omnibus test (p-value)				15194.249 (0.000***)		

Table 8.23.: Combined standardised velocity model

Velocity

Both PR and VR paradigms indicate a constant predictor variable for velocity in Table 8.23, while there is a significantly higher constant within VR environments. Additionally, both paradigms show a negative relationship with age, and a positive relationship with self-assessed exercise.

There are multiple predictor variables that are present solely in the VR model, including distance, distance^2, X, X^3, Y and gender. There is also an exaggeration of the dependency on abs(X), and the dependency on Y^2 is only present in the PR model (shown by the VR-linked variable cancelling out any effect).

Tangential acceleration

Both paradigms show a positive constant component (seen in Table 8.24, while this is exaggerated in the VR experiments. Both paradigms also show a significant relationship with X^2 and Y^3, as well as with a relationship with distance, which is exaggerated in the VR experiments. Tangential acceleration is shown to be reduced with age and increased with self-assessed exercise level.

The VR model shows a unique dependence on distance^2, X, abs(X) and Y. Finally, the VR model shows a reversal of the dependence on gender, where female participants accelerate more in PR experiments and male participants accelerate more in VR experiments.

Tangential velocity

The tangential velocity model (Table 8.25) shows that both paradigms have a positive constant term within the regression, although this is significantly higher in the VR model. Both PR and VR experiments see tangential velocity decreasing with distance, and increasing with self-assessed exercise level. Both experimental paradigms also see older participants moving more slowly, however this effect is reduced in VR.

There are significant differences based on positional variables, indicating that the dependence on X^2, abs(X) and Y is only present in VR, and that the dependence on X is reversed in VR. The model suggests that female participants move tangentially slower than male participants within VR experiments only, with no effect based on gender seen in PR experiments.

Radial acceleration

Both datasets produce a positive constant term within the radial acceleration model (Table 8.26), although this is significantly exaggerated within the VR model. Both datasets suggest a negative correlation with Abs(Y), the non-linear positional variables X^2 and Y^3, and with age.

The model suggests that there is a VR specific negative relationship between radial acceleration and gender, as well as X, abs(X) and Y.

Radial velocity

Finally, the model for radial velocity (Table 8.27) shows that both paradigms see a positive constant term within the regression, which is again significantly

Combined data: Standardised tangential acceleration						
Subset		**Variable**	**Coeff.**	**Std error**	**t-value**	**p-value**
Const.	Base	Constant	0.1833	0.0514	3.5667	3.62e-04***
	VR	VR constant	1.0657	0.0483	22.0827	2.29e-107***
Positional	Base	Distance	-0.0218	0.0034	-6.4497	1.13e-10***
		X^2	0.0006	0.0003	2.3185	0.0204*
		Y^3	-3.72e-05	9.40e-06	-3.9507	7.81e-05***
	VR	VR distance	-0.0837	0.0096	-8.6836	3.99e-18***
		VR distance^2	0.0042	0.0004	10.6143	2.79e-26***
		VR X	-0.0152	0.0009	-17.5705	7.90e-69***
		VR abs(X)	-0.0533	0.0040	-13.3323	1.86e-40***
		VR Y	0.0322	0.0016	20.6954	1.31e-94***
Demog.	Base	Age	-0.0068	0.0017	-4.0291	5.61e-05***
		Exercise	0.0225	0.0046	4.8431	1.28e-06***
		F	0.1062	0.0270	3.9364	8.29e-05***
	VR	VR F	-0.1481	0.0303	-4.8851	1.04e-06***
Adjusted r^2				0.103		
F-statistic (p-value)				328.126 (0.000***)		
Omnibus test (p-value)				24767.403 (0.000***)		

Table 8.24.: Combined standardised tangential acceleration model

Combined data: Standardised tangential velocity						
Subset		**Variable**	**Coeff.**	**Std error**	**t-value**	**p-value**
Const.	Base	Constant	8.00e-01	8.53e-02	9.3715	7.54e-21***
	VR	VR constant	0.7276	0.0881	8.2585	1.52e-16***
Positional	Base	Distance	-0.0573	0.0058	-9.9077	4.12e-23***
		Distance^2	0.0014	0.0003	4.6409	3.48e-06***
		X	0.0108	0.0021	5.1848	2.17e-07***
		X^3	-4.10e-05	1.06e-05	-3.8796	1.05e-04***
		Y^3	-5.05e-05	8.88e-06	-5.6903	1.28e-08***
	VR	VR X	-0.0196	0.0020	-9.8614	6.53e-23***
		VR X^2	0.0027	0.0004	7.0718	1.56e-12***
		VR abs(X)	-0.1020	0.0056	-18.3096	1.48e-74***
		VR Y	0.0385	0.0015	26.1793	1.06e-149***
Demog.	Base	Age	-0.0295	0.0031	-9.4387	3.99e-21***
		Exercise	0.0512	0.0046	11.2203	3.61e-29***
	VR	Age	0.0082	0.0038	2.2197	0.0264*
		VR F	-0.0694	0.0128	-5.4149	6.17e-08***
Adjusted r^2				0.168		
F-statistic (p-value)				535.752 (0.000***)		
Omnibus test (p-value)				16936.736 (0.000***)		

Table 8.25.: Combined standardised tangential velocity model

Combined data: Standardised radial acceleration						
Subset		**Variable**	**Coeff.**	**Std error**	**t-value**	**p-value**
Const.	Base	Constant	0.2852	0.0441	6.4606	1.06e-10***
	VR	VR constant	0.4583	0.0230	19.9184	8.13e-88***
Positional	Base	X^2	-0.0007	0.0002	-3.4510	0.0006***
		Y^3	-0.0001	8.55e-06	-8.7611	2.01e-18***
		Abs(Y)	-0.0169	0.0017	-9.9846	1.91e-23***
	VR	VR X	-0.0176	0.0009	-20.4694	1.31e-92***
		VR abs(X)	-0.0265	0.0033	-8.0681	7.35e-16***
		VR Y	0.0291	0.0015	19.1579	2.06e-81***
Demog.	Base	Age	-0.0079	0.0017	-4.6384	3.52e-06***
	VR	VR F	-0.0631	0.0129	-4.8793	1.07e-06***
Adjusted r^2				0.059		
F-statistic (p-value)				257.275 (0.000***)		
Omnibus test (p-value)				25877.904 (0.000***)		

Table 8.26.: Combined standardised radial acceleration model

Combined data: Standardised radial velocity						
Subset		**Variable**	**Coeff.**	**Std error**	**t-value**	**p-value**
Const.	Base	Constant	0.7419	0.0450	16.4741	9.28e-61***
	VR	VR constant	0.5713	0.0318	17.9498	9.70e-72***
Positional	Base	Distance	-0.0182	0.0034	-5.2829	1.28e-07***
		Y^3	-0.0001	9.00e-06	-8.2950	1.12e-16***
		Abs(Y)	-0.0178	0.0036	-4.8817	1.06e-06***
	VR	VR X	-0.0061	0.0017	-3.4869	4.89e-04***
		VR X 3	-0.0001	1.20e-05	-6.3395	2.33e-10***
		VR abs(X)	-0.0441	0.0056	-7.8998	2.87e-15***
		VR Y	0.0415	0.0015	27.6734	7.32e-167***
		VR Y 2	0.0005	0.0002	2.5215	0.0117*
Demog.	Base	Age	-0.0225	0.0017	-13.5648	8.15e-42***
	VR	VR exercise	0.0204	0.0052	3.9380	8.23e-05***
		VR F	-0.0911	0.0133	-6.8529	7.35e-12***
Adjusted r^2				0.123		
F-statistic (p-value)				399.230 (0.000***)		
Omnibus test (p-value)				22206.357 (0.000***)		

Table 8.27.: Combined standardised radial velocity model

Subset	Predictor variable	Acceleration	Velocity	Tangential acceleration	Tangential velocity	Radial acceleration	Radial velocity
Const.	Constant	Both- VR exagg.	Both- VR exagg.	Both- VR exagg.	Both- VR exagg.	Both- VR exagg	Both- VR exagg.
Positional	X	VR only	VR only	VR only	Both- VR reversed	VR only	VR only
	X^2	VR only	Neither	Both	VR only	Both	Neither
	X^3	Neither	VR only	Neither	Both	Neither	VR only
	Abs(X)	Both- VR exagg.	Both- VR exagg.	VR only	VR only	VR only	VR only
	Y	VR only	VR only	VR only	VR only	VR only	VR only
	Y^2	VR only	PR only	Neither	Neither	Neither	VR only
	Y^3	Both	Both	Both	Both	Both	Both
	Abs(Y)	Both	Neither	Neither	Neither	Both	Both
	Distance	VR only	VR only	Both- VR exagg.	Both	Neither	Both
	Distance^2	VR only	VR only	VR only	Both	Neither	Neither
Demog.	Age	Both	Both	Both	Both- VR reduced	Both	Both
	F	Both- VR reversed	VR only	Both- VR reversed	VR only	VR only	VR only
	Exercise	Both	Both	Both	Both	Neither	VR only

Table 8.28.: Summary of commonalities and differences between predictors of movement types within PR and VR datasets.

higher in the VR model. Both models also see a negative relationship with distance, Y^3, abs(Y) and age.

The VR model suggests a unique relationship between radial velocity and numerous predictors, including X, X^3, abs(X), Y^2, self-assessed exercise level and gender.

8.3.8. Continuous analysis summary

Table 8.28 combines all of the models above into a single colour-coded chart, showing the level of overlap between the paradigms. It uses the same colour-coding system as detailed above in Table 8.1.

This is a simple method of summarising the quantitative information shown in subsubsection 8.3.7, by movement type and by predictor variable subset. The numerous "VR only" boxes within the positional subset are likely a product of the lack of noise in the VR measurements, which resulted in higher explanatory power for the VR models.

When comparing the positional elements within this table, it can be seen that there are significant overlaps between the VR and PR experimental paradigms, while there is only a singular reversed behavioural response within the X variable. The demographic predictor subset shows overlap in both age and exercise variables, while there are differences in gender-specific responses across all responses.

It can be concluded from this table that there are significant overlaps in the data observed, and that the VR environments utilised in these experiments are a reasonable approximation of real-world continuous movement. However, it is also concluded that there is a significant difference in response type based on gender between the experimental paradigms, and all future work should be aware of this difference when inferring any relationships.

Overall, it is seen that there is a large degree in overlap between the responses between VR and PR experimental paradigms. Examples of this are in the local maxima for almost all models that occur approximately at the origin, and the levels of symmetry observed within these models.

Improving upon this analysis will require several further elements, including:

- More data
- More predictor variables, including nearby neighbour participant states (including speed, acceleration and direction)
- A more sophisticated modelling approach, incorporating time-series dependence (see Section 10.4)

8.4. Discussion and conclusion

This chapter has analysed the external data obtained during the two experimental paradigms (PR and VR). As a result of this analysis, there are several key results detailed below.

1) Initial datasets and analysis techniques for movement behaviour in response to a hostile aggressor.

Previous work has primarily modelled assumed movement responses to hostile aggressors, and very little experimental work or analysis of real-world data is available to provide insight into the validity of these assumptions.

However, based on the analysis performed in Chapter 6 and Chapter 7, it is possible to conclude that these experiments did produce a state of negative emotions in the participants, as a direct result of the introduction of the hostile aggressor. Therefore, the external datasets produced by the experiments and the analysis performed in this chapter provide some of the first insights into the movement responses of pedestrians as a result of a hostile aggressor.

These insights are split between discrete choices and continuous responses, and are based on external data (which could be inferred or measured from typical data sources, such as CCTV), and demographic data (which provides a useful way of improving future modelling capabilities).

2) Identification of important features within these datasets.

By performing both discrete and continuous analysis, this chapter has identified the important factors in the movement of participants in response to hostile aggressors.

In the discrete analysis, it was seen that there was a low initial probability of a participant running from the aggressor, shown by the strongly significant negative constant coefficients in Table 8.5 and Table 8.7. However, strongly significant positive coefficients were also seen in the Number Reacted and Number Caught variables, indicating that participants were influenced to a large extent by social interactions. Interestingly, no significance was observed for either the 'Within FOV', or 'Distance to aggressor' predictors, suggesting that participant responses were far more influenced by their neighbours' responses than by the aggressor. It is worth noting, however, that the presence of the aggressor was required to produce the psychological and physiological responses, therefore placing the participants into a state where they are strongly influenced by their neighbours.

The continuous data analysis initially performed a qualitative approach for comparing the movement speed of participants with respect to the perspective of the aggressor. This analysis showed analogous results for VR and PR experimental paradigms. The initial result suggested that the movement speed of the participants was related to the distance from the aggressor, with 'running' datapoints being clustered closer to the aggressor than 'walking' datapoints. Further to this, an asymmetrical response was seen, with 'running' datapoints presenting more often in front of the aggressor than behind.

This analysis was extended with a forward-backward stepwise selection process on a range of external and demographic predictor variables, providing relationships on movement speed and acceleration in terms of absolute, radial, and tangential components. The regression model results showed significant relationships between movement responses and positional and demographic predictor variables. There were also qualitatively similar results between VR and PR experimental responses. Owing to the non-linearity of the residuals, the same analysis was then performed on a log-transformed dataset, providing similar levels of significance and qualitative similarity, but with higher model explanatory power. However, the residuals were still significantly non-linear, suggesting that a different approach could be used. These models can be found in Appendix E.

While these models are only valid for a small subset of scenarios (i.e. indoors and within 20x20m areas), and only for nearby responses (e.g. within a 15m radius of the aggressor), there is a notable overlap with previous work that examines real-world data. For example, Li et al. [76] modelled threat levels of different attack vectors based on distance, nearby obstructions, diffusion coefficients, and risk perception. There are obvious overlaps between this approach and the logistic model governing discrete choices, and the continuous model suggesting reductions in velocity and acceleration with increasing distance from the aggressor.

3) Significant similarity in discrete activity choices between PR and VR experimental paradigms.

The discrete data analysis specifically provides some of the first quantitative results that characterise the fight-flight-freeze response as a function of both external data and demographic data.

Using this assessment methodology, it was revealed that the discrete datasets were very similar between paradigms. The only significant differences seen (Table 8.9) were:

- An exaggerated dependence on social observations in VR
- A removed dependence on exercise level in VR
- A reversed gender effect in VR

This provides strong evidence for the ecological validity of VR as an experimental paradigm. However, it is also noted that the experimental procedure led participants to believe that the other avatars were controlled by human participants, and that they would meet each other later. This likely led to an increased level of social dependence within the environment, which may not have occurred otherwise.

4) Assessment of validity for VR for both continuous data collection methods.

The PR Stress and VR datasets were then standardised and combined for continuous analysis, in order to find any significant differences between the experimental paradigms. These differences were assessed by creating a model based on basic variables, then adding in VR-linked variables that are only non-zero in VR datasets.

This methodology also revealed that the continuous datasets had many significant similarities between paradigms. With the majority of the differences resulting from predictor variables being significant only within the VR paradigm. It is suggested that this is at least partially a result of the noise within the PR paradigm. Without this noise, therefore, there is significant overlap between VR and PR-generated datasets.

One major difference is seen in the responses of different gender participants, which was similarly seen in the discrete analysis. This should be considered in all VR experiments in future.

8.4.1. Conclusion

This PhD concludes that VR can be used to obtain discrete data that will accurately mimic discrete data from PR experiments. However, when obtaining this data, the experimental team should be aware of potential confounding effects from increased social dependence and from demographic effects.

Additionally, this PhD concludes that VR can be used to obtain continuous data which will mimic continuous data from PR experiments. However, as with discrete data, the experimental team should be aware of the different demographic effects seen within VR. This PhD only used a single locomotion technique (armswinging), and therefore also concludes that future research should investigate the validity of different locomotion techniques, including omnidirectional treadmills and free movement.

8.5. Chapter summary

This chapter provides insight into the individual and group movement responses to hostile aggressors, from both a discrete and a continuous perspective. These results can be used for pedestrian dynamics models, as well as for a host of other applications, including improving crowd management and developing theoretical elements of group responses to emergencies.

This chapter also provides quantitative models from two different data collection methodologies – PR and VR – providing a comparison between the two experimental paradigms. The result of this comparison is to show that PR and VR experimental paradigms are both well suited to obtaining discrete data information for behavioural responses to hostile aggressors. Furthermore, both paradigms can be used to obtain continuous data information for these responses, but the locomotion technique in VR must be continuous and calibrated before any generated data is used for further purposes.

The next chapter will detail the development of, and results from, the participant survey on recall quality to provide insight into how reliable testimony is as a method of generating data on responses to hostile emergencies.

9. Results: Recall Quality

Chapter 6 and Chapter 7 focused on the analysis of the psychological and physiological data obtained from the experiments performed in December 2018 (PR) and January-February 2020 (VR), establishing the ecological validity of the experiment. Chapter 8 focused on investigating the movement responses of participants within these experiments, providing some of the first datasets on human behaviour in response to hostile aggressors. This movement investigation also compared the PR and VR paradigms as data-generating methods.

Subsection 4.7.4 detailed the survey that was performed on the participants of the PR and VR experiments, aiming to identify their ability to remember their experiences during the experiments.

This chapter, outlined in Figure 9.1, will first detail the development of the survey in which the participants' memory was assessed. This includes the domain definitions, the scoring of each individual's recall, and the survey distribution methodology. Next, this chapter will assess the results of the surveys from the Control, PR-Stress and VR experiments with respect to these domains, including: domain scores, confidence scores, confidence-weighted domain scores, regression models and confabulation proportions. Finally, this chapter will discuss the key results from this survey, any limitations on the data obtained, and how these results can be used in future.

This results in this chapter will provide insight into the reliability of individual recall as a datasource, using surveys as a data-gathering method for human behaviour in emergencies. This reliability will be quantified using a number of numerical criteria, primarily comparing the experimental branches (i.e.: PR-Stress and VR participants) against the Control experiments.

Throughout this chapter significance levels will be reported using the following codes: *$p<0.05$, **$p<0.01$, ***$p<0.001$. Since many of the datasets used in this chapter were found to be non-normally distributed (see Section 5.3 for further details), the reported significance codes are based on the Mann-Whitney U test.

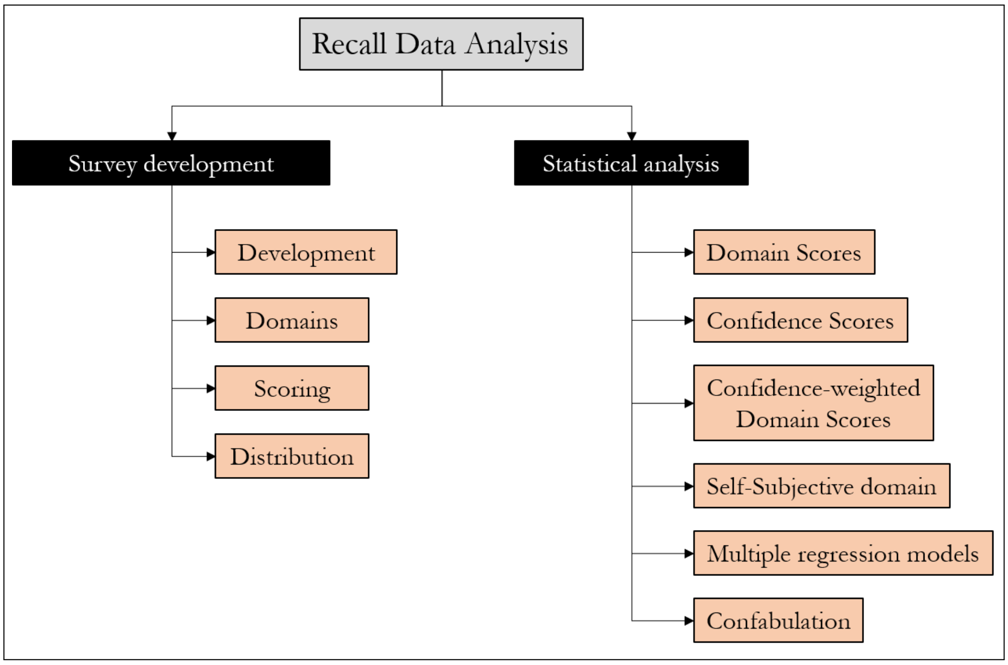


Figure 9.1.: Chapter outline

9.1. Survey development

As detailed in Section 2.2, the majority of data surrounding human behavioural responses has come from interviews. Interviews and surveys are a commonly used data-gathering tool that suffer from several limitations for the purposes of pedestrian dynamics modelling. Interview and surveys are primarily qualitative, leading to a lack of the quantitative data surrounding movement responses in emergencies that is required for model inputs, and for prediction of responses to hypothetical scenarios. It is also broadly accepted that quality of an individual's recollection of an event deteriorates over time, but little quantitative data is available on the extent of this deterioration, or whether it changes depending on context.

The survey was developed in conjunction with psychologists at University College London, utilising best practices for survey design, and aiming to quantitatively answer some of the outstanding questions surrounding recollection. The development process was iterative, with key literature resources (see Subsection 4.7.4 for more information) used to inform the question design.

The survey utilised a mixture of free choice and forced choice styles, including qualitative and quantitative questions, and consequently the analysis required a range of techniques, such as interrater reliability and regression modelling. The results of this survey would therefore provide insight into the reliability of recall across multiple contexts, including examining the level of confabulation (or: memory generation).

Another element of the survey design was to mitigate the potential response bias by providing participants with both positive and negative forms of questions. This was implemented to prevent the responses either being biased towards agreeing or disagreeing with the framing of the question. In order to achieve this, the online form utilised two separate branches, each containing the same questions, but framed either as positive or negative statements (where applicable), and participants would be allocated randomly to these branches at the beginning of the online survey. For example, a positive question form might be 'I tried to evade the aggressor', while the negative form would be 'I did not try to evade the aggressor'.

The final consideration of the design of this survey was to ensure as high a completion rate as possible. As the potential sample was limited to the participants of the PR (80) and VR (55) experiments, there was a requirement for a majority of the participants to respond, as otherwise there would have been insufficient data to generalise any results. Therefore, the survey was designed to last approximately ten minutes in total. It was sent to the participants of both experimental paradigms three months after the participants' individual sessions, through a Qualtrics online form. As the VR experiments were performed approximately a year after the PR experiments, the survey was updated prior to distribution to the VR experiment participants, to include questions relevant to the VR environment. The full surveys, provided in Appendix F, achieved a relatively high completion rate, with 55/80 PR participants and 45/55 VR participants completing the survey.

9.2. Impact of time

As noted earlier, recall of an event can deteriorate over time (see Section 4.7). This study was not designed to investigate *how* this deterioration occurs over time, but rather to quantify this deterioration *at a specific moment* in time. It was therefore necessary to ensure that all participants completed the survey after approximately the same period following their respective experiments.

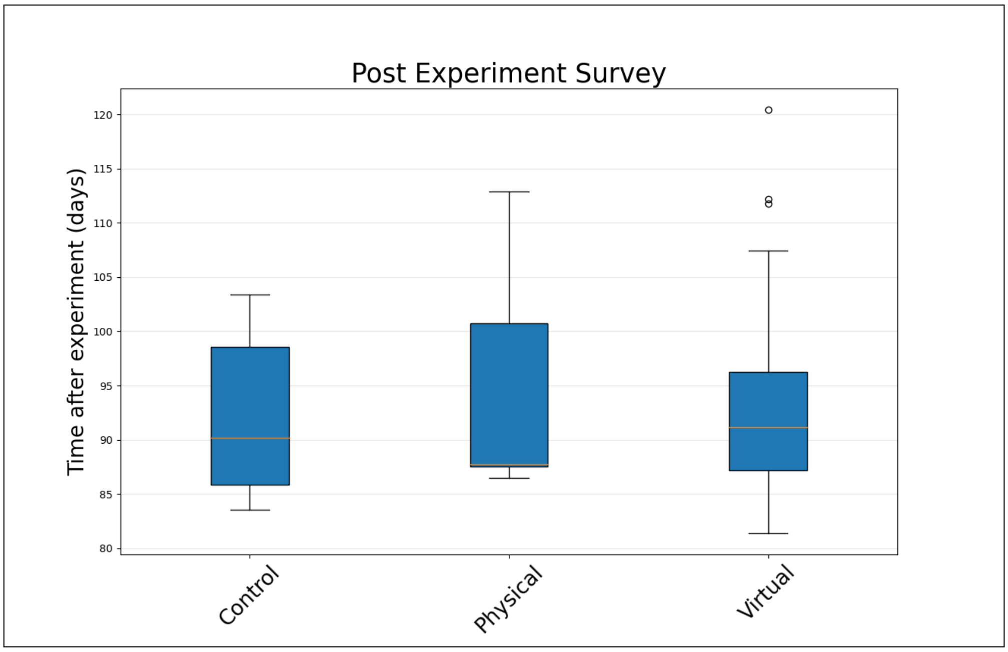


Figure 9.2.: Duration of time after experiments for survey completion

The implication is that the survey questions should be time-invariant (i.e.: they are equally applicable at all moments in time after the experiment), whereas the results of the survey might not be time-invariant (i.e.: participant recall quality might vary from three months to three years, for example).

Each participant was contacted three months after their experiment and asked to complete this survey. Any participants who failed to complete the survey were then contacted again one week later, in an effort to increase the response rate. The time between experiment and survey completion was recorded and is shown below in Figure 9.2. No significant differences were found between any experimental paradigms. The VR participants completed their surveys during April-May 2020, during the first nationwide UK COVID-19 lockdown. It is not possible to ascertain if this had an impact on the results.

9.3. Scores

This chapter will quantitatively assess the deterioration of an individual's recollection of their experiment, and will compare this deterioration between experimental paradigms. In order to do this, it is necessary to be able to measure the accuracy of an individual's responses, specifically with respect to the record of what actually occurred, the 'ground truth'. Consequently it was decided that this study would quantify any deviation from this ground truth. If a participant gave a correct answer, they received a score of 0. However, if they gave an incorrect answer, they received a score that was dependent on their response. Each question had different individual rating scales, depending on the question itself. This flexibility provided the option to quantify correct and incorrect answers, or the option to quantify over- or underestimation of specific recalled values.

For example, if a participant was asked 'What colour was the aggressor's hair?', their response would be marked on a binary scale, where 0 corresponds to a correct answer, and 1 corresponds to an incorrect answer. In contrast to this, any responses to the question 'On a scale of 1-7, how much did you feel scared?' would be marked on a relative scale, as these responses were able to under- or overestimate the ground truth value that was obtained at the time of the experiment. If the participant actually answered 5 at the time of the experiment (the ground truth), but recalled a 4, they would receive a raw score of -1. Conversely, if they recalled a 7, they would receive a raw score of 2.

In order to combine these scores in a meaningful way, as well as scores from multiple questions, a normalisation procedure was adopted that scaled the scores by the maximum of the absolute raw scores.

$$Raw\ Score = deviation\ from\ truth\ (quantitative,\ or\ coded\ qualitative) \tag{9.1}$$

$$Normalised\ score,\ Q_i = Abs\left(\frac{Raw\ Score_i}{Max\,(\ Abs(Raw\ scores_i))}\right) \tag{9.2}$$

In order to accommodate all types of question scoring, the absolute value of the raw scores was used, to quantify the deviation from ground truth. However, a sub-analysis was also performed, utilising all possible relative scores (taking a value between -1 and 1) combined into a single domain. This is explained further below.

Domain	Details
Overall	Recall of the entire scenario
Descriptive	Recall of qualitative descriptions of surroundings and people
Self-behaviour	Recall of own behaviour and choices
Self-subjective	Recall of own recorded subjective responses
Aggressor-focused	Recall of aggressor-related information
Neighbour-focused	Recall of information relating to other participants
Spatial	Recall of spatially orientated information

Table 9.1.: Recall domains

Narrative Element	Description	Calculation
Confabulation	How many times a memory was created for each correct answer	$$C = \frac{FP}{TP + TN}$$ Where: FP= False positives, TP=True positives, TN=True negatives

Table 9.2.: Confabulation calculation

9.4. Domains

This PhD investigated participant recall accuracy within several domains, which were selected to encompass the broad range of relevant pedestrian dynamics topics (see Chapter 2 for more information).

Furthermore, the questions also had an associated confidence rating, from 0-100 (low to high), allowing participants to indicate their level of certainty surrounding their responses. This confidence rating was not used for the Self-subjective domain, as the questions for this domain were combined to make the composite scores for PANAS, STAI-S and group identification.

The surveys also investigated the potential for confabulation among the participants. This was quantified as the proportion of created memories to correct memories within the True/False questions.

As survey duration should be optimised to balance a high enough completion rate with sufficient obtained data, the number of questions in the survey needs to be controlled. The number of questions in each domain is detailed below in Table 9.3, and the full surveys are provided in Appendix F. The survey was adapted after being distributed to the PR participants, with more questions

Domain	PR Stress and Control survey questions	VR survey questions
Overall	28	35
Descriptive	6	7
Self-subjective	9	10
Self-behaviour	6	8
Aggressor-focused	13	15
Neighbour-focused	9	10
Spatial	2	8

Table 9.3.: Number of questions per domain

Cross tabulation		**Rater 1 Classification**		
		1	0	Total
Rater 2 Classification	1	97	1	98
	0	10	43	53
	Total	107	44	151

Table 9.4.: Interrater reliability scores

that were specific to the VR experiment, leading to the differences shown in the table.

To accommodate the limitation on the survey length while still obtaining sufficient data, the survey questions were allocated to multiple domains. Each question could contribute to the domain score of up to a maximum of four separate domains, including the Overall domain which was used as a summary measure of participant recall quality. For example, the question 'Where were you when you noticed the aggressor?' would fall into Spatial and Aggressor domains, as well as the Overall domain.

9.5. Data analysis

For the subjective questions an inter-reliability rating process was performed, the scores of which are presented in Table 9.4, producing a Cohen's $\kappa = 0.834$. The full coding methodology is provided alongside the list of questions in Appendix F.

9.5.1. Score calculation

The domain scores are initially taken as an absolute value, providing a comparison between all domains. Further analysis was then performed on the Self-subjective domain, which was the only domain that included signed-value responses.

To create the domain scores, the raw responses were all normalised, leading to a score between 0 to 1 for each question, which allowed the scores from different questions to be averaged using an arithmetic mean to produce a single domain score. The normalised scores were calculated using the following procedure:

$$j^{th}\ Domain\ score = mean(Q_i) \forall i \in j \tag{9.3}$$

$$j^{th}\ Domain\ Confidence\ score = median(C_i) \forall i \in j \tag{9.4}$$

Each domain score was then recalculated using the confidence scores that participants had recorded for each question. The confidence-weighted domain scores were calculated using the following formula:

$$Confidence - weighted\ normalised\ score = Q_i C_i \forall i \in j \tag{9.5}$$

$$j^{th}\ Confidence - weighted\ domain\ score = mean(Q_i C_i) \forall i \in j \tag{9.6}$$

This confidence-weighting procedure maintains the assessment of deviation from ground truth. If, for example, a participant produces a wrong answer (i.e. obtaining a high Q_i score), but has low confidence in their response (i.e. $C_i = 0$), then the confidence-weighted score will be reduced accordingly, resulting in an adapted score that can then be analysed in an identical manner to the base domain scores.

Finally, all results were then analysed using multiple regression models to see if there was any relationship between recall quality and participant demographic, self-assessed emotional data, and physiological responses, providing insight into the reliability of memory as a function of the participant's internal state.

9.6. Domain scores

The average participant scores in each domain are shown graphically below in Figure 9.3, and tabulated in Table 9.5, with all differences between the paradigms are shown in Table 9.6.

Significant differences were observed between Control and VR participants for several domains, with a higher quality recall for VR participants in Overall, Descriptive and Neighbour domains. Further to this, there is a marginal significance ($p<0.1$) in the difference in Self-Subjective domain, with Control participants responding more accurately. Finally, the quality of Self-behavioural recall appears to deteriorate significantly in VR experiments when compared to Control; however, given the relatively low significance level ($p<0.05$), it is possible that this is an artefact of the study.

Differences were also observed between PR and VR participants, with the VR participants recalling more accurately within Descriptive and Neighbour domains. There is also a marginally significant ($p<0.1$) difference present in the Overall and Self-Behaviour domains, again with VR participants recalling more accurately in Overall and less accurately in Self-Behaviour.

Control and PR participants only differed significantly in the Neighbour domain, with PR participants recalling this information more accurately.

Average absolute normalised score per domain								
Paradigm	**Metric**	**Overall**	**Descriptive**	**Self-subjective**	**Self-behaviour**	**Aggressor**	**Neighbour**	**Spatial**
Control	Mean	0.236	0.281	0.188	0.096	N/A	0.305	0.184
	SD	0.078	0.158	0.072	0.128	N/A	0.186	0.248
	Min	0.138	0.000	0.110	0.000	N/A	0.083	0.000
	Max	0.405	0.500	0.368	0.333	N/A	0.710	0.500
PR Stress	Mean	0.205	0.269	0.184	0.111	0.162	0.207	0.278
	SD	0.075	0.151	0.075	0.144	0.141	0.124	0.326
	Min	0.066	0.167	0.000	0.000	0.000	0.049	0.000
	Max	0.415	0.667	0.309	0.667	0.615	0.503	1.000
VR	Mean	0.182	0.189	0.207	0.160	0.155	0.140	0.182
	SD	0.047	0.117	0.061	0.125	0.080	0.077	0.141
	Min	0.108	0.000	0.000	0.000	0.000	0.011	0.000
	Max	0.293	0.429	0.322	0.500	0.400	0.309	0.500

Table 9.5.: Domain scores

Mann Whitney test results on domain scores								
Paradigm	**Metric**	**Overall**	**Descriptive**	**Self-subjective**	**Self-behaviour**	**Aggressor**	**Neighbour**	**Spatial**
PR-VR	Statistic	665	370	709.5	668	771	587	809.5
	p-value	0.064†	5.51e-06***	0.135	0.062†	0.297	0.012*	0.431
Control-VR	Statistic	256	198.5	323	326	N/A	172	365.5
	p-value	0.005**	0.00018***	0.051†	0.0495*	N/A	6.81e-05***	0.145
Control-PR	Statistic	274	321	309	324.5	N/A	226.5	296
	p-value	0.116	0.343	0.283	0.369	N/A	0.021*	0.179

Table 9.6.: Differences between experimental paradigm domain scores

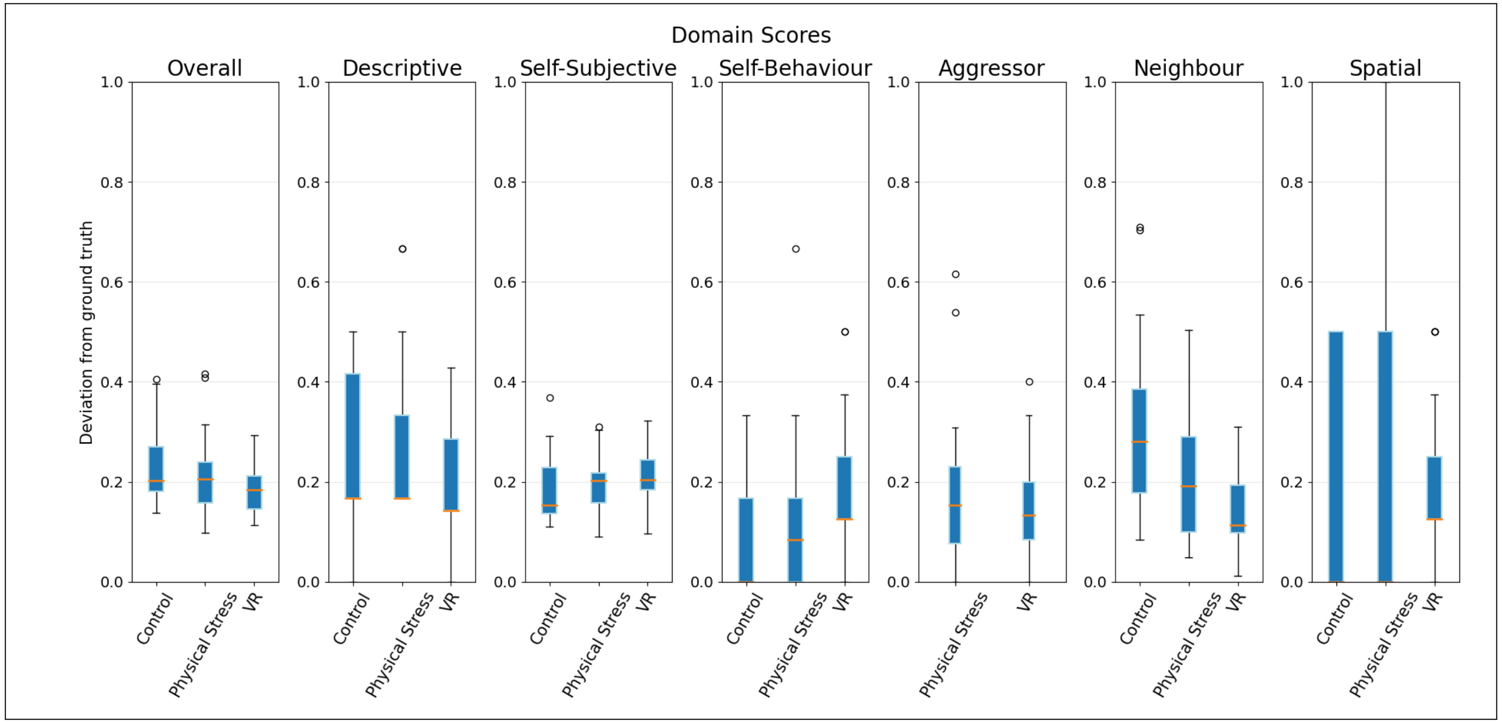


Figure 9.3.: Domain scores

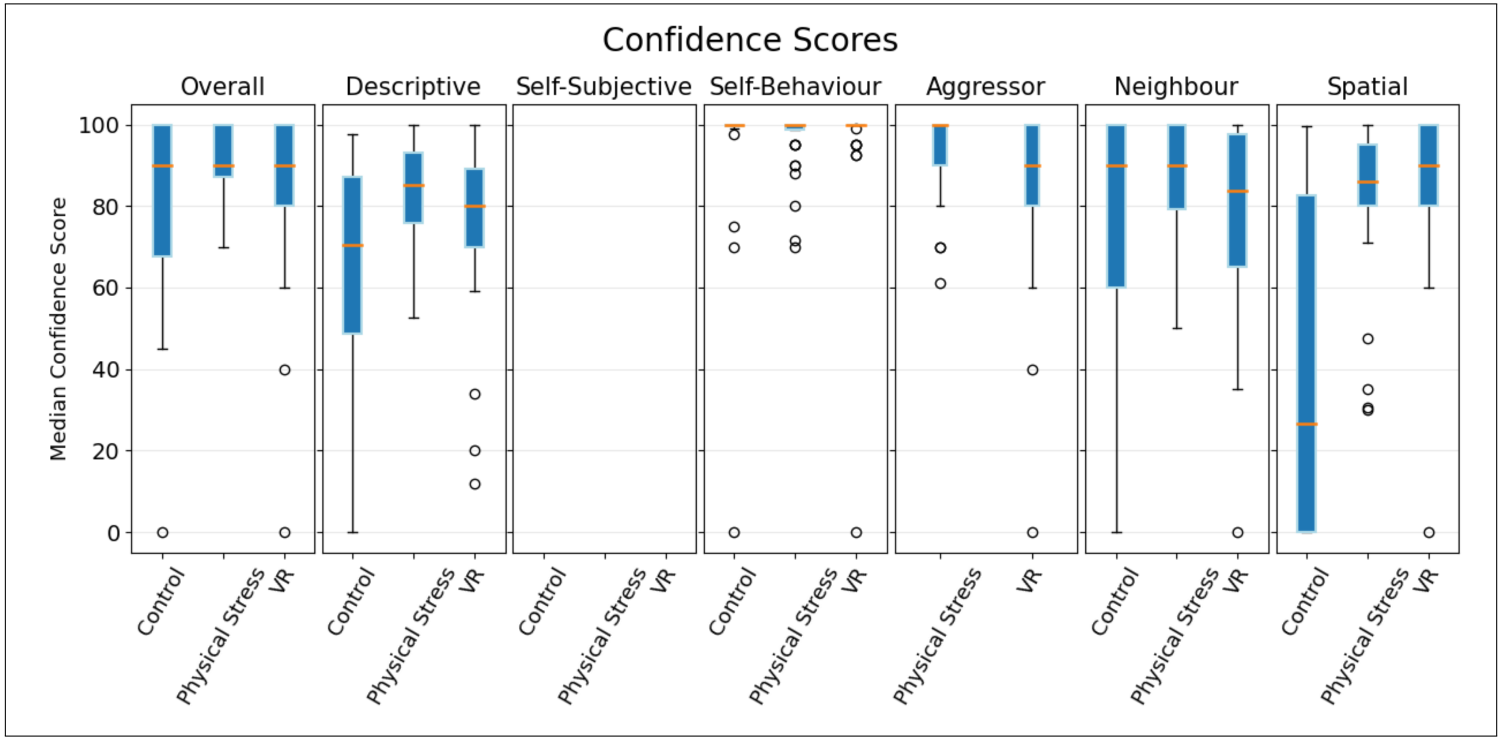


Figure 9.4.: Domain confidence scores

Mann Whitney test results on median confidence scores								
Paradigm	**Metric**	**Overall**	**Descriptive**	**Self-subjective**	**Self-behaviour**	**Aggressor**	**Neighbour**	**Spatial**
PR-VR	Statistic	765	576	N/A	735	618.5	704.5	677
	p-value	0.275	0.009**	N/A	0.106	0.0199*	0.123	0.0787†
Control-VR	Statistic	426.5	358.5	N/A	386	N/A	402.5	160.5
	p-value	0.441	0.130	N/A	0.141	N/A	0.310	2.97e-05***
Control-PR	Statistic	315.5	209	N/A	335.5	N/A	326.5	155
	p-value	0.318	0.009**	N/A	0.445	N/A	0.393	0.000467***

Table 9.7.: Differences between experimental paradigm confidence scores

9.6.1. Confidence scores

These questions also asked the participants their confidence in their responses. An analysis was performed to understand the individual participants confidence in their recall, independent of the accuracy. For each domain, each participant's median confidence value was recorded, to minimise any impact of outliers (such as a confidence rating of 0). The results are shown in Figure 9.4 and Table 9.7. As there were no confidence ratings for the Self-Subjective domain, there are no values for this domain within this figure, however it is shown for consistency.

The participants had higher confidence in their recall of their own behaviour than any other domain, and that this is independent of experimental paradigm. Both VR and PR Stress participants had higher confidence in their spatial recall than the Control participants, and PR Stress participants had the highest confidence in their Descriptive recall. Finally, the PR Stress participants had significantly higher confidence than VR participants regarding their recall of the aggressor. However, when combining the confidence scores into the Overall domain, no significant differences are seen between paradigms. This is likely a result of the combination of the separate domains.

9.6.2. Confidence-weighted results

The confidence-weighted domain scores are shown below in Figure 9.5, with the scores and differences tabulated in Table 9.8 and Table 9.9 respectively. As the confidence scores were not used for questions within the Self-subjective domain, there is no corresponding confidence-weighted domain score, but the axis is included for consistency.

The results are very similar to the base domain scores, replicating the significant differences initially seen between Control and VR participants in the base domains. That is, there is a higher quality recall for VR participants in Overall, Descriptive and Neighbour domains, and a higher quality recall for Control participants in Self-Behaviour.

The same differences are also observed again between PR and VR participants, with the VR participants recalling more accurately within Overall, Descriptive, and Neighbour domains, and less accurately in Self-Behaviour. The only change observed when incorporating confidence weighting is the level of significance, as the differences in Overall and Self-behaviour recall are now

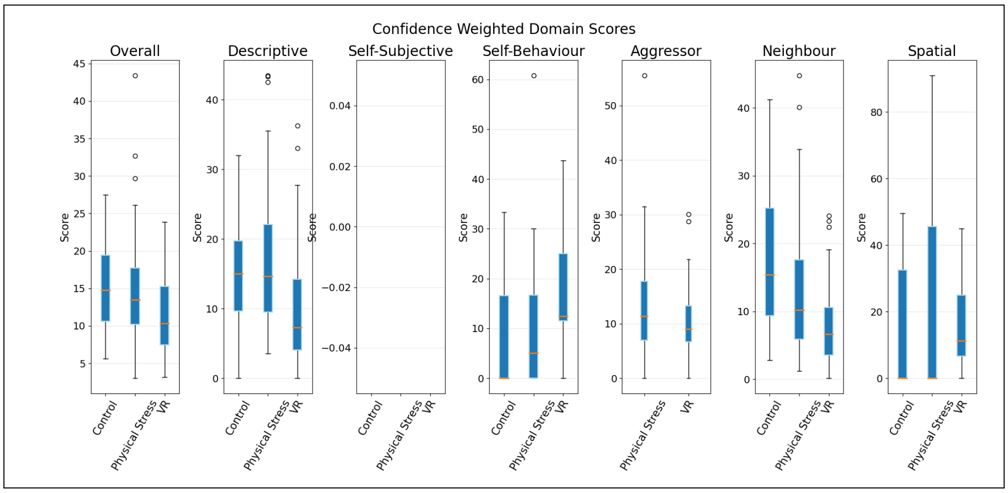


Figure 9.5.: Confidence weighted domain scores

significant (p<0.05).

The significance level of the difference between Control and PR participants in the Neighbour domain has reduced to a marginal level (p<0.1), with PR participants recalling this information more accurately.

Average absolute normalised confidence-weighted score per domain								
Paradigm	**Metric**	**Overall**	**Descriptive**	**Self- subjective**	**Self-behaviour**	**Aggressor**	**Neighbour**	**Spatial**
Control	Mean	15.026	15.474	N/A	7.088	N/A	17.617	14.474
	SD	6.137	8.372	N/A	10.194	N/A	9.934	20.166
	Min	5.645	0.000	N/A	0.000	N/A	2.806	0.000
	Max	27.491	32.000	N/A	33.333	N/A	41.259	49.500
PR Stress	Mean	14.719	16.935	N/A	9.745	12.487	13.872	23.486
	SD	8.623	11.280	N/A	12.888	11.330	10.669	28.346
	Min	3.023	3.500	N/A	0.000	0.000	1.216	0.000
	Max	43.434	43.500	N/A	60.833	55.538	44.806	91.000
VR	Mean	11.437	10.099	N/A	15.136	10.341	8.110	14.345
	SD	5.293	8.988	N/A	11.601	6.413	6.179	11.826
	Min	3.132	0.000	N/A	0.000	0.000	0.161	0.000
	Max	23.844	36.286	N/A	43.750	30.067	24.099	45.000

Table 9.8.: Confidence-weighted domain scores

Mann Whitney test results on confidence-weighted domain scores								
Paradigm	**Metric**	**Overall**	**Descriptive**	**Self- subjective**	**Self-behaviour**	**Aggressor**	**Neighbour**	**Spatial**
PR-VR	Statistic	647	497.5	N/A	608	747.5	541	805
	p-value	0.046*	0.001**	N/A	0.017*	0.227	0.004**	0.4151
Control-VR	Statistic	292	262.5	N/A	283	N/A	179	362.5
	p-value	0.019*	0.006**	N/A	0.011*	N/A	0.0001***	0.139
Control-PR	Statistic	309	341	N/A	314	N/A	260	284.5
	p-value	0.283	0.497	N/A	0.297	N/A	0.075†	0.133

Table 9.9.: Differences between experimental paradigm confidence-weighted domain scores

Mann Whitney U test		3. Self-subjective
PR-VR	Statistic	-1.644
	p-value	0.104
Control-VR	Statistic	-2.908
	p-value	0.0058**
Control-PR	Statistic	1.425
	p-value	0.162

Table 9.11.: Differences in signed subjective scores

9.6.3. Signed domain scores

The only domain that included questions with signed responses was the Self-Subjective domain. Here, participants were asked to recall their responses to the PANAS and STAI-S post-experiment surveys, as well as their level of group identification. The domain results including the sign of the domain scores are shown in Table 9.10, and the respective differences are tabulated in Table 9.11.

Signed domain scores: Constituent questions

This domain consists of the participants' recall of their responses to the PANAS, STAI-S and social response questionnaires that were given immediately after their experiment. The scores from these individual elements were then investigated, identifying which components were more reliable when examining an individual's recall. This investigation also included an assessment of the percentage change between the individual's recall and their initial scores. This provides another assessment of the recall quality which is only applicable to the quantitatively assessed questions present within this domain.

This investigation, detailed in Table 9.12, shows that the PANAS deteriorates less over time than the STAI-S – both in terms of the average percentage change between actual and recalled values, and the normalised scores – but that the STAI deteriorates more consistently across paradigms.

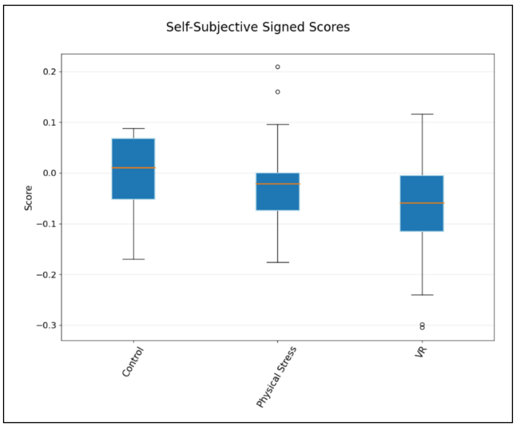


Figure 9.6.: Signed subjective scores

Average signed normalised score		
Paradigm	**Parameter**	**3: Self-subjective**
Control	Mean	0.001
	SD	0.078
	Min	-0.170
	Max	0.088
PR	Mean	-0.032
	SD	0.085
	Min	-0.176
	Max	0.209
VR	Mean	-0.066
	SD	0.097
	Min	-0.304
	Max	0.116

Table 9.10.: Signed subjective scores

Paradigm	PR			Control			VR		
Parameter	**Raw**	**%**	**Normalised**	**Raw**	**%**	**Normalised**	**Raw**	**%**	**Normalised**
PANAS-P	2.56	13	0.08	4.00	17	0.11	-0.78	0	-0.04
PANAS-N	4.72	34	0.20	3.47	29	0.13	3.31	21	0.13
STAI-S	-25.61	-63	-0.55	-22.61	-64	-0.45	-29.78	-65	-0.62
Solidarity	0.48	12	-0.02	-0.11	6	-0.01	0.31	15	0.02
Satisfaction	-1.15	-4	0.06	0.42	2	0.02	-0.29	4	-0.01
Centrality	1.26	42	-0.06	0.05	11	0.00	-0.42	6	-0.02
Individual self-stereotyping	0.45	20	-0.04	0.53	0	0.04	0.04	6	0.00
In group- homogeneity	0.03	4	0.00	0.63	-3	0.06	-0.49	-4	-0.04
Shared Identity	-1.30	-1	0.05	2.58	-10	0.09	-0.24	3	-0.01

Table 9.12.: Constituent components of Self-Subjective domain (signed) scores, by experimental paradigm

Differences in raw scores of constituent component measures of subjective recall by paradigm										
Paradigms		**PANAS-P**	**PANAS-N**	**STAI-S**	**Solidarity**	**Satisfaction**	**Centrality**	**Individual self-stereotyping**	**In-group homogeneity**	**Shared Identity**
PR-VR	Statistic	474.0	720.5	685.5	694.0	677.5	601.0	669.5	606.5	684.5
	p-value	0.0007***	0.198	0.119	0.313	0.256	0.076	0.230	0.082	0.280
Control-VR	Statistic	195.0	419.5	262.0	409.0	422.5	364.5	383.0	402.0	333.5
	p-value	0.0003***	0.456	0.008**	0.395	0.474	0.176	0.257	0.355	0.084
PR-Control	Statistic	280.0	313.0	282.0	275.0	282.5	284.5	243.5	273.5	274.5
	p-value	0.137	0.306	0.146	0.233	0.279	0.293	0.091	0.221	0.232

Table 9.13.: Differences in paradigms by constituent components of Self-subjective domain (signed)

Relationships between recall of emotional scores							
Measures		**Percentage scores**			**Normalised scores**		
		Control	**PR**	**VR**	**Control**	**PR**	**VR**
PANAS-N and STAI-S	Statistic	5.5	6.5	14.5	15	4	0
	p-value	2.54e-07***	6.47e-12***	1.87e-16***	7.04e-07***	2.14e-12***	1.51e-16***
PANAS-P and STAI-S	Statistic	159	309.5	611.5	171.5	407	641
	p-value	0.270	0.019*	0.0002***	0.402	0.039*	0.001**
PANAS-P and PANAS-N	Statistic	2.5	8	26.5	5	18	40
	p-value	1.59e-07***	7.44e-12***	3.99e-16***	1.59e-07***	7.38e-12***	2.12e-15***

Table 9.14.: Differences in recall of positive and negative emotional scores

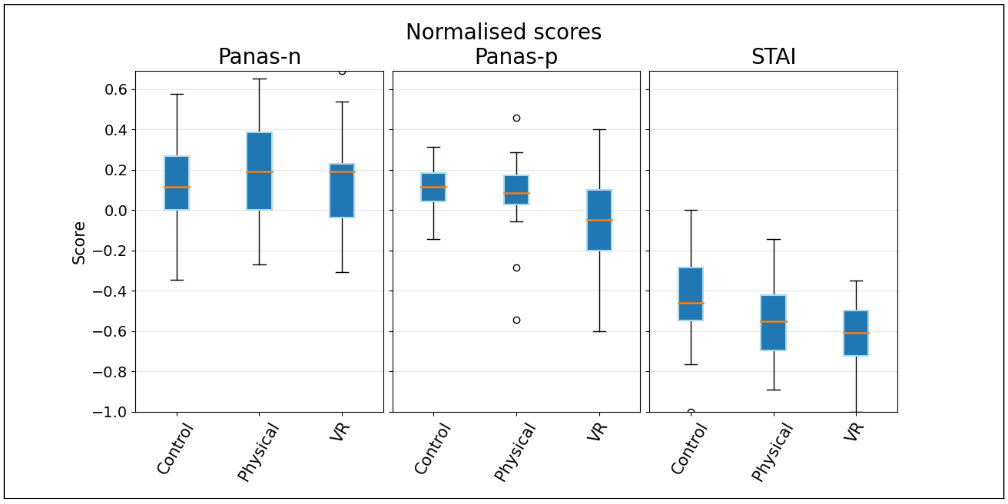


Figure 9.7.: Self-subjective normalised emotional measures

Table 9.12 shows that participants generally recall positive emotions more accurately than negative emotions, and that there is a variation in recall of group identification, with PR experiment participants underestimating social responses, VR participants tending to overestimate social responses, and Control participants ranging from under- to overestimating their stated levels of social response. This is shown graphically in Figure 9.7, Figure 9.8, Figure 9.9, and Figure 9.10.

Table 9.14 shows the differences in recall of positive emotions between paradigms. However, it also indicates that there is far less of an impact on negative emotions across paradigms. Indeed, it is possible that the difference identified in STAI-S recall between Control and VR could be an artefact of the experiment, as no such difference is identified in the equivalent test for PANAS-N. However, this result is not excluded here, owing to the higher level of significance ($p<0.01$) identified in this test.

When comparing the PANAS and STAI-S scores, it is apparent that there are strongly significant ($p<0.001$) differences across all measures (both positive and negative scores) and across all experimental paradigms. In all cases, the participants' recalled STAI-S score are significantly underestimated, both in terms of normalised and percentage scores, while the PANAS scores range from significantly overestimated to approximately accurate. This could be a systematic difference between the two measures, and will be discussed further below.

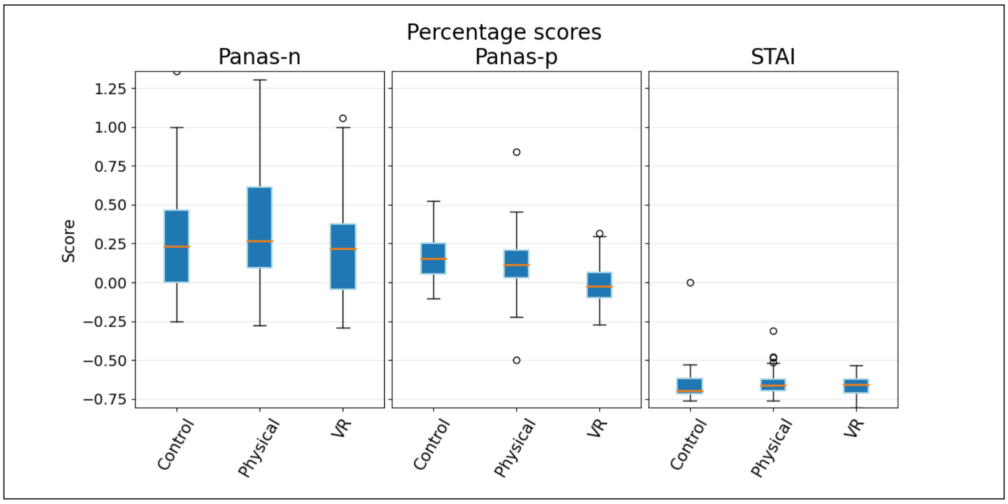


Figure 9.8.: Self-subjective percentage emotional measures

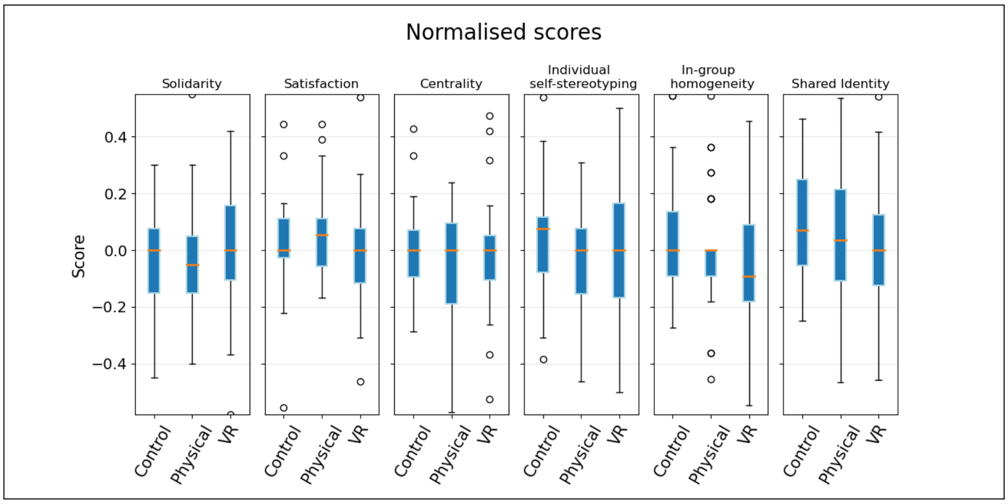


Figure 9.9.: Self-subjective normalised social measures

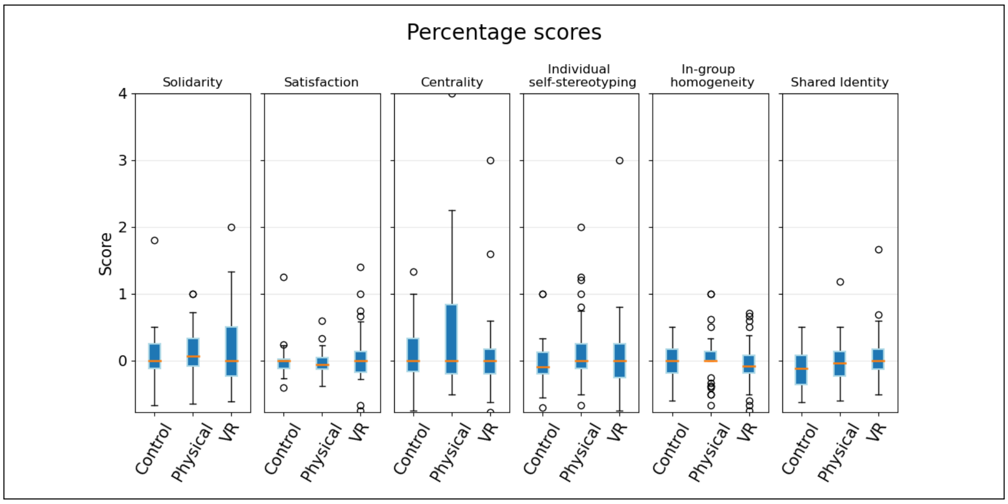


Figure 9.10.: Self-subjective percentage social measures

Additionally this study shows significant differences in the quality of recall between positive and negative emotional states. The recalled PANAS-N scores were significantly overestimated relative to the PANAS-P, when measured by both normalised and percentage scores, with the exception of the Control paradigm. However, the PANAS-N scores were significantly underestimated relative to the alternative measure of negative emotional state, the STAI-S.

No significant differences between the paradigms were observed for recalled scores for social responses.

9.6.4. Results: regression models

Domain score regression models: stress measures and demographic							
Coefficient	**Overall**	**Descriptive**	**Self-Subjective**	**Self-Behaviour**	**Aggressor**	**Neighbour**	**Spatial**
Constant	0.1679*	0.0799	0.2129***	0.4408**	0.2854*	0.1003	-0.0396
Delta-STAI	-3.44e-05	0.0008	-0.0007	-0.0008	0.0001	-0.0008	0.0032
PANAS-N	0.0001	-0.0005	0.0013	-0.0025	0.0004	0.0018	-0.0016
PANAS-P	0.0015	0.0016	-9.17e-05	5.87e-05	0.0018	0.0016	0.0061
Max Diff Cortisol	-0.0046*	-0.0080	0.0022	-0.0066	-0.0088*	-0.0066	-0.0113
Age	0.0006	0.0035	-0.0023	-0.0050	-0.0038	0.0044	0.0043
Exercise	-0.0080	0.0054	0.0060	-0.0301*	-0.0258*	-0.0266*	0.0018
Female	0.0152	0.0453	-0.0105	-0.0024	0.0362	0.0058	0.0165
Adjusted r^2	0.0153	-0.0105	0.0145	0.0338	0.0649	0.0866	-0.0277
F Statistic (p-value)	1.1625 (0.3363)	0.8855 (0.5226)	1.1531 (0.3417)	1.3847 (0.2255)	1.7634 (0.1086)	2.0434 (0.0615)	0.7034 (0.6691)
Omnibus test (p-value)	10.4466 (0.0054)	10.8924 (0.0043)	2.1749 (0.3371)	14.2503 (0.0008)	16.7150 (0.0002)	7.8060 (0.0202)	17.3663 (0.0002)

Table 9.15.: Regression models for base domain scores against stress measures and demographic

Paradigm	Mean confabulation proportion
PR	0.153
VR	0.09
Control	0.167

Table 9.16.: Confabulation proportions between paradigms

Table 9.15 reports the results from an investigation on the correlations between recall quality and measures of emotional state and social response. This study observed a slight relationship between cortisol and recall quality, where an increase in maximum cortisol difference over the experiment led to a reduction in recall quality in the Overall and Aggressor domains. Further differences were observed based on demographic, where participants who exercised more frequently had lower recall quality across Self-behaviour, Aggressor and Neighbour domains.

9.6.5. Confabulation

The final analysis was to investigate the level of confabulation within participants. Specifically, within the surveys where participants were asked a series of True/False questions, the proportion of generated memories (i.e.: where a participant reported something that did not occur) against correct answers (True positive or True negative) was recorded. The results of this a shown graphically in Figure 9.11 and tabulated in Table 9.16, indicating that approximately one confabulation is recorded every 10 correct answers, with no significant differences being observed between paradigms.

9.7. Discussion and conclusion

This chapter has identified several key results within the data obtained from the survey which will be discussed below.

9.7.1. Impact of stress

It was observed that stress can have a varying effect on an individual's recall quality, depending on the subject domain. When comparing the impact of stress between Control and both VR and PR paradigms, this study identified a single domain where stress improved recall quality (Neighbour), and a further

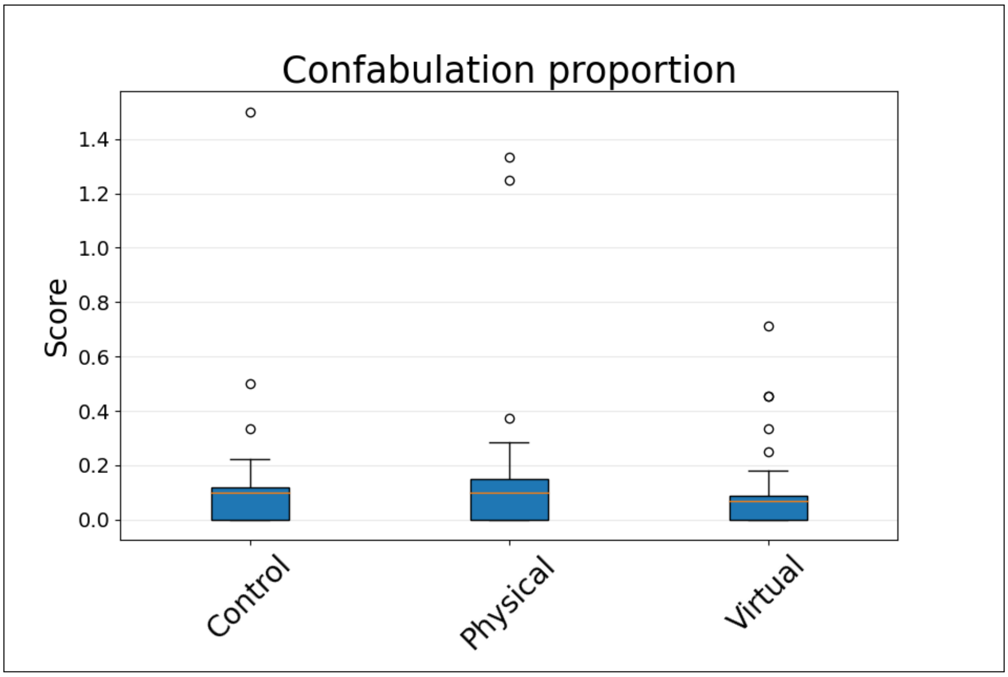


Figure 9.11.: Confabulation scores

domain where stress appeared to have no effect (Spatial). All other domains showed potentially mixed effects, with differences between VR-Control and PR-Control comparisons.

- Improvement under stress

The Neighbour-focused recall shows significant improvement with stressed individuals, where both PR ($p<0.05$) and VR ($p<0.001$) scored better than the Control branch participants. There is also a difference observed between VR and PR participants ($p<0.05$).

- No impact

No differences were observed between stressful and non-stressful paradigms in the Spatial recall domain.

The differences in Neighbour-focused recall could be a result of an increased importance placed on social interactions during emergencies, similar to that as seen in Section 8.2.

9.7.2. Comparison of VR and PR experiments

There are significant overlaps between participant recall within PR and VR paradigms. There was no significant difference observed between the recall of these paradigms within the following domains:

- Self-subjective
- Aggressor
- Spatial

Furthermore, there were only marginally significant differences in the following domains. However, given the number of tests performed, it is possible that this is an artefact of the study, and not a real effect.

- Overall
- Self-behaviour

This shows that there are multiple areas in which an individual's recollection of their experience in VR would be identical to that of a PR experiment. As this PhD is investigating pedestrian behaviour in hostile emergencies, the overlap in aggressor-centric recall is highly pertinent.

Improvement within VR

There was a strongly significant ($p<0.001$) improvement in VR recall within the Descriptive domain and a significant ($p<0.05$) difference in the Neighbour-focused. However, as with the difference in In-group homogeneity identified in the Chapter 6, it is possible that this difference is a product of the VR environment being a simplistic representation of the PR environment.

9.7.3. Relationships with participant state and demographic

The domain scores were investigated with respect to the initial emotional measures taken during the experiment. These measures (PANAS, delta-STAI and max cortisol difference) indicate the level of stress the participants underwent during the experiments. Possible relationships were observed based on maximum cortisol difference, providing some evidence to suggest that an increase in stress causes a change in recall quality. However, it is probable that this experiment did not induce a sufficient fear stimulus to impact the recall quality in an obvious manner. No relationships were observed between demographic and recall quality, across any domain.

Additionally, given that no effect was seen in the social responses as a result of the experimental approach (see Section 6.3 for more information), there was not any expected relationship between perceived shared identity and recall quality.

9.7.4. Recall of subjective concepts

The deterioration in Self-Subjective recall is only present when comparing the signed responses of VR and Control experiments. Additionally, during the investigation of the Self-Subjective domain, this study observed several specific effects. When considering the PANAS recall scores, participants tended to overestimate both positive and negative emotional states. However, there was also wide variation in the level of this overestimation. Participants consistently underestimated their STAI-S responses, but there was less variation within these recall scores than with the PANAS, with participant scores lying within approximately 50-70% underestimation change of their original stated scores, while the PANAS scores varied by up to 50% underestimation and 120% overestimation. This study also observed a strongly significant effect where in participants' recall of emotional state, with PANAS-N being overestimated significantly ($p<0.001$) more than PANAS-P across all paradigms.

When investigating subjective memories specifically, this study has seen that, on average, while the PANAS is on average more reliable through time than the STAI, the STAI deteriorates more consistently across paradigms. Therefore the STAI would be more appropriate for assessing the emotional state of participants through recall after approximately three months.

Finally, the confabulation scores were equivalent across paradigms, suggesting that the degree of memory creation does not depend on the level of stress. However, this may not be the case with more extreme stress levels.

9.8. Chapter summary

This chapter shows an assessment of participant recall of a stressful experiment procedure across PR and VR paradigms, investigating the relative accuracy of interviews as a data gathering methodology for stressful emergencies.

This chapter has successfully quantified how participants recall changes through a stressful experience, showing that the recall quality improves in specific areas that participants might deem more relevant (such as their neighbours), and deteriorates in areas that could be deemed irrelevant (such as their own behaviour). Future work should incorporate this information in any data gathering exercises, and also when assessing pre-existing data.

Finally, there was no significant difference in aggressor-centric recall between PR and VR paradigms, indicating a similar phenomenological experience of the participants. This is a strong piece of evidence supporting the use of VR in emergency-related data-gathering.

Overall, this chapter has shown some of the first results comparing recall quality between stressful and non-stressful scenarios. While these results are particularly useful for future research into emergency scenarios, the explanation of the determining factors in recall is most likely dependent on numerous variables, which are out of the scope of this thesis.

The next chapter will conclude this thesis, identifying the key results of the PhD and establishing the next stages of research in this topic.

10. Conclusion

Terrorism remains one of the most high-profile dangers in modern society. Incidents have far-reaching consequences and large impacts on modern living, from the death and injury of civilians to an ever-present level of fear within a populace. Incidents are also characterised by their extreme nature, and consequently there is little understanding surrounding how people respond within these scenarios.

It is necessary to understand these reactions if we are to make any progress in mitigating these attacks. This understanding should be based on valid data, obtained in a controlled and ecologically valid manner. However, terrorist attacks are notoriously difficult to study, largely owing to logistical and ethical limitations, and there is a significant lack of quantitative research surrounding the impact of these incidents.

This PhD approached this problem by combining a pedestrian dynamics mindset with experimental approaches, providing an ethical method of generating valid scenarios to produce the behavioural responses that would be seen in terrorist attacks. This PhD then implemented this methodology to perform two separate experiments (PR and VR), obtaining quantitative data on movement responses to hostile aggressors. This methodology was described in Chapter 4, and the resulting data was analysed in Chapter 6, Chapter 7, Chapter 8, and Chapter 9.

To conclude this thesis, this chapter will discuss the summary results from the previous chapters with respect to the project's objectives as originally stated in Chapter 1, and repeated below for reference.

10.1. Objectives revisited

Objective 1: Critically analyse the current state-of-the-art relating to pedestrian dynamics in hostile emergencies.

The initial objective of this thesis was to detail the current and potential further use for pedestrian dynamics in the case of hostile emergencies such as terrorist attacks. This understanding was established by performing interviews with key stakeholders in academia and industry, alongside a full literature review.

The results from the literature review and the interviews showed that there is a significant need for theoretical and computational models that can predict responses to emergency situations, as this would underpin resilience plans, inform design codes, and improve crowd management techniques. The results from this work also established that the main barrier to realising these benefits was the lack of readily available quantitative data, limiting the reliability of any current theory or simulations.

Objective 2: Develop a robust methodology for experimental investigation of behavioural responses to hostile emergencies.

The second objective of the thesis was to describe the methodology that was developed during this PhD. This methodology can generate the datasets identified as necessary in Objective 1 and, as part of this, defines various concepts necessary to differentiating between hostile emergencies and experimental paradigms. This methodology was designed after consultation with experts in numerous areas, including psychologists, doctors, sociologists, terrorism researchers, and pedestrian dynamics professionals. Further to this, this PhD built on established fields of academic research to define the required measurements for these experiments, notably emotional assessment and sociological identification.

A generalised methodology was presented in Chapter 3, providing an assessment of the experimental design considerations for the following concepts:

- Ethical considerations
- Threat characterisation
- Positional tracking

- Emotional and physiological assessment
- Sociological assessment
- Different data generating paradigms (VR and PR experiments, and surveys)

An implementation of this methodology was provided in Chapter 4, detailing how this PhD investigated behavioural responses to knife-based hostile aggressors.

Objective 3: Detail the design and implementation of this methodology.

This PhD then utilised this methodology to implement experimental procedures for investigating behavioural responses to knife-based hostile emergencies.

The first set of experimental procedures was carried out in December 2018, with 80 participants over 10 separate experiments carried out on 5 separate days. Each of the 10 experiments used chest-based and wrist based physiological monitors, ultra-wideband indoor position-tracking tags, and obtained saliva samples from all participants. Additionally, each participant was provided with surveys to record their self-assessed responses.

The second set of experimental procedures was a VR-based experiment, developed during a three-month research fellowship at Massey University, New Zealand, in July-August 2019. It was carried out in January-February 2020, with 55 participants taking part in individual experiments over the period. Each individual experiment lasted approximately 1.5 hours, and produced positional, physiological and survey-based datasets.

Both experiment sets were assessed by the Imperial College Research Ethics Committee, who agreed that this data was both necessary and likely to be obtained by the experimental set-ups. The data gathered by these experiments extended across several different areas, including:

- Emotional assessment
  - Physiological data
  - Self-assessed data
- Movement responses

- Recall quality

These datasets will be hosted on Imperial College London's open-source data repository.

Objective 4: Detail appropriate datasets and resulting analysis.

The fourth objective of this thesis was to describe the analysis performed on the data obtained during the experimental procedures. This analysis generated provide insight into many different avenues that had previously been theoretical. The results and analyses fall into the following four sets:

- Internal Subjective data
  - Self-assessed emotional surveys
  - Self-assessed sociological surveys
- Internal Objective data
  - Endocrine data
  - Autonomic Nervous System responses
- External data
  - Discrete responses
  - Continuous responses
- Recall quality

The analysis of these datasets included techniques such as multiple regression and machine-learning classification algorithms. Across the thesis, the effect of demographic and participant state was used as a predictor to understand whether the responses were dependent on the individual. Finally, this analysis included a comparison of different methodological paradigms (PR vs VR), providing evidence of the overlap in the results obtained by both PR and VR experimental procedures. However, it must be noted that this analysis has been performed solely within a single VR type (namely HMDs), and a single locomotion methodology (armswinging). While this is an indication of the high potential for VR to be used as a data generating paradigm, further work is necessary for a full validation, as the chosen type of VR and locomotion will have significant impacts on immersion and interaction, and therefore could potentially impact on any results.

The Internal Subjective and Internal Objective results aimed to initially provide internal validation for the experiments, assessing the emotional responses of participants. This confirmed the ecological validity of the experiments, as it showed that not only was there a detectable difference between experimental and control groups, but also that a specifically negative emotional change occurred as a result of the addition of the hostile aggressor.

The External data produced results that were specifically tailored to be easily incorporated into pedestrian dynamics simulation models, including discrete choice models (e.g. run/hide/fight/freeze), and the continuous responses (e.g. acceleration and velocity) of participants moving away from the aggressor. Both datasets are required for incorporation into a pedestrian dynamics simulation model. The analysis from this chapter showed the importance of social influence on discrete responses, which supports previous theories on socially-determined behaviour (see Section 2.2).

Finally, the Recall quality produced results that provided insight into the use of interviews and surveys for data gathering on hostile scenarios. To the best knowledge of the author, all of these datasets and analysis results are the first ever ecologically valid, quantitative datasets on human movement in response to hostile aggressors.

10.2. Hypothetical scenario revisited

In Chapter 1 this thesis suggested a hypothetical example of a terrorist attack, shown below in Figure 10.1. Four pedestrian members of the public (in blue) were being attacked by one hostile aggressor (in red).

Initially this thesis indicated the variety of actions available to the pedestrians in this scenario, including fight-flight-freeze and hiding responses, but also noted that it was impossible to know which of these actions each individual would perform, and further, that a typical modern modelling method would be to assume certain action sets and to predict the consequent outcome. This approach, however, is theoretical at best, and misleading at worst.

After the research done in this PhD, an engineer hoping to model this scenario can make informed choices over the likely actions of each individual in the scenario, and therefore have a far greater confidence in any results. For

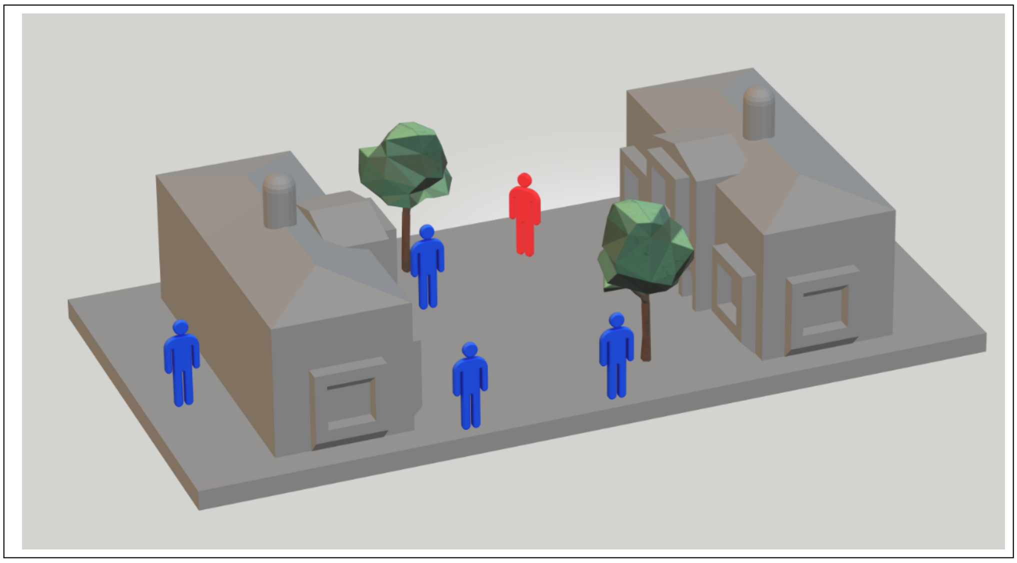


Figure 10.1.: Hypothetical scenario revisited

example, each blue pedestrian would select their actions using the discrete choice analysis results detailed in Section 8.2. For any pedestrian that has decided to move their movement could be modelled using the continuous analysis detailed in Section 8.3. The pedestrian behind the building might move according to day-to-day relationships until they confirm the presence of the threat (either through audio or visual stimuli), at which point they will follow the same discrete and continuous relationships as the other pedestrians.

Furthermore, the data provided in this thesis provides further insight into the likely responses of the blue pedestrians based on their demographic. If common parameters (such as age, gender and exercise level) are known about these pedestrians, any modelling performed can be further refined.

Result	**Chapter**	**Detail**	**Impact**
A generalised methodology that can be adapted for arbitrary terrorist attacks.	Methodology: Key Principles	A detailed view of the key measurements and logistical approaches for an experimental team to consider when designing an experiment that investigates human behavioural responses to hostile emergencies such as terrorist attacks.	Provides a theoretical basis to generate experimental designs for hostile emergency investigations.
An implementation of this methodology, adapted for knife-based attacks.	Methodology: Implementation	An example implementation of the theoretical basis, compared between Physical and virtual paradigms.	Provides a practical example of this type of research.

Result	Chapter	Detail	Impact
These experiments produced self-assessed 'fear'.	Internal Subjective	When measuring the emotions of participants using standard, validated self-assessed survey measurements, participants in the experimental groups (both Physical and VR paradigms) experienced more negative emotions (fear, stress etc.) than those in the control groups. This is broadly independent of personality type, and does not correlate strongly with the level of group identification observed.	Shows that it is practical to ethically produce a level of emotional reaction that mimics that of hostile emergencies, an issue that was previously considered intractable.
These experiments produced physiological 'fear'.	Internal Objective	The objective measurements of emotional response (endocrine and ANS responses), shows that there was a significant difference between pre- and post-hostile aggressor. This was quantified by both difference testing and machine learning classification algorithms.	Provides internal validation of the experimental procedure, as well as indication of the relative benefits of each measure of emotion.

Result	Chapter	Detail	Impact
There are significant discrete and continuous relationships in movement	External	Understanding movement behaviour in emergencies is key to informing pedestrian dynamics models that can be used for design processes. The behaviour observed indicates a different response to hostile emergencies than had previously been assumed, especially with the level of social dependence.	Previous simulation models were not based on valid, controlled data, so the provision of this data represents a huge leap in the predictive capability of these models.
There are significant differences in recall quality based on question domain, with a consequent implication for survey- and interview-based data collection.	Recall	The recall quality of individuals who have gone through stressful scenarios is known to be variable, with inconclusive and non-quantitative data on the level of quality. This recall quality was investigated by comparing the responses of participants to a survey distributed three months after their experiments with what occurred during their experiments.	It was seen that recall quality depends significantly on the domain.

Result	**Chapter**	**Detail**	**Impact**
Different paradigm comparisons	Internal Subjective Internal Objective External Memory	This research performed experiments across two major paradigms (Physical and Virtual Reality), to understand the relative benefits and potential validity of both. By performing a near-identical experiment, the results of these paradigms can be compared, and this comparison showed small differences between the two. Key differences include a heightened dependence on social behaviours in Virtual Reality, and differences based on gender.	Provides insight into the relative benefits of different methods of generating quantitative data, including Physical and Virtual Reality, and provides detail on the potential drawbacks of using interviews for qualitative data generation. As a result of this comparison, this PhD concludes that Virtual Reality is an appropriate tool for data gathering, with the caveats that experiments must be designed to ensure social interactions are present between participants, and that demographic differences are tracked.

Table 10.1.: PhD key results

10.3. Impact of this PhD

This section covers the impacts of this PhD, and the outline is shown graphically in Figure 10.2.

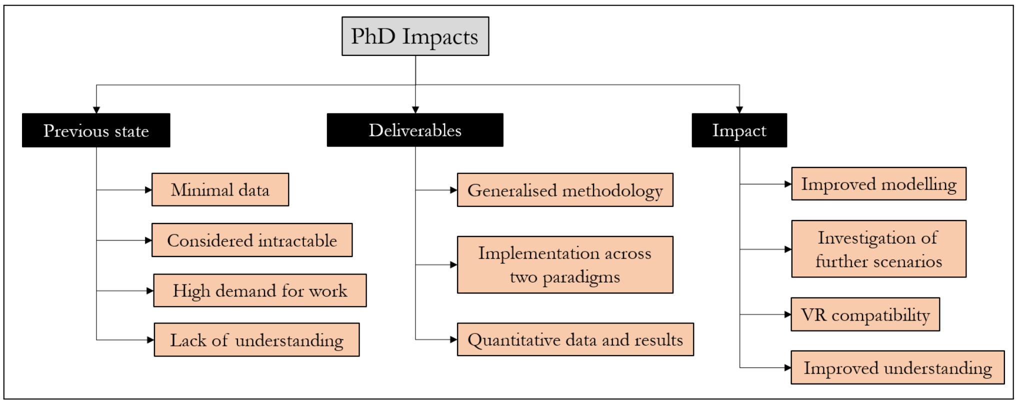


Figure 10.2.: Impact of this PhD

Previous state

The first chapters of this thesis described the initial research performed as part of this PhD, establishing several key points covering the limitations of the state of this topic, in academia, industry, and governance.

1) There is little data surrounding human behavioural responses to hostile emergencies.

There have been large amounts of research aiming to understand the psychology of people involved in emergencies, leading to the development of numerous models of behaviour and response.

However, this research has not yet produced a quantitative, predictive model of behavioural responses, nor a model that predicts movements. This is at least partially due to a lack of any useful data on the movement of pedestrians in the scenarios. For pedestrian dynamics approaches, this data is crucial in validating such models.

2) A large proportion of previous approaches have suggested that this problem is intractable.

The majority of previous approaches have been computational, based on either modeller assumptions or datasources of questionable relevance (e.g. ants, mice, etc.). Further to this, a large number of previous approaches have stated that performing the appropriate types of experiment would be impossible, due to ethical and logistical reasons (see [3] for further information).

3) There is a demand for this work.

Understanding the movement of pedestrians in hostile emergencies will lead to a significant improvement in the survivability of such situations. There is ongoing work within this sphere, within areas such as security, resilience, crowd management, emergency planning and infrastructure design. For example, [55] was performed in conjunction with airport management, noting that stakeholder engagement suggested that this type of work would be particularly useful.

4) There is a widespread misunderstanding surrounding outdated theories.

There are examples of modern policy provided to inform members of the public of what to do during an emergency, such as 'Run, Hide, Tell' [193]. However, policy documents distributed to government officials have been found to be based on outdated theoretical approaches, such as panic theory [86]. Moreover, there is a commonplace misconception on the topic of panic, and its prevalence within individuals involved in emergencies.

Deliverables

Having defined the limitations and requirements in this area, this PhD performed several key pieces of research to provide solutions (detailed in Table 10.1).

1) A generalisable methodology.

2) A successful implementation of this methodology, in two separate paradigms.

3) Initial results and analysis.

Impact

To summarise, these deliverables will have a significant impact on the wider area of pedestrian dynamics, including:

1) Improved simulation capability.

As detailed in the hypothetical scenario, the discrete and continuous models produced in this PhD provide insight into the movement of pedestrians in response to hostile aggressors. Future work can can now use the results established in this thesis.

This improved simulation must also be supported by research into multimodal movement responses. This PhD found that typical computational methods were inappropriate for modelling this type of emergency response, largely because they designate a maximum movement speed, corresponding to an individual walking unimpeded at low local densities. This, however, is not the case when the individual changes from walking to running, as a result of a perceived threat, or otherwise. These simulation approaches could be improved upon by adding in further psychological variables, risk perception models, motivation levels, or any further appropriate method.

2) Potential expansions to further, arbitrary scenarios.

The generalised methodology allows for different scenarios to be considered. Examples are provided below.

3) Validity of multiple paradigms.

Prior to this, there was significant concern that behavioural responses (both discrete choices and continuous movement) within VR environments were not similar to those in real world environments. One of the major findings of this PhD is the analysis that suggests VR can function as an acceptable data collection paradigm for emergency scenarios.

4) Improved understanding surrounding human behaviour in emergencies.

As an example, the discrete choice model showed that social factors are among the most important in determining the choices of individuals. This increased understanding could extend to improvements for current theory, for legislation and policy, for security planning and crowd management.

10.4. Future work for pedestrian dynamics research and industry

The final part of this PhD discusses the application of this methodology for arbitrary scenarios and investigates proposed future research topics, with the specific aim of assessing potential implementation of the above results for real-world applications.

10.4.1. Arbitrary hostile emergencies

The generalised methodology produced by this PhD provides a way of generating experimental designs for arbitrary hostile emergencies. The example implementation focused on knife-based emergencies, owing to the relative frequency and high-impact of such emergencies. However, there are further attack types that should be considered and analysed in future, including:

- Firearm-based incidents

This type of incident is particularly prevalent in the USA [187], and represents a slightly different threat characterisation that that of knife-based incidents. Specifically, the physical nature of the event would be remote rather than tactile (see Section 3.3), but otherwise could be characterised identically. This difference, however, might produce significantly different behavioural responses, given the far-reaching potential of the stressor.

- Vehicle-based incidents

There have been many examples of vehicle-based incidents (see Section 2.5), and they represent a very different threat characterisation to knife- and firearm-based incidents. However, the logistical complexities of representing a hostile vehicle within a Physical experiment may be intractable, and so instead a Virtual Reality paradigm should be considered for investigating these behavioural responses.

- Multiple vector-based incidents

A key feature of many modern terrorist attacks is the number of attack vectors and number of stressing agents involved. For example, the 2017 London Borough Market incident involved three hostile individuals, with vehicular and knife-based attack vectors. The hostile individuals were also wearing fake explosive vests, further complicating the incident. While this PhD investigated a single attack vector (i.e. knives) with a single stressing agent, it would also be a worthwhile area of research to investigate the behavioural responses to multiple vectors.

10.4.2. Simulation methods

As part of this PhD, and with funding obtained from the ISST Champion's Fund, the author spent time in Munich, Germany, working with Accu:rate, a company specialising in pedestrian dynamics consultancy. Accu:rate own a proprietary software ("Crowd:it", application available commercially), based on the Optimal Steps Model [5]. This period was spent simulating the results from the PR experiment, and significant progress was made in the ability to simulate the responses observed in the experimental procedures, specifically the results from the discrete choice model (see Section 8.2). However, there were limitations concerning implementing the continuous response models, owing to an inability to simulate multimodal responses. Section 8.3 details how participants would run when closer to the aggressor. The majority of simulation models do not accommodate running behaviour, instead focusing on variable-speed walking relationships. It was therefore not possible to fully replicate the results from the PR experiment.

Thus a fruitful area of future research would be to develop theoretical and computational models that can produce multimodal pedestrian responses, potentially based on perceived urgency. As part of this, several requirements will need to be satisfied:

- Experimental data

Fundamental diagrams and other basic data sets are routinely generated for walking behaviour. However to the author's best knowledge these do not exist for running responses, nor for transitions between walking and running movement styles.

- Theoretical models

The above experimental data will then underpin theoretical constructs that can explain when pedestrians transition from walking to running, as well as the movement within both of these modes.

This PhD proposes a simple theoretical model based on the data generated experimentally: that participants run when closer to an aggressor, and that this behaviour is strongly socially determined. However, it is not suggested that this is the only occasion when pedestrians will run, and alternative density-flow-speed relationships are required before a fully-formed theoretical model can be realised.

10.4.3. Further analysis, including classification and prediction of hostile scenarios

The external results chapter commented that the positional data obtained during the experiments actually forms a time-series, meaning that adjacent datapoints are not independent, which violates the independence requirement for linear regression models. While the overall validity of the models is ensured by the high number of datapoints, as well as by the overall duration of the experiments (two datapoints separated by enough time can be considered independent in this context), there are further approaches might offer fruitful areas of research.

To analyse time-series data, there are numerous established approaches, with machine learning techniques being possibly the most high-profile at the time of writing. Suggested examples of possible machine learning techniques include:

- Long-Short-Term Memory (LSTM) models [277]
- Transformers [278]
- Generative Adversarial Network (GAN) models [279]

These approaches are suited to time-series data, and are commonly implemented in natural language processing (NLP) tasks (e.g. speech recognition). They have also been used in previous pedestrian dynamics research [280], however, there has not been widespread uptake of their use in simulation models.

Using the data generated in this experiment and future experiments, these techniques could be used to investigate several areas of research and industry. Specifically, these models could be implemented to:

- Understand the important features of trajectory prediction

Linear regression models are limited in that they require the modeller to specify the predictive variables that are considered important. This PhD has attempted to overcome this limitation by specifying numerous predictive variables, and then using a forward-backward stepwise selection procedure to only include the relevant variables. However, utilising an LSTM or a Transformer approach would allow a vast number of predictive variables and interactions to be assessed, and then highlighted if they proved to be relevant. Understanding these features would enable engineers to produce user-focused infrastructure designs to accommodate hostile scenarios.

- Predict movement responses to specific scenarios

GAN models are well positioned to generate models that can predict the movement responses to the types of scenario considered in this PhD. GAN models initially produce a predictive model, and then attempt to differentiate between its prediction and the real data. If it can tell the difference between the two datasets, it adapts the generative model, producing another dataset. This continues until the model produces data that is indistinguishable from the real data, and as a result the model can be used with sufficient confidence for further arbitrary scenarios.

- Classify hostile scenarios

State classification is a key issue within machine learning, and this PhD used various algorithms to classify the state of participants based on their physiological signals (Section 7.3). This approach could be extended to classify the movement state of pedestrians, to see how closely they resemble the movements commonly seen during hostile emergencies. The level of accuracy of such a classification system would be dependent on the data used in training the model, and therefore it would be necessary to perform more experiments in alternative scenarios.

The ability to classify the occurrence of hostile scenarios would have a strong potential impact on the management of hostile emergencies. For example, it could lead to an automated early detection system that identifies areas where emergencies are occurring and informs the emergency services, rather than relying on the people involved calling, which itself can be a dangerous activity.

Overall, this PhD has sought to extend the understanding of how people react in hostile emergencies, and to improve the capability of researchers to investigate differing scenarios and thus allow future work to mitigate against some of the most horrific incidents in modern life. While much future work research remains to be done, the output of this PhD has defined and extended the limits of current expertise in this area, providing a solid basis upon which future engineers can build.

10.5. Contributions

This PhD produced several pieces of peer-reviewed work throughout, including:

Journal papers

- "Ethical implications of designing infrastructure around human behaviour." Shipman, IEEE, 2019
- "A glossary for research on human crowd dynamics." Adrian et al, Collective Dynamics, 2019.
- "Fear in Humans, a glimpse into the computer modelling perspective." Shipman and Majumdar, Transport Research Record, 2018.

Conference papers

- Transport Research Board, Washington DC, USA, "A practical, low intensity methodology to optimise evacuations from existing buildings." 2019. Poster presentation.
- Physics and Psychology of Crowd Dynamics, Leiden, Netherlands. 'A methodology for the experimental investigation of human behaviour in terrorist attacks'. 2018, Oral presentation
- Pedestrian and Evacuation Dynamics, Lund, Sweden. "A test to ascertain the optimal fire evacuation procedure, and the efficacy of interventions." 2018. Oral presentation.
- Transport Research Board, Washington DC, USA, "Fear in humans, a glimpse into the computer modelling perspective." 2018. Poster presentation.

- Values in Medicine, Science and Technology, Dallas USA. "Ethical considerations for data gathering for humans in extreme emergencies". 2017. Roundtable presentation.

Bibliography

[1] M. Kinateder, E. Ronchi, D. Nilsson, M. Kobes, M. Müller, P. Pauli, and A. Mühlberger, "Virtual reality for fire evacuation research," in *2014 Federated Conference on Computer Science and Information Systems*, Conference Proceedings, pp. 313–321.

[2] G. Chaliand and A. Blin, *The history of terrorism : from antiquity to al Qaeda.* Berkeley, Calif. ; London: University of California Press, 2007. [Online]. Available: http://www.loc.gov/catdir/toc/ecip072/2006032389.html

[3] A. Shipman and A. Majumdar, "Fear in humans: A glimpse into the crowd-modeling perspective," *Transportation Research Record*, vol. 2672, no. 1, pp. 183–197, 2018. [Online]. Available: https://doi.org/10.1177/0361198118787343

[4] D. Helbing and A. Johansson, *Pedestrian, Crowd and Evacuation Dynamics*, 2010, vol. 16, pp. 697–716.

[5] I. von Sivers and G. Köster, *Realistic stride length adaptation in the optimal steps model.* Springer, 2015, pp. 171–178.

[6] M. Boltes, M. Chraibi, W. Liao, A. Tordeux, W. Mehner, J. Zhang, S. Holl, A. U. K. Wagoum, and G. Lämmel, "Experimentation, data collection, modeling and simulation of pedestrian dynamics," Conference Proceedings.

[7] X. Li, I. Hijazi, R. Koenig, Z. Lv, C. Zhong, and G. Schmitt, "Assessing essential qualities of urban space with emotional and visual data based on gis technique," *ISPRS International Journal of Geo-Information*, vol. 5, no. 11, 2016.

[8] S. S. Dickerson and M. E. Kemeny, “Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research,” *Psychological Bulletin*, vol. 130, no. 3, pp. 355–391, 2004.

[9] S. Spivak, “Articular: an augmented reality game concept for a museum setting,” Thesis, 2015, auckland University of Technology.

[10] C. Santifort, T. Sandler, and P. T. Brandt, “Terrorist attack and target diversity: Changepoints and their drivers,” *Journal of Peace Research*, vol. 50, no. 1, pp. 75–90, 2012. [Online]. Available: https://doi.org/10.1177/0022343312445651

[11] Global terrorism database. Accessed on 09/11/2017. [Online]. Available: https://www.start.umd.edu/gtd/

[12] E. Quarantelli, “Images of withdrawal behavior in disasters: Some basic misconceptions,” *Social Problems*, vol. 8, pp. 68–79, 1960.

[13] G. Naudet and J. Naudet, “November 13: Attack on Paris,” 2018, distributed by Netflix.

[14] London bridge attack: Skateboard hero receives bravery award. Accessed on 19/01/2021. [Online]. Available: https://www.bbc.co.uk/news/uk-45809954?ns_source=facebook&ns_mchannel=social&ns_campaign=bbcnews&ocid=socialflow_facebook&fbclid=IwAR1NtY_O_K_ijP3lCU8miTTbkDjA67nCrLo07mwcMZXX9wU5YraPScNCgLU

[15] S. Bennett. Paddington station legion modelling. Accessed on 10/06/2019. [Online]. Available: https://learninglegacy.crossrail.co.uk/documents/station-pedestrian-modelling-archive/attachment/paddington-station-legion-modelling-c130-swn-t1-xmo-b071-50001_rev3-0-pdf/

[16] Using pedestrian library to create a new look for an art gallery. Accessed on 19/01/2021. [Online]. Available: https://www.anylogic.com/the-ametria-project-using-pedestrian-library-to-create-a-new-look-for-an-art-gallery/

[17] D. Marin. Modelling pedestrian movement with sense: Old science, new approach. Accessed on 19/1/2021. [Online]. Available-

able: https://www.movementstrategies.com/blog-posts/modelling-pedestrian-movement-with-sense-old-science-new-approach.

[18] E. Carattin. Dynamic crowd modelling for better stadium design. Accessed on 19/01/2021. [Online]. Available: https://www.oasys-software.com/news/dynamic-crowd-modelling-better-stadium-design/

[19] G. Wilkinson. Crowd safety at hajj- simulation modeling. Accessed on 19/01/2021. [Online]. Available: https://www.anylogic.com/blog/crowd-safety-at-hajj-simulation-modeling/

[20] A. Cuesta and S. M. V. Gwynne, "The collection and compilation of school evacuation data for model use," *Safety Science*, vol. 84, pp. 24–36, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925753515002805

[21] R. Swamy, J. E. Kang, R. Batta, and Y. Chung, "Hurricane evacuation planning using public transportation," *Socio-Economic Planning Sciences*, vol. 59, pp. 43–55, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038012116300210

[22] H. Dong, M. Zhou, Q. Wang, X. Yang, and F. Wang, "State-of-the-art pedestrian and evacuation dynamics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 5, pp. 1849–1866, 2020.

[23] S. Li, J. Zhuang, S. Shen, and J. Wang, "Driving-forces model on individual behavior in scenarios considering moving threat agents," *Physica A: Statistical Mechanics and its Applications*, vol. 481, pp. 127–140, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437117303114

[24] M. Awada, R. Zhu, B. Becerik-Gerber, G. Lucas, and E. Southers, "An integrated emotional and physiological assessment for vr-based active shooter incident experiments," *Advanced Engineering Informatics*, vol. 47, p. 101227, 2021. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474034620301968

[25] J. Tsai, E. Bowring, S. Marsella, and M. Tambe, "Empirical evaluation of computational fear contagion models in crowd dispersions," *Autonomous Agents and Multi-Agent Systems*, vol. 27, no. 2, pp.

200–217, 2013. [Online]. Available: https://doi.org/10.1007/s10458-013-9220-6

[26] J. M. Chertkoff and R. H. Kushigian, *Don't panic : the psychology of emergency egress and ingress.* Westport, Conn: Praeger, 1999.

[27] C. McPhail, *The Myth of the Madding Crowd.* Walter de Gruyter, Inc., 1991.

[28] H. J. P. Timmermans, *Pedestrian behavior : models, data collection and applications.* Emerald Group Publishing Limited, 2009.

[29] E. Cristiani, B. Piccoli, and A. Tosin, "Multiscale modeling of granular flows with application to crowd dynamics," *Multiscale Modeling and Simulation*, vol. 9, no. 1, pp. 155–182, 2011.

[30] P. R. Neumann, *Old and new terrorism : late modernity, globalization and the transformation of political violence.* Cambridge, UK: Polity, 2009.

[31] N. Pelechano, J. M. Allbeck, M. Kapadia, and N. I. Badler, *Simulating heterogeneous crowds with interactive behaviors.* A K Peters CRC Press, 2016.

[32] A. Cuesta, O. Abreu, and D. Alvear, *Evacuation Modeling Trends.* Springer Publishing Company, Incorporated, 2015.

[33] R. Iphofen, *Ethical Issues in surveillance and privacy*, 1st ed., 2015, pp. 59–75.

[34] D. C. Duives, W. Daamen, and S. P. Hoogendoorn, "State-of-the-art crowd motion simulation models," *Transportation Research Part C-Emerging Technologies*, vol. 37, pp. 193–209, 2013. [Online]. Available: ⟨GotoISI⟩://WOS:000329419500012

[35] A. Templeton, J. Drury, and A. Philippides, "From mindless masses to small groups: Conceptualizing collective behavior in crowd modeling," *Rev Gen Psychol*, vol. 19, no. 3, pp. 215–229, 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/26388685

[36] J. Adrian, M. Amos, M. Baratchi, M. Beermann, N. Bode, M. Boltes, A. Corbetta, G. Dezecache, J. Drury, Z. Fu, R. Geraerts, S. Gwynne, G. Hofinger, A. Hunt, T. Kanters, A. Kneidl, K. Konya, G. Köster, M. Küpper, G. Michalareas, F. Neville, E. Ntontis, S. Reicher, E. Ronchi, A. Schadschneider, A. Seyfried, A. Shipman, A. Sieben, M. Spearpoint, G. B. Sullivan, A. Templeton, F. Toschi, N. v. d. Wal, F. v. Schadewijk, C. v. Krüchten, N. Wijermans, Z. Yücel, F. Zanlungo, and I. Zuriguel, "A glossary for research on human crowd dynamics," *Collective Dynamics; Vol 4 (2019)*, 2019. [Online]. Available: https://collective-dynamics.eu/index.php/cod/article/view/A19

[37] B. D. Hankin and R. A. Wright, "Passenger flow in subways," *Journal of the Operational Research Society*, vol. 9, no. 2, pp. 81–88, 1958. [Online]. Available: https://doi.org/10.1057/jors.1958.9

[38] J. J. Fruin, *Pedestrian planning and design*, 1st ed. New York, Metropolitan Association of Urban Designers and Environmental Planners, 1971.

[39] J. J. Fruin, "Designing for pedestrians. a level- of- service concept," *Highw Res Rec*, no. 35, pp. 1–15, 1971.

[40] E. Kuligowski, "Review of 28 egress models," NIST SP 1032, Report, 2005.

[41] V. Predtechensky and A. Milinsky, "Planning for foot traffic flow in buildings," *Amerind Publishing Co., New Delhi*, 1978.

[42] V. Predtechensky, "Calculative design of foot traffic flows in mass public buildings," *Higher School News – (Architecture and Construction) №7*, 1958.

[43] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, 1995. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.51.4282

[44] F. Zanlungo, Z. Yucel, and T. Kanda, "The effect of social roles on group behaviour," in *Pedestrian and Evacuation Dynamics*, Conference Proceedings.

[45] S. A. Doychinov and A. S. Dederichs, "The development of drunk behaviour during evacuation," *Collective Dynamics; Vol 5 (2020)DO - 10.17815/CD.2020.48*, 2020. [Online]. Available: https://collective-dynamics.eu/index.php/cod/article/view/A48

[46] S. Gwynne, E. R. Galea, M. Owen, P. J. Lawrence, and L. Filippidis, "A review of the methodologies used in the computer simulation of evacuation from the built environment," *Building and Environment*, vol. 34, no. 6, pp. 741–749, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360132398000572

[47] Ucrowds. Accessed on 08/02/2021. [Online]. Available: https://www.ucrowds.com/

[48] R. L. Hughes, "A continuum theory for the flow of pedestrians," *Transportation Research Part B: Methodological*, vol. 36, no. 6, pp. 507–535, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0191261501000157

[49] L. Huang, S. C. Wong, M. Zhang, C.-W. Shu, and W. H. K. Lam, "Revisiting hughes' dynamic continuum model for pedestrian flow and the development of an efficient solution algorithm," *Transportation Research Part B: Methodological*, vol. 43, no. 1, pp. 127–141, 2009.

[50] R. Lovreglio, D. Borri, L. dell'Olio, and A. Ibeas, "A discrete choice model based on random utilities for exit choice in emergency evacuations," *Safety Science*, vol. 62, pp. 418–426, 2014.

[51] C. Hollmann, P. J. Lawrence, and E. R. Galea, *Introducing Emotion Modelling to Agent-Based Pedestrian Circulation Simulation*. Boston, MA: Springer US, 2011, pp. 799–803. [Online]. Available: http://dx.doi.org/10.1007/978-1-4419-9725-8_77

[52] D. Helbing, I. J. Farkas, P. Molnár, and T. Vicsek, "Simulation of pedestrian crowds in normal and evacuation situations," in *Pedestrian and Evacuation Dynamics, Duisburg, Germany, April 4-6, 2001*, M. Schreckenberg and S. D. Sharma, Eds. Springer, Conference Proceedings, p. 58. [Online]. Available: http://hdl.handle.net/20.500.11850/123980

[53] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic," *Nature*, vol. 407, no. 6803, pp. 487–490, 2000. [Online]. Available: http://dx.doi.org/10.1038/35035023

[54] H. Zou, N. Li, and L. Cao, *Emotional Response–Based Approach for Assessing the Sense of Presence of Subjects in Virtual Building Evacuation Studies*, 2017.

[55] J. Tsai, N. Fridman, E. Bowring, M. Brown, S. Epstein, G. Kaminka, S. Marsella, A. Ogden, I. Rika, A. Sheel, M. E. Taylor, X. Wang, A. Zilka, and M. Tambe, "Escapes: evacuation simulation with children, authorities, parents, emotions, and social comparison," in *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*. International Foundation for Autonomous Agents and Multiagent Systems, Conference Proceedings, p. 457–464.

[56] L. Ma, B. Chen, S. Qiu, Z. Li, and Z. Song, "Agent-based modeling of emergency evacuation in a railway station square under sarin terrorist attack," *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 8, 2016.

[57] Z. Fa, W. Songchun, and S. Zhihua, "Evacuation during violent attacks: Agent-based modeling and simulation," in *IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Conference Proceedings, pp. 412–415.

[58] O. Abreu, A. Cuesta, A. Balboa, and D. Alvear, "On the use of stochastic simulations to explore the impact of human parameters on mass public shooting attacks," *Safety Science*, vol. 120, pp. 941–949, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753519305296

[59] C. Arteaga and J. Park, "Building design and its effect on evacuation efficiency and casualty levels during an indoor active shooter incident," *Safety Science*, vol. 127, p. 104692, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753520300898

[60] C. Cho, J. Park, and S. Sakhakarmi, "Emergency response: Effect of human detection resolution on risks during indoor mass shooting

events," *Safety Science*, vol. 114, pp. 160–170, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753518316205

[61] K. J. Krizek and H. Timmermans, *TRB 2010 Transportation Research Board : Location choice models*. Springer, 2010. [Online]. Available: 1850-9999http://www.springer.com/gb/BLDSS

[62] Q. Liu, "A social force approach for the defensive strategy of security guards in a terrorist attack," *International Journal of Disaster Risk Reduction*, vol. 46, p. 101605, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212420919313937

[63] Q. Liu, "A social force model for the crowd evacuation in a terrorist attack," *Physica A: Statistical Mechanics and its Applications*, vol. 502, pp. 315–330, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437118302607

[64] C. Chen, Y. Tong, C. Shi, and W. Qin, "An extended model for describing pedestrian evacuation under the threat of artificial attack," *Physics Letters A*, vol. 382, no. 35, pp. 2445–2454, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037596011830673X

[65] H. Faroqi and S. Mesgari, "Agent-based crowd simulation considering emotion contagion for emergency evacuation problem," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1-W5, pp. 193–196, 2015.

[66] M. Zhou, H. Dong, J. Zhang, X. Sun, and X. Yao, "Effect of assailants on crowd evacuation: A fuzzy logic approach," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Conference Proceedings, pp. 1098–1103.

[67] L. Zheng, X. Peng, L. Wang, and D. Sun, "Simulation of pedestrian evacuation considering emergency spread and pedestrian panic," *Physica A: Statistical Mechanics and its Applications*, vol. 522, pp. 167–181, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437119301359

[68] E. M. J. Huis in 't Veld and B. de Gelder, "From personal fear to mass panic: The neurological basis of crowd perception," *Human Brain Mapping*, vol. 36, no. 6, pp. 2338–2351, 2015. [Online]. Available: http://dx.doi.org/10.1002/hbm.22774

[69] T. Elzie, E. Frydenlund, A. J. Collins, and R. M. Robinson, "Panic that spreads sociobehavioral contagion in pedestrian evacuations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2586, pp. 1–8, 2016. [Online]. Available: https://doi.org/10.3141/2586-01

[70] W. B. Cannon, *Bodily changes in pain, hunger, fear and rage: An account of recent researches into the function of emotional excitement.* New York, NY, US: D Appleton and Company, 1915.

[71] B. E. Aguirre, S. El-Tawil, E. Best, K. B. Gill, and V. Fedorov, "Contributions of social science to agent-based models of building evacuation," *Contemporary Social Science*, vol. 6, no. 3, pp. 415–432, 2011. [Online]. Available: http://dx.doi.org/10.1080/21582041.2011.609380

[72] X. Pan, C. Han, K. Law, and J. Latombe, "A computational framework to simulate human and social behaviors for egress analysis," Conference Proceedings.

[73] M. D'Orazio, L. Spalazzi, E. Quagliarini, and G. Bernardini, "Agent-based model for earthquake pedestrians' evacuation in urban outdoor scenarios: Behavioural patterns definition and evacuation paths choice," *Safety Science*, vol. 62, pp. 450–465, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925753513002233

[74] M. D'Orazio, E. Quagliarini, G. Bernardini, and L. Spalazzi, "Epes – earthquake pedestrians' evacuation simulator: A tool for predicting earthquake pedestrians' evacuation in urban outdoor scenarios," *International Journal of Disaster Risk Reduction*, vol. 10, pp. 153–177, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2212420914000648

[75] X. Yang, Z. Wu, and Y. Li, "Difference between real-life escape panic and mimic exercises in simulated situation with implications

to the statistical physics models of emergency evacuation: The 2008 wenchuan earthquake," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 12, pp. 2375–2380, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437110008770

[76] S. Li, J. Zhuang, and S. Shen, "A three-stage evacuation decision-making and behavior model for the onset of an attack," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 119–135, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X17300840

[77] J. Wang, S. Ni, S. Shen, and S. Li, "Empirical study of crowd dynamic in public gathering places during a terrorist attack event," *Physica A: Statistical Mechanics and its Applications*, vol. 523, pp. 1–9, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037843711930127X

[78] M. J. Kinsey, S. M. V. Gwynne, E. D. Kuligowski, and M. Kinateder, "Cognitive biases within decision making during fire evacuations," *Fire Technology*, 2018. [Online]. Available: https://doi.org/10.1007/s10694-018-0708-0

[79] N. J. Smelser, *Theory of Collective Behavior.* Collier-MacMillan Canada, Ltd., 1962.

[80] Thresher's off license fire cctv. Accessed on 12/02/2021. [Online]. Available: https://www.youtube.com/watch?v=kyP6RCfVCjI

[81] E. Quarantelli, "The nature and conditions of panic," *American Journal of Sociology*, vol. 60, no. 3, pp. 267–275, 1954.

[82] J. Keating, "The myth of panic," *Fire*, vol. 147, 1982.

[83] R. F. Fahy, G. Proulx, and L. Aiman, "Panic or not in fire: Clarifying the misconception," *Fire and Materials*, vol. 36, no. 5-6, pp. 328–338, 2012.

[84] C. E. Fritz and E. S. Marks, "The norc studies of human behavior in disaster," *Journal of Social Issues*, vol. 10, no. 3, pp. 26–41, 1954. [Online]. Available: http://dx.doi.org/10.1111/j.1540-4560.1954.tb01996.x

[85] J. Drury, D. Novelli, and C. Stott, "Psychological disaster myths in the perception and management of mass emergencies," *Journal of Applied Social Psychology*, vol. 43, no. 11, pp. 2259–2270, 2013.

[86] J. Drury, D. Novelli, and C. Stott, "Representing crowd behaviour in emergency planning guidance: 'mass panic' or collective resilience?" *Resilience*, vol. 1, no. 1, pp. 18–37, 2013.

[87] D. P. Schultz, "Theories of panic behavior: A review," *The Journal of Social Psychology*, vol. 66, no. 1, pp. 31–40, 1965. [Online]. Available: https://doi.org/10.1080/00224545.1965.9919618

[88] D. P. Schultz, "Panic in the military," Department of Psychology, University of North Carolina, Report, 1971.

[89] N. R. Johnson, "Panic and the breakdown of social order: Popular myth, social theory, empirical evidence," *Sociological Focus*, vol. 20, no. 3, pp. 171–183, 1987. [Online]. Available: http://dx.doi.org/10.1080/00380237.1987.10570950

[90] P. Gantt and R. Gantt, "Disaster psychology: Dispelling the myths of panic," *Professional Safety*, 2012.

[91] R. d. S. Franca, M. d. G. B. Marietto, and M. B. Steinberger, "Proposing a cognitive multi-agent model for the panic in crowds phenomenon," in *2009 Second International Conference on the Applications of Digital Information and Web Technologies*, Conference Proceedings, pp. 737–742.

[92] A. R. Mawson, "Understanding mass panic and other collective responses to threat and disaster," *Psychiatry: Interpersonal and Biological Processes*, vol. 68, no. 2, pp. 95–113, 2005.

[93] J. D. Sime, "Movement toward the familiar: Person and place affiliation in a fire entrapment setting," *Environment and Behavior*, vol. 17, no. 6, pp. 697–724, 1985. [Online]. Available: https://doi.org/10.1177/0013916585176003

[94] J. C. Turner, P. J. Oakes, S. A. Haslam, and C. McGarty, "Self and collective: Cognition and social context," *Personality and Social Psychology Bulletin*, vol. 20, no. 5, pp. 454–463, 1994.

[95] J. Drury, C. Cocking, and S. D. Reicher, "The nature of collective resilience: Survivor reactions to the 2005 london bombings," *International Journal of Mass Emergencies and Disasters*, vol. 27, pp. 66–95, 2009.

[96] C. Stott and J. Drury, "Crowds, context and identity: Dynamic categorization processes in the 'poll tax riot'," *Human Relations*, vol. 53, no. 2, pp. 247–273, 2000. [Online]. Available: https://doi.org/10.1177/a010563

[97] M. K. Lindell and R. W. Perry, "The protective action decision model: Theoretical modifications and additional evidence," *Risk Analysis*, vol. 32, no. 4, pp. 616–632, 2012. [Online]. Available: http://dx.doi.org/10.1111/j.1539-6924.2011.01647.x

[98] W. E. Feinberg and N. R. Johnson, "The ties that bind: A macro-level approach to panic," *International journal of mass emergencies and disasters*, vol. 19, no. 3, pp. 269–295, 2001. [Online]. Available: http://dx.doi.org/

[99] A. D. Baddeley, "Selective attention and performance in dangerous environments," *Journal of Human Performance in Extreme Environments*, vol. 5, no. 1, 2000.

[100] E. Kuligowski, "Terror defeated: Occupant sensemaking, decision-making and protective action in the 2001 world trade center disaster," Thesis, 2011. [Online]. Available: https://scholar.colorado.edu/concern/graduate_thesis_or_dissertations/6t053g11g

[101] W. James, "II—what is an emotion ?" *Mind*, vol. os-IX, no. 34, pp. 188–205, 1884. [Online]. Available: http://dx.doi.org/10.1093/mind/os-IX.34.188

[102] R. McCarty, *The Fight-or-Flight Response: A Cornerstone of Stress Research*, 2016, pp. 33–37.

[103] P. Ekman, *Emotion in the human face*, 2nd ed. Cambridge: Cambridge University Press, 1982.

[104] S. Notzon, S. Deppermann, A. Fallgatter, J. Diemer, A. Kroczek, K. Domschke, P. Zwanzger, and A. C. Ehlis, "Psychophysiological

effects of an itbs modulated virtual reality challenge including participants with spider phobia," *Biological Psychology*, vol. 112, no. Supplement C, pp. 66–76, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301051115300648

[105] M. M. Bradley and P. J. Lang, *Measuring emotion: Behavior, feeling, and physiology.* Oxford University Press, 2000.

[106] L. M. Romero, *Fight or Flight Responses.* [Amsterdam]: Elsevier B.V., 2010, pp. 710–714.

[107] M. Mortillaro and M. Mehu, *Emotions: Methods of Assessment*, 2015, pp. 519–525.

[108] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0005791694900639

[109] K. Starcke and M. Brand, "Decision making under stress: a selective review," *Neurosci Biobehav Rev*, vol. 36, no. 4, pp. 1228–48, 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/22342781

[110] J. A. Healey, "Wearable and automotive systems for affect recognition from physiology," Thesis, 2000, massachusetts Institute of Technology. [Online]. Available: http://hdl.handle.net/1721.1/9067

[111] R. W. Levenson, "Autonomic nervous system differences among emotions," *Psychological Science*, vol. 3, no. 1, pp. 23–27, 1992. [Online]. Available: https://doi.org/10.1111/j.1467-9280.1992.tb00251.x

[112] J. T. Cacioppo and L. G. Tassinary, "Inferring psychological significance from physiological signals," *Am Psychol*, vol. 45, no. 1, pp. 16–28, 1990.

[113] S. Kreibig, F. Wilhelm, W. T Roth, and J. J Gross, *Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films*, 2007, vol. 44.

[114] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cognition and emotion*, vol. 23, no. 2, pp. 209–237, 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2756702/

[115] N. Holt, *Psychology : the science of mind and behaviour*, 2nd ed. London: McGraw-Hill Higher Education, 2012.

[116] D. H. Hellhammer, S. Wüst, and B. M. Kudielka, "Salivary cortisol as a biomarker in stress research," *Psychoneuroendocrinology*, vol. 34, no. 2, pp. 163–171, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306453008002990

[117] C. Harmon-Jones, B. Bastian, and E. Harmon-Jones, "The discrete emotions questionnaire: A new tool for measuring state self-reported emotions," *PLOS ONE*, vol. 11, no. 8, p. e0159915, 2016. [Online]. Available: https://doi.org/10.1371/journal.pone.0159915

[118] T. M. Marteau and H. Bekker, "The development of a six-item short-form of the state scale of the spielberger state-trait anxiety inventory (stai)," *Br J Clin Psychol*, vol. 31, no. 3, pp. 301–6, 1992.

[119] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the panas scales," *J Pers Soc Psychol*, vol. 54, no. 6, pp. 1063–70, 1988.

[120] M. Hersen, "Self-assessment of fear," *Behavior Therapy*, vol. 4, no. 2, pp. 241–257, 1973. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005789473800346

[121] J. H. Geer, "The development of a scale to measure fear," *Behaviour Research and Therapy*, vol. 3, no. 1, pp. 45–53, 1965. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0005796765900409

[122] J. W. Paul, "Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala," *Current Directions in Psychological Science*, vol. 7, no. 6, pp. 177–188, 1998. [Online]. Available: http://journals.sagepub.com/doi/abs/10.1111/1467-8721.ep10836912

[123] K. J. Ressler, "Amygdala activity, fear, and anxiety: modulation by stress," *Biol Psychiatry*, vol. 67, no. 12, pp. 1117–9, 2010.

[124] M. Yang, Z. Yang, T. Yuan, W. Feng, and P. Wang, "A systemic review of functional near-infrared spectroscopy for stroke: Current application and future directions," *Frontiers in Neurology*, vol. 10, p. 58, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fneur.2019.00058

[125] L. Hernández-Ronquillo, L. Thorpe, D. Dash, T. Hussein, G. Hunter, K. Waterhouse, P. Laboni Roy, and J. F. Téllez-Zenteno, "Diagnostic accuracy of the ambulatory eeg vs. routine eeg for first single unprovoked seizures and seizure recurrence: The dx-seizure study," *Frontiers in neurology*, vol. 11, pp. 223–223, 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32328023https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7160330/

[126] S. Balters and M. Steinert, "Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices," *Journal of Intelligent Manufacturing*, vol. 28, no. 7, pp. 1585–1607, 2017. [Online]. Available: https://doi.org/10.1007/s10845-015-1145-2

[127] C. F. Camerer and R. M. Hogarth, "The effects of financial incentives in experiments: A review and capital-labor-production framework," *Journal of Risk and Uncertainty*, vol. 19, no. 1, pp. 7–42, 1999. [Online]. Available: https://doi.org/10.1023/A:1007850605129

[128] M. Haghani and M. Sarvi, "Crowd behaviour and motion: Empirical methods," *Transportation Research Part B: Methodological*, vol. 107, pp. 253–294, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0191261517303788

[129] G. Lawson, "Predicting human behaviour in emergencies." Thesis, 2011, university of Nottingham.

[130] M. Haghani, "Empirical methods in pedestrian, crowd and evacuation dynamics: Part i. experimental methods and emerging topics," *Safety Science*, vol. 129, p. 104743, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753520301405

[131] M. Haghani, "Empirical methods in pedestrian, crowd and evacuation dynamics: Part ii. field methods and controversial topics," *Safety Science*, vol. 129, 2020.

[132] D. Kučera and J. Haviger, "Using mood induction procedures in psychological research," *Procedia - Social and Behavioral Sciences*, vol. 69, pp. 31–40, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877042812053669

[133] A. P. Allen, P. J. Kennedy, J. F. Cryan, T. G. Dinan, and G. Clarke, "Biological and psychological markers of stress in humans: Focus on the trier social stress test," *Neuroscience and Biobehavioral Reviews*, vol. 38, pp. 94–124, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0149763413002728

[134] M. Buckert, J. Oechssler, and C. Schwieren, "Imitation under stress," *Journal of Economic Behavior and Organization*, vol. 139, pp. 252–266, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167268117301154

[135] M. Buckert, C. Schwieren, B. M. Kudielka, and C. J. Fiebach, "How stressful are economic competitions in the lab? an investigation with physiological measures," *Journal of Economic Psychology*, vol. 62, pp. 231–245, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167487016304238

[136] J. Marín-Morales, C. Llinares, J. Guixeres, and M. Alcañiz, "Emotion recognition in immersive virtual reality: From statistics to affective computing," *Sensors*, vol. 20, no. 18, 2020.

[137] Second life. Accessed on 08/02/2021. [Online]. Available: https://secondlife.com/

[138] A. Kageyama and A. Tomiyama, "Visualization framework for cave virtual reality systems," *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 07, no. 04, p. 1643001, 2016. [Online]. Available: https://doi.org/10.1142/S1793962316430017

[139] S. Manjrekar, S. Sandilya, D. Bhosale, S. Kanchi, A. Pitkar, and M. Gondhalekar, "Cave: An emerging immersive technology – a review,"

in *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, Conference Proceedings, pp. 131–136.

[140] T. J. Buker, D. A. Vincenzi, and J. E. Deaton, "The effect of apparent latency on simulator sickness while using a see-through helmet-mounted display: reducing apparent latency with predictive compensation," *Hum Factors*, vol. 54, no. 2, pp. 235–49, 2012.

[141] E. C. Regan and K. R. Price, "The frequency of occurrence and severity of side-effects of immersion virtual reality," *Aviat Space Environ Med*, vol. 65, no. 6, pp. 527–30, 1994.

[142] D. Treanor, N. Jordan-Owers, J. Hodrien, J. Wood, P. Quirke, and R. A. Ruddle, "Virtual reality powerwall versus conventional microscope for viewing pathology slides: an experimental comparison," *Histopathology*, vol. 55, no. 3, pp. 294–300, 2009.

[143] R. Ball, C. North, and D. A. Bowman, "Move to improve: promoting physical navigation to increase user performance with large displays," p. 191–200. [Online]. Available: https://doi.org/10.1145/1240624.1240656

[144] Z. Feng, V. A. González, R. Amor, R. Lovreglio, and G. Cabrera-Guerrero, "Immersive virtual reality serious games for evacuation training and research: A systematic literature review," *Computers and Education*, vol. 127, pp. 252–266, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360131518302380

[145] M. Panahi-Shahri, A. Fathi-Ashtiani, P. Azadfallah, and A. Gh., "Reliability and validity of igroup presence questionnaire (ipq)," *International Journal of Behavioral Sciences*, vol. 3, pp. 27–34, 2009.

[146] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence: Teleoper. Virtual Environ.*, vol. 7, no. 3, p. 225–240, 1998. [Online]. Available: https://doi.org/10.1162/105474698565686

[147] M. Meehan, "Physiological reaction as an objective measure of presence in virtual environments," University of North Carolina at Chapel Hill, Report, 2001.

[148] M. Meehan, S. Razzaque, B. Insko, M. Whitton, and J. F. Brooks, *Review of Four Studies on the Use of Physiological Reaction as a Measure of Presence in StressfulVirtual Environments*, 2005, vol. 30.

[149] A. Brogni, V. Vinayagamoorthy, A. Steed, and M. Slater, "Variations in physiological responses of participants during different stages of an immersive virtual environment experiment," in *Proceedings of the ACM symposium on Virtual reality software and technology.* 1180572: ACM, Conference Proceedings, pp. 376–382.

[150] V. Vinayagamoorthy, "Participant responses to virtual agents in immersive virtual environments," Thesis, 2006, University of London.

[151] M. Slater, A. Rovira, R. Southern, D. Swapp, J. J. Zhang, C. Campbell, and M. Levine, "Bystander responses to a violent incident in an immersive virtual environment," *PLOS ONE*, vol. 8, no. 1, p. e52766, 2013. [Online]. Available: https://doi.org/10.1371/journal.pone.0052766

[152] M. Moussaïd, M. Kapadia, T. Thrash, R. W. Sumner, M. Gross, D. Helbing, and C. Hölscher, "Crowd behaviour during high-stress evacuations in an immersive virtual environment," *Journal of the Royal Society, Interface*, vol. 13, no. 122, 2016. [Online]. Available: http://europepmc.org/abstract/MED/27605166http://europepmc.org/articles/PMC5046946?pdf=renderhttp://europepmc.org/articles/PMC5046946https://doi.org/10.1098/rsif.2016.0414

[153] M. van den Berg, "The influence of herding on departure choice in case of an evacuation: Design and analysis of a serious gaming experimental set-up," Thesis, 2016, delft University of Technology. [Online]. Available: https://books.google.co.uk/books?id=EWVxAQAACAAJ

[154] J. Lin, R. Zhu, N. Li, and B. Becerik-Gerber, "Do people follow the crowd in building emergency evacuation? a cross-cultural immersive virtual reality-based study," *Advanced Engineering Informatics*, vol. 43, p. 101040, 2020.

[155] A. Mintz, "Non-adaptive group behavior," *Journal of Anormal Psychology*, vol. 46, no. 2, pp. 150–159, 1951.

[156] H. Kelley, J. Condry, A. Dahlke, and A. Hill, "Collective behavior in a simulated panic situation," *Journal of Experimental Social Psychology*, vol. 1, no. 1, pp. 20–54, 1965. [Online]. Available: http://dx.doi.org/10.1016/0022-1031(65)90035-1

[157] G. Keinan, N. Friedland, and Y. Ben-Porath, "Decision making under stress: Scanning of alternatives under physical threat," *Acta Psychologica*, vol. 64, no. 3, pp. 219–228, 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0001691887900084

[158] J. Drury, C. Cocking, S. Reicher, A. Burton, D. Schofield, A. Hardwick, D. Graham, and P. Langston, "Cooperation versus competition in a mass emergency evacuation: a new laboratory simulation and a new theoretical model," *Behav Res Methods*, vol. 41, no. 3, pp. 957–70, 2009. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/19587213

[159] Y. Bar-Haim, A. Kerem, D. Lamy, and D. Zakay, "When time slows down: The influence of threat on time perception in anxiety," *Cognition and Emotion*, vol. 24, no. 2, pp. 255–263, 2010.

[160] W. Daamen and S. P. Hoogendoorn, "Calibration of pedestrian simulation model for emergency doors by pedestrian type," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2316, pp. 69–75, 2012. [Online]. Available: https://doi.org/10.3141/2316-08

[161] W. Daamen and S. P. Hoogendoorn, "Emergency door capacity: Influence of door width, population composition and stress level," *Fire Technology*, vol. 48, no. 1, pp. 55–71, 2012. [Online]. Available: https://doi.org/10.1007/s10694-010-0202-9

[162] M. Haghani and M. Sarvi, "Stated and revealed exit choices of pedestrian crowd evacuees," *Transportation Research Part B: Methodological*, vol. 95, pp. 238–259, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0191261516306762

[163] M. Haghani and M. Sarvi, "Social dynamics in emergency evacuations: Disentangling crowd's attraction and repulsion effects," *Physica A: Statistical Mechanics and its Applications*, vol. 475, pp. 24–34, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437117301243

[164] C. Dias, M. Sarvi, O. Ejtemai, and M. Burd, "Elevated desired speed and change in desired direction," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2490, pp. 65–75, 2015. [Online]. Available: https://doi.org/10.3141/2490-08

[165] N. Shiwakoti, M. Sarvi, G. Rose, and M. Burd, "Biologically inspired modeling approach for collective pedestrian dynamics under emergency conditions," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2196, pp. 176–184, 2010.

[166] E. Altshuler, O. Ramos, Y. Núñez, J. Fernández, A. J. Batista-Leyva, and C. Noda, "Symmetry breaking in escaping ants," *The American Naturalist*, vol. 166, no. 6, pp. 643–649, 2005. [Online]. Available: http://www.journals.uchicago.edu/doi/abs/10.1086/498139

[167] C. Dias, M. Sarvi, N. Shiwakoti, and M. Burd, "Turning angle effect on emergency egress," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2312, pp. 120–127, 2012. [Online]. Available: https://doi.org/10.3141/2312-12

[168] N. Shiwakoti, M. Sarvi, G. Rose, and M. Burd, "Consequence of turning movements in pedestrian crowds during emergency egress," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2234, pp. 97–104, 2011. [Online]. Available: https://doi.org/10.3141/2234-11

[169] C. Dias, M. Sarvi, N. Shiwakoti, O. Ejtemai, and M. Burd, "Investigating collective escape behaviours in complex situations," *Safety Science*, vol. 60, pp. 87–94, 2013.

[170] Z. Shahhoseini, M. Sarvi, and M. Saberi, "Insights toward characteristics of merging streams of pedestrian crowds based on experiments with panicked ants," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2561, pp. 81–88, 2016. [Online]. Available: https://doi.org/10.3141/2561-10

[171] Z. Shahhoseini, M. Sarvi, M. Saberi, and M. Haghani, "Pedestrian crowd dynamics observed at merging sections," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2622, pp. 48–57, 2017. [Online]. Available: https://doi.org/10.3141/2622-05

[172] A. Sobhani, M. Sarvi, D. Duives, O. Ejtemai, K. Aghabayk, and S. Hoogendoorn, "Exploring the relationship of exit flow and jam density in panic scenarios using animal dynamics," *Transportation Research Procedia*, vol. 2, pp. 745–751, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352146514001197

[173] I. Zuriguel, J. Olivares, J. M. Pastor, C. Martín-Gómez, L. M. Ferrer, J. J. Ramos, and A. Garcimartín, "Effect of obstacle position in the flow of sheep through a narrow door," *Physical Review E*, vol. 94, no. 3, p. 032302, 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.94.032302

[174] N. Shiwakoti and M. Sarvi, "Enhancing the panic escape of crowd through architectural design," *Transportation Research Part C: Emerging Technologies*, vol. 37, pp. 260–267, 2013.

[175] M. Kobes, I. Helsloot, B. de Vries, J. G. Post, N. Oberijé, and K. Groenewegen, "Way finding during fire evacuation; an analysis of unannounced fire drills in a hotel at night," *Building and Environment*, vol. 45, no. 3, pp. 537–548, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360132309001759

[176] G. Proulx and J. D. Sime, "To prevent 'panic' in an underground emergency: Why not tell people the truth?" in *Fire Safety Science- Proceedings of the Third International Symposium*, Conference Proceedings, pp. 843–852.

[177] R. Zinke, G. Hofinger, and L. Künzer, *Psychological Aspects of Human Dynamics in Underground Evacuation: Field Experiments*. Cham: Springer International Publishing, 2014, pp. 1149–1162. [Online]. Available: https://doi.org/10.1007/978-3-319-02447-9_94

[178] M. Boltes and A. Seyfried, "Collecting pedestrian trajectories," *Neurocomputing*, vol. 100, pp. 127–133, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231212003189

[179] Z. Yucel, F. Zanlungo, T. Ikeda, T. Miyashita, and N. Hagita, "Deciphering the crowd: modeling and identification of pedestrian group motion," *Sensors (Basel)*, vol. 13, no. 1, pp. 875–97, 2013.

[180] N. Shiwakoti, "Understanding differences in emergency escape and experimental pedestrian crowd egress through quantitative comparison," *International Journal of Disaster Risk Reduction*, vol. 20, pp. 129–137, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2212420916303946

[181] Z. Gu, Z. Liu, N. Shiwakoti, and M. Yang, "Video-based analysis of school students' emergency evacuation behavior in earthquakes," *International Journal of Disaster Risk Reduction*, vol. 18, pp. 1–11, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2212420916300371

[182] G. Bernardini, E. Quagliarini, and M. D'Orazio, "Towards creating a combined database for earthquake pedestrians' evacuation models," *Safety Science*, vol. 82, pp. 77–94, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925753515002301

[183] M. Haghani and M. Sarvi, "Human exit choice in crowded built environments: Investigating underlying behavioural differences between normal egress and emergency evacuations," *Fire Safety Journal*, vol. 85, pp. 1–9, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0379711216300844

[184] N. Shiwakoti, R. Tay, P. Stasinopoulos, and P. J. Woolley, "Likely behaviours of passengers under emergency evacuation in train station," *Safety Science*, vol. 91, no. Supplement C, pp. 40–48, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925753516301588

[185] M. Haghani, M. Sarvi, Z. Shahhoseini, and M. Boltes, "How simple hypothetical-choice experiments can be utilized to learn humans' navigational escape decisions in emergencies," *PloS one*, vol. 11, no. 11, pp. e0 166 908–e0 166 908, 2016. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27870880https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5117746/

[186] W. Laqueur, *No end to war : terrorism in the twenty-first century*. New York ; London: Continuum, 2003.

[187] S. G. Jones, C. Doxsee, N. Harrington, G. Hwang, and J. Suber, "The war comes home: The evolution of domestic terrorism in the united states," Center for Strategic and International Studies, Report, 2020. [Online]. Available: https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/201021_Jones_War_Comes_Home_v2.pdf

[188] D. Niewart, D. Ankrom, E. Kaplan, and S. Pham. Homegrown terror. Accessed on 11/2/2021. [Online]. Available: https://apps.revealnews.org/homegrown-terror/

[189] C. Heath-Kelly and L. Fernández de Mosteyrín, "The political use of victimhood: Spanish collective memory of eta through the war on terror paradigm," *Review of International Studies*, vol. 47, no. 1, pp. 1–18, 2021. [Online]. Available: https://www.cambridge.org/core/article/political-use-of-victimhood-spanish-collective-memory-of-eta-through-the-war-on-terror-paradigm/9BF28C54D60976921666C28F6DD960C7

[190] A. P. Schmid, *The Routledge Handbook of Terrorism Research*, ser. The Routledge Handbook of Terrorism Research. Routledge, 2011. [Online]. Available: https://www.routledgehandbooks.com/doi/10.4324/9780203828731

[191] D. O. Dusenberry, *Handbook for blast-resistant design of buildings.* Hoboken, N.J.: J. Wiley, 2010.

[192] P. Eachus, A. Stedmon, and L. Baillie, "Hostile intent in public crowded spaces: A field study," *Applied Ergonomics*, vol. 44, no. 5, pp. 703–709, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003687012000737

[193] NPCC. Run, hide, tell. Accessed on 05/12/2017. [Online]. Available: http://www.npcc.police.uk/NPCCBusinessAreas/WeaponAttacksStaySafe.aspx

[194] A. P. Schmid and A. J. Jongman, *Political terrorism : a new guide to actors, authors, concepts, data bases, theories and literature.* Amsterdam: North-Holland Publ. Co. [u.a.], 1984.

[195] A. Silke, “The devil you know: Continuing problems with research on terrorism,” *Terrorism and Political Violence*, vol. 13, no. 4, pp. 1–14, 2001. [Online]. Available: https://doi.org/10.1080/09546550109609697

[196] P. R. C. Limited, “Terrorism threat and mitigation report (january - july 2017),” Pool Re’s Terrorism Research and Analysis Centre, Report, 25 September 2017 2017.

[197] A. Bryman and E. Bell, *Social research methods*, 4th ed. Oxford University Press, 2016.

[198] A. Milivojevich and T. Barker, *Quality by Experimental Design, Fourth Edition*. Taylor & Francis, 2015. [Online]. Available: https://books.google.co.uk/books?id=3_BYrgEACAAJ

[199] R. Gersten, S. Baker, and J. W. Lloyd, “Designing high-quality research in special education: Group experimental design,” *The Journal of Special Education*, vol. 34, no. 1, pp. 2–18, 2000. [Online]. Available: https://doi.org/10.1177/002246690003400101

[200] B. E. R. A. (BERA), *Ethical guidelines for educational research.*, 4th ed., 2018. [Online]. Available: https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018

[201] G. Comstock, *Research Ethics: A Philosophical Guide to the Responsible Conduct of Research*. Cambridge: Cambridge University Press, 2013. [Online]. Available: https://www.cambridge.org/core/books/research-ethics/B97EB5CDFDAB6079ECDF0ECE12894D4B

[202] M. Humphreys, “Reflections on the ethics of social experimentation,” *Journal of Globalization and Development*, vol. 6, no. 1, pp. 87–112, 2015. [Online]. Available: https://doi.org/10.1515/jgd-2014-0016

[203] C. f. P. o. N. Infrastructure. Terrorism. Accessed on 08/02/2021. [Online]. Available: https://www.cpni.gov.uk/terrorism

[204] BBC. Salisbury declared decontaminated after novichok poisoning. Accessed on 08/02/2021. [Online]. Available: https://www.bbc.co.uk/news/uk-england-wiltshire-47412390

[205] P. G. Zimbardo, R. L. Johnson, and V. McCann, *Psychology : core concepts*, 2017.

[206] S. Milgram, "Behavioral study of obedience," *The Journal of Abnormal and Social Psychology*, vol. 67, no. 4, pp. 371–378, 1963.

[207] B. von Dawans, C. Kirschbaum, and M. Heinrichs, "The trier social stress test for groups (tsst-g): A new research tool for controlled simultaneous social stress exposure in a group format," *Psychoneuroendocrinology*, vol. 36, no. 4, pp. 514–522, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.psyneuen.2010.08.004

[208] S. Arias, J. Wahlqvist, D. Nilsson, E. Ronchi, and H. Frantzich, "Pursuing behavioral realism in virtual reality for fire evacuation research," *Fire and Materials*, vol. n/a, no. n/a, 2020. [Online]. Available: https://doi.org/10.1002/fam.2922

[209] B. Latané and J. M. Darley, "Bystander 'apathy'," *American Scientist*, vol. 57, no. 2, pp. 244–268, 1969. [Online]. Available: http://www.jstor.org/stable/27828530

[210] R. Lovreglio, V. Gonzalez, Z. Feng, R. Amor, M. Spearpoint, J. Thomas, M. Trotter, and R. Sacks, "Prototyping virtual reality serious games for building earthquake preparedness: The auckland city hospital case study," *Advanced Engineering Informatics*, vol. 38, 2018.

[211] J. K. Haas, "A history of the unity game engine," Conference Proceedings. [Online]. Available: https://digital.wpi.edu/show/tx31qh96p

[212] E. Games. Unreal engine. Accessed on 12/02/2021. [Online]. Available: https://www.unrealengine.com

[213] R. Lovreglio. How to create, rig and animate an human character using fuse, mixamo and unity - part 1. Accessed on 12/02/2021. [Online]. Available: https://www.youtube.com/watch?v=0W_81PjS3Zc

[214] S. Lay, N. Brace, G. Pike, and F. Pollick, "Circling around the uncanny valley: Design principles for research into the relation between human likeness and eeriness," *i-Perception*, vol. 7, no. 6, 2016. [Online]. Available: https://doi.org/10.1177/2041669516681309

[215] S. Weech, S. Kenny, and M. Barnett-Cowan, "Presence and cybersickness in virtual reality are negatively related: A review," *Frontiers in Psychology*, vol. 10, p. 158, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2019.00158

[216] M. Nabiyouni, A. Saktheeswaran, D. Bowman, and A. Karanth, *Comparing the Performance of Natural, Semi-Natural, and Non-Natural Locomotion Techniques in Virtual Reality*, 2015.

[217] D. Saredakis, A. Szpak, B. Birckhead, H. A. D. Keage, A. Rizzo, and T. Loetscher, "Factors associated with virtual reality sickness in head-mounted displays: A systematic review and meta-analysis," *Frontiers in human neuroscience*, vol. 14, pp. 96–96, 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32300295https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7145389/

[218] M. E. S. Pierre, S. Banerjee, A. W. Hoover, and E. R. Muth, "The effects of 0.2 hz varying latency with 20–100 ms varying amplitude on simulator sickness in a helmet mounted display," *Displays*, vol. 36, pp. 1–8, 2015.

[219] M. Meehan, S. Razzaque, M. Whitton, and J. F. Brooks, *Effect of Latency on Presence in Stressful Virtual Environments*, 2003, vol. 2003.

[220] C. Boletsis, "The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology," *Multimodal Technologies and Interaction*, vol. 1, no. 4, 2017.

[221] T. J. Renshaw, N. A. Sonnenfeld, and M. D. Meyers, "Fundamentals for a turing test of virtual reality," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1, pp. 2113–2117, 2016. [Online]. Available: https://doi.org/10.1177/1541931213601478

[222] M. Roohi-Azizi, L. Azimi, S. Heysieattalab, and M. Aamidfar, "Changes of the brain's bioelectrical activity in cognition, consciousness, and some mental disorders," *Medical journal of the Islamic Republic of Iran*, vol. 31, pp. 53–53, 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29445682https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5804435/

[223] Accessed on 12/02/2021. [Online]. Available: https://nirx.net/nirsport

[224] J. Shi, Y. Yao, C. Zhan, Z. Mao, F. Yin, and X. Zhao, "The relationship between big five personality traits and psychotic experience in a large non-clinical youth sample: The mediating role of emotion regulation," *Frontiers in psychiatry*, vol. 9, pp. 648–648, 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30564151https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6288374/

[225] A. Furnham, *Myers-Briggs Type Indicator (MBTI)*, ser. Encyclopedia of Personality and Individual Differences. Cham: Springer International Publishing, 2017. [Online]. Available: https://doi.org/10.1007/978-3-319-28099-8_50-1

[226] D. P. McAdams, *Personality Psychology.* Dordrecht: Springer Netherlands, 2013, pp. 1634–1641. [Online]. Available: https://doi.org/10.1007/978-1-4020-8265-8_857

[227] H. Carter, J. Drury, G. J. Rubin, R. Williams, and R. Amlot, "Public experiences of mass casualty decontamination," *Biosecur Bioterror*, vol. 10, no. 3, pp. 280–9, 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/22823588

[228] W. Leach Colin, M. R. Mosquera Patricia, L. W. Vliek Michael, and E. Hirt, "Group devaluation and group identification," *Journal of Social Issues*, vol. 66, no. 3, pp. 535–552, 2010. [Online]. Available: https://doi.org/10.1111/j.1540-4560.2010.01661.x

[229] S. S. Khan, N. Hopkins, S. Reicher, S. Tewari, N. Srinivasan, and C. Stevenson, "Shared identity predicts enhanced health at a mass gathering," *Group Processes & Intergroup Relations*, vol. 18, no. 4, pp. 504–522, 2014. [Online]. Available: https://doi.org/10.1177/1368430214556703

[230] e. Empatica. Accessed on 12/02/2021. [Online]. Available: https://www.empatica.com/

[231] E. Equivital. Accessed on 12/02/2021. [Online]. Available: https://www.equivital.com/

[232] P. H10. Accessed on 12/02/2021. [Online]. Available: https://www.polar.com/uk-en/products/accessories/polar_h10_heart_rate_sensor

[233] Shimmer. Accessed on 12/02/2021. [Online]. Available: http://www.shimmersensing.com/products/consensys-ecg-development-kits-update

[234] Movisens. Accessed on 12/02/2021. [Online]. Available: https://www.movisens.com/en/products/eda-and-activity-sensor/

[235] K. Schwab, G. Heubel, and H. Bartels, "Free epinephrine, norepinephrine and dopamine in saliva and plasma of healthy adults," *European journal of clinical chemistry and clinical biochemistry : journal of the Forum of European Clinical Chemistry Societies*, vol. 30 9, pp. 541–4, 1992.

[236] E. T. Solovey, D. Afergan, E. M. Peck, S. W. Hincks, and R. J. K. Jacob, "Designing implicit interfaces for physiological computing: Guidelines and lessons learned using fnirs," *ACM Trans. Comput.-Hum. Interact.*, vol. 21, no. 6, p. Article 35, 2015. [Online]. Available: https://doi.org/10.1145/2687926

[237] B. Burle, L. Spieser, C. Roger, L. Casini, T. Hasbroucq, and F. Vidal, "Spatial and temporal resolutions of eeg: Is it really black and white? a scalp current density view," *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, vol. 97, no. 3, pp. 210–220, 2015. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/25979156https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4548479/

[238] S. P. Kyathanahally, Y. Wang, V. D. Calhoun, and G. Deshpande, "Investigation of true high frequency electrical substrates of fmri-based resting state networks using parallel independent component analysis of simultaneous eeg/fmri data," *Frontiers in Neuroinformatics*, vol. 11, p. 74, 2017. [Online]. Available: https://www.frontiersin.org/article/10.3389/fninf.2017.00074

[239] iPhone user guide. Accessed on 16/03/2021. [Online]. Available: https://support.apple.com/en-gb/guide/iphone/iph771fd0aad/ios

[240] M. Cominelli, P. Patras, and F. Gringoli, "Dead on arrival: An empirical study of the bluetooth 5.1 positioning system," in *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization.* Association for Computing Machinery, Conference Proceedings, p. 13–20. [Online]. Available: https://doi.org/10.1145/3349623.3355475

[241] B. Huang, J. Liu, W. Sun, and F. Yang, "A robust indoor positioning method based on bluetooth low energy with separate channel information," *Sensors (Basel)*, vol. 19, no. 16, 2019.

[242] G. Bradski, "The opencv library," *Dr Dobb's J. Software Tools*, vol. 25, pp. 120–125, 2000. [Online]. Available: https://ci.nii.ac.jp/naid/10028167478/en/

[243] Codamotion. Accessed on 12/02/2021. [Online]. Available: https://codamotion.com/

[244] M. Shadow. Accessed on 12/02/2021. [Online]. Available: https://www.motionshadow.com/vr

[245] L. Toleikyte and P. H. England, "Equality in public health england," Public Health England, Report, 2017.

[246] I. Perogamvros, L. J. Owen, B. G. Keevil, G. Brabant, and P. J. Trainer, "Measurement of salivary cortisol with liquid chromatography-tandem mass spectrometry in patients undergoing dynamic endocrine testing," *Clin Endocrinol (Oxf)*, vol. 72, no. 1, pp. 17–21, 2010.

[247] T. Van de Mortel, "Faking it: Social desirability response bias in self-report research," *Australian Journal of Advanced Nursing*, vol. 25, pp. 40–48, 2008.

[248] P. M. Rodriguez Mosquera, A. K. Uskul, and S. E. Cross, "The centrality of social image in social psychology," *European Journal of Social Psychology*, vol. 41, no. 4, pp. 403–410, 2011. [Online]. Available: https://doi.org/10.1002/ejsp.820

[249] J. Drury, C. Cocking, and S. Reicher, "Everyone for themselves? a comparative study of crowd solidarity among emergency survivors," *Br*

J Soc Psychol, vol. 48, no. Pt 3, pp. 487–506, 2009. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/18789185

[250] C. W. E. M. Quaedflieg, "Memory dynamics under stress," *Memory*, vol. 26, pp. 1–13, 2017.

[251] J. D. Payne, E. D. Jackson, S. Hoscheidt, L. Ryan, W. J. Jacobs, and L. Nadel, "Stress administered prior to encoding impairs neutral but enhances emotional long-term episodic memories," *Learning & memory (Cold Spring Harbor, N.Y.)*, vol. 14, no. 12, pp. 861–868, 2007. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18086830https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2151024/

[252] L. Schwabe and O. Wolf, "Learning under stress impairs memory formation," *Neurobiology of learning and memory*, vol. 93, pp. 183–8, 2009.

[253] K. A. Deffenbacher, B. H. Bornstein, S. D. Penrod, and E. K. McGorty, "A meta-analytic review of the effects of high stress on eyewitness memory," *Law and Human Behavior*, vol. 28, no. 6, pp. 687–706, 2004. [Online]. Available: https://doi.org/10.1007/s10979-004-0565-x

[254] E. F. Loftus, "The malleability of human memory: Information introduced after we view an incident can transform memory," *American Scientist*, vol. 67, no. 3, pp. 312–320, 1979. [Online]. Available: http://www.jstor.org/stable/27849223

[255] E. F. Loftus, "Eyewitness testimony," *Applied Cognitive Psychology*, vol. 33, no. 4, pp. 498–503, 2019. [Online]. Available: https://doi.org/10.1002/acp.3542

[256] S.-k. Christianson, "Emotional stress and eyewitness memory: A critical review," *Psychological Bulletin*, vol. 112, no. 2, pp. 284–309, 1992.

[257] S.-k. Christianson and B. Hübinette, "Hands up! a study of witnesses' emotional reactions and memories associated with bank robberies," *Applied Cognitive Psychology*, vol. 7, no. 5, pp. 365–379, 1993. [Online]. Available: https://doi.org/10.1002/acp.2350070502

[258] T. Smeets, T. Giesbrecht, M. Jelicic, and H. Merckelbach, "Context-dependent enhancement of declarative memory performance following

acute psychosocial stress," *Biological psychology*, vol. 76, pp. 116–23, 2007.

[259] O. Hardt, K. Nader, and L. Nadel, "Decay happens: The role of active forgetting in memory," *Trends in Cognitive Sciences*, vol. 17, 2013.

[260] S. Hoscheidt, K. LaBar, L. Ryan, W. Jacobs, and L. Nadel, "Encoding negative events under stress: High subjective arousal is related to accurate emotional memory despite misinformation exposure," *Neurobiology of learning and memory*, vol. 112, 2013.

[261] C. Laney and E. F. Loftus, "Traumatic memories are not necessarily accurate memories," *The Canadian Journal of Psychiatry*, vol. 50, no. 13, pp. 823–828, 2005. [Online]. Available: https://doi.org/10.1177/070674370505001303

[262] D. Strange and M. K. T. Takarangi, "Memory distortion for traumatic events: The role of mental imagery," *Frontiers in Psychiatry*, vol. 6, no. 27, 2015. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyt.2015.00027

[263] J. Shenk and R. Busche, "justinshenk/traja: v0.1.1," https://doi.org/10.5281/zenodo.3237827, p. Pages, online; accessed on 12/02/2021. [Online]. Available: https://doi.org/10.5281/zenodo.3237827

[264] R. Lovreglio, E. Ronchi, and D. Nilsson, "Calibrating floor field cellular automaton models for pedestrian dynamics by using likelihood function optimization," *Physica A: Statistical Mechanics and its Applications*, vol. 438, pp. 308–320, 2015.

[265] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with Python," *Proceedings of the 9th Python in Science Conference*, vol. 2010, 2010.

[266] W. McKinney, "Data structures for statistical computing in Python," *Proceedings of the 9th Python in Science Conference*, 2010.

[267] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and

d. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

[268] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," http://www.scipy.org/, p. Pages, online; accessed on 12/02/2021. [Online]. Available: http://www.scipy.org/

[269] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with numpy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[270] R. van den Bos, M. Harteveld, and H. Stoop, "Stress and decision-making in humans: Performance is related to cortisol reactivity, albeit differently in men and women," *Psychoneuroendocrinology*, vol. 34, no. 10, pp. 1449–1458, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306453009001486

[271] D. Bozovic, M. Racic, and N. Ivkovic, "Salivary cortisol levels as a biological marker of stress reaction," *Medical archives (Sarajevo, Bosnia and Herzegovina)*, vol. 67, no. 5, pp. 374–377, 2013. [Online]. Available: http://europepmc.org/abstract/MED/24601177https://doi.org/10.5455/medarh.2013.67.374-377

[272] C. Carreiras, A. P. Alves, A. Lourenço, F. Canento, H. Silva, A. Fred *et al.*, "BioSPPy: Biosignal processing in Python," https://github.com/PIA-Group/BioSPPy/, p. Pages, online; accessed on 12/02/2021. [Online]. Available: https://github.com/PIA-Group/BioSPPy/

[273] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," in *Proceedings of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*. IOS Press, Conference Proceedings, p. 3–24.

[274] S. Boonnithi and S. Phongsuphap, "Comparison of heart rate variability measures for mental stress detection," in *2011 Computing in Cardiology*, Conference Proceedings, pp. 85–88.

[275] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, pp. 258–258, 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29034226https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5624990/

[276] D. A. Hensher and L. W. Johnson, "Applied discrete-choice modelling," 2018. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1729015

[277] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, p. 1735–1780, 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[278] M. Dehghani, S. Gouws, O. Vinyals, J. Uszkoreit, and L. Kaiser, *Universal Transformers*, 2019. [Online]. Available: https://openreview.net/pdf?id=HyzdRiR9Y7

[279] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[280] J. Amirian, J.-B. Hayet, and J. Pettre, "Social ways: Learning multi-modal distributions of pedestrian trajectories with gans," Conference Proceedings, pp. 2964–2972.

Appendices

A. Normality tests

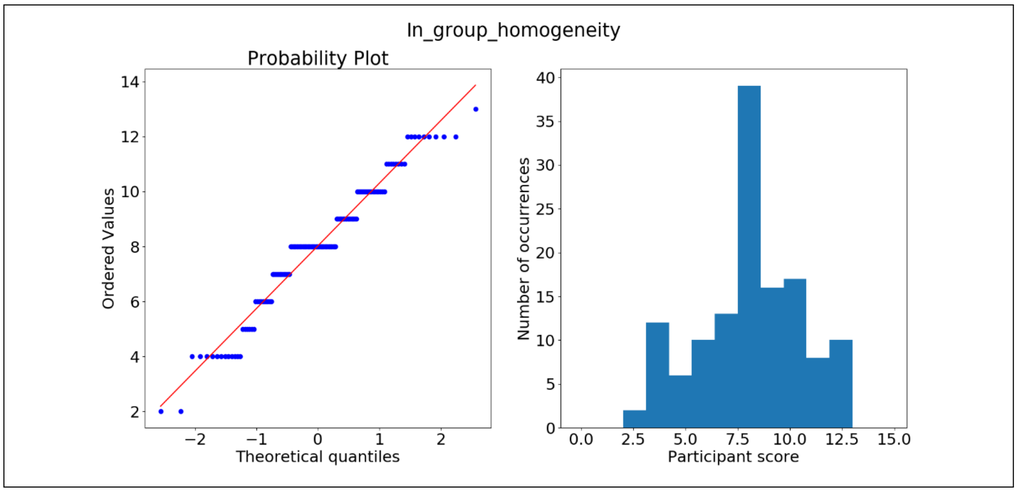


Figure A.1.: Q-Q plot for In-group homogeneity

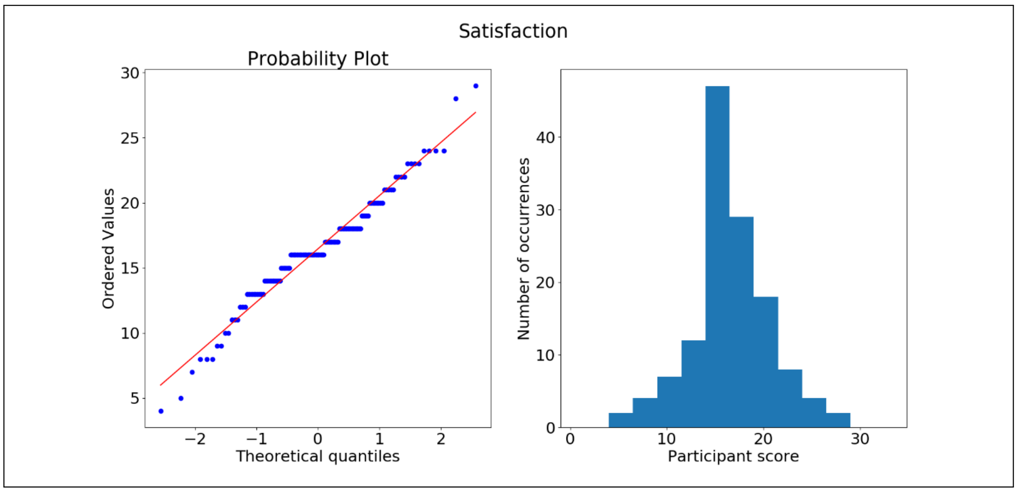


Figure A.2.: Q-Q plot for Satisfaction

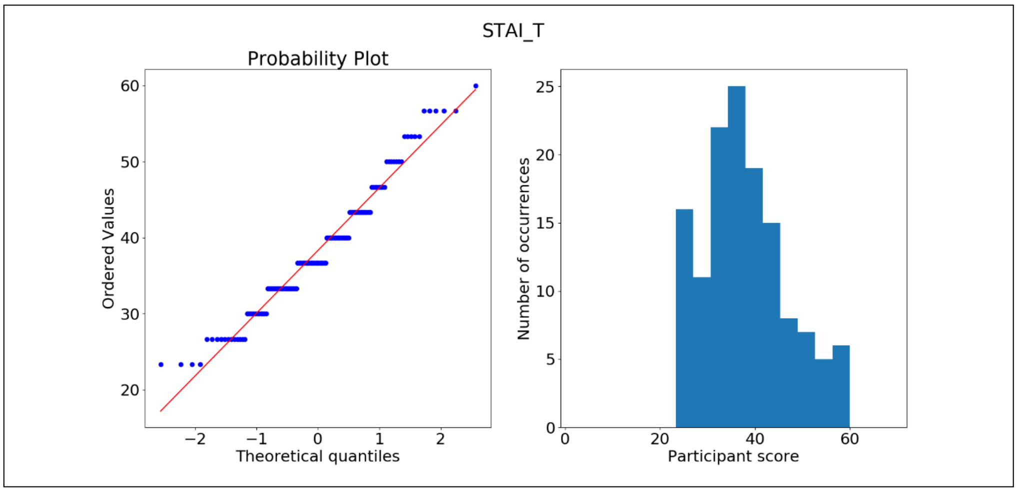


Figure A.3.: Q-Q plot for STAI-T

B. Transcripts

Initials are used to identify individuals, and the timings are provided at key moments throughout the transcript.

B.1. Stakeholder engagement

B.1.1. Int 1

What is your current main section of research/industry?

Consulting specifically consulting in the area of resilience for businesses and cities. My background (also in [workplace]) leading the resilience and counter-terrorism team.

How does your work relate to crowd/pedestrian movement and emergency evacuations?

On the resilience side, it's quite a broad area. That marries up mitigation before an incident, but also response and recovery from incidents. Evacuations would be a small part of that solution. I have someone who works for me in that area. At [workplace] we had a full team working on people flow modelling. Very often as part of the design team you will have a people flow modeller. They have been doing more evacuation as well as just how people get in and out of venues, whether it's in an emergency of whether it's just things like stadia, just trying to manage those crowd flows. But I can see it's certainly very pertinent for looking at in an incident how they would react and move. Definitely the tools are there but I'm not sure about the decision modelling, and how that's done as part of that.

That sort of thing can influence the design, that's the whole point really, you change the size of corridors, you change the number of exits, you can change things to improve the situation. I know you're focusing on MTFA.

That's a real challenge, we actually had a session talking about that, because what can you do? Because traditional guidance regarding counter-terrorism is you want to be able to see and have no places to hide threats and bombs and stuff, but now, because of marauding firearms, you kind of need places to hide. So it's quite hard to know what's the best thing to do, especially with these big open spaces like airports.

How do you incorporate Human behaviour into your work?

More specifically, pedestrian dynamics?

More specifically, pedestrian group dynamics?

I wouldn't have said I do do that much really, I think the only thing I've really touched on it is when we looked at crowd flow in stadia, but it wasn't specific to a terrorist incident. We don't routinely connect the two pieces. But obviously the design is all about people and how people respond to the design, but I think other things like aesthetics tends to takes over, and other things like atmosphere and comfort and those things. But certainly its definitely an area of interest, but I'm not sure if it's something that can currently be done in the sense of decision making and how you respond in that situation.

Might the discussion between human behaviour and aesthetic choices be skewed, as one area is well developed? I'm not sure, as it's out of my realm of expertise... At [workplace] they were using people modelling for optimisation of operations. Like in a hospital looking at the flow of people from one place to another, and how long they wait for certain things, so I wouldn't have said it was about people decision making or group dynamics. However they also do do crowd flow and queueing and stuff like that for getting in and out of stadia.

The most I've seen it for, as it pertains to security, is modelling how long it takes people to get in and out of stadia if they're going through security search... I don't know how much real decision they use as part of that. It is slightly more well developed in the special effects (e.g.: star wars people running away from threats).

How do you incorporate fear in your work?

If your work is related to emergencies, how do you incorporate behavioural responses with respect to this?

I don't.

This is a point of discussion, and has been over the recent months with terrorist attacks (with MTFA). I protect buildings from falling down from bomb blasts, that's a very specific physical mitigation, and it is more difficult to physically mitigate somebody with a gun. And so it does become much more about how people behave, and how can we help that through the design process. Because at the end of the day we're designers of that space. That needs to be factored in, because we need to be like 'what are people likely to do, where can we provide somewhere for them to go'.

I guess we have been incorporating invacuation to quite a few of the buildings we have been designing, so that's actually a requirement or recommendation of the counter-terrorism security advisers. So from that point of view I'd say we do, but we don't specifically think about people being afraid and what they're going to do.

<u>I was talking to [Int 2], who says that understanding of behavioural responses are now required as part of the response planning</u>

As we were talking about before, it's an amygdala reaction, it's a fight or flight response. So you're either going to stand there or you're going to run.

I think from [Int 2's] perspective the fear aspect, it's about managing fear, it's about calming people down, and messaging to make people feel safe. Because propagation of fear is what the terrorists want. The messages that come out after these events is about 'us not hating them, but caring for ourselves', which I think is right.

<u>If you think that you could make use of an improved model of behavioural response to fear, What would be the most useful tool you would use?</u>

I don't think I would use it. It sounds to me like a people flow issue. I would imagine you would build in that decision or behaviour into that model, showing how people move given certain scenarios. You would then use that information to help design the building. So I imagine an architect would be looking at 'well, actually we need another exit from that space or we need to provide security'.

I can see from a security designer and architect point of view you can use that to incorporate this into the design, and it can help influence the design.

It's not well understood at the at all. We're not doing that kind of modelling at the moment in terms of decision making, and basically there is not really any guidance out there on what you should incorporate into your design on the MTFA design case. There is secure by design guidance for general counter-terrorism, which is a requirement on certain projects. There is good practice guidance. There are a few levels here. There is potential for qualitative guidance, there might be some conclusions you can draw that would be very useful guidance for architects. There is another level of sophistication for high profile or crowded spaces, you might be doing some modelling on how people are interacting in really crowded spaces (even in suburban places). That would be quite interesting.

Will the models ever be trusted enough to make a change in design?

I think those types of models have a precedent for being believed, because we do use them on projects, and quite often I would be thinking 'Well how do you know that they're going to do that, or how do you know they're going to go that way?' so there is some level of that in the model at the moment.

I think with more technology, people are used to this type of model (especially if it is visually appealing), so that type of modelling will be embraced from that perspective. You have to understand the uncertainty and sensitivity, because you don't want to get yourself sued, you have to be very clear about that in that type of model.

B.1.2. Int 2

NB: this is not a transcript, this is a summary of notes taken down during interview. Any quotations are identified in quote marks.

What is your current main section of research/industry?

How does your work relate to crowd/pedestrian movement and emergency evacuations? Resilience planning manager at [city] council, responsible for planning major emergencies that can impact the city, and ensuring organisation (council) is able to respond to them.

How do you incorporate Human behaviour into your work?

More specifically, pedestrian dynamics?

More specifically, pedestrian group dynamics?

[Int 2] doesn't use computer pedestrian dynamics software models. However, she does use human behaviour in her work, [basing] planning assumptions on available papers and experiential learning from peers. It is informally part of training for people working in her office.

Not PEDESTRIAN specific, more a general sense (e.g. remember being taught that panic is not a thing, that it is often very rational movement).

It is part of the training in that part of CPD is to keep up with behavioural science. [The amount of human behavioural considerations] depends on which bit of resilience planning [is being considered]. Some people/work fronts are interested in organisational and structural response, others in people centric approach. Both are required to have a rounded approach. The people centric approach is harder and usually gets left out.

There is [now] an expectation in guidance (coming from the Civil Contingencies Act, 2004) that human behaviours is taken into consideration. There is also an expectation that Julia Pearce's work is written into the plan.

Also- public outrage needs to be considered when forming a plan for resilience (eg. Local resilience forum needs to assess various risks in likelihood and impact. Eg. Aircraft crash- assess the impact -economic, environmental, public outrage)

There is a need to consider vulnerable communities and humanitarian assistance- bringing back some of human behaviours.

It is really easy to focus on fixing infrastructure, as 'human stuff is much more hidden'

Therefore human behaviours are harder to keep at the forefront of the approach, easy to ignore the 'softer side'.

GROUP EVACUATION

Evacuation usually means from home, therefore fewer the better. Depending on the event, differing emergency services will have precedence: command and control of the incident is always the police, mgmt. of site is event dependent. Therefore group dynamics will always be at forefront, but differently for different events.

Police would have a very definite approach- treat people often as witnesses. Cf terrorism, risk of perpetrator mixed in people.

MFTA- police in control- ambulance and firemen can (now) follow in. look at DEREK BIRD as a case study.

Ambulances want to triage based on injuries.

How do you incorporate fear in your work?

Human behaviours are taken into consideration- [in the resilience planning process].

It is considered during communication with the public- twitter can now create that fear (through rumour etc). Several different agencies have official twitter feeds that can be used to disseminate accurate information. [There are also multiple different ways of disseminating information, not necessarily to the public- see PubWatch, ShopWatch).

Sharing information (accurate and timely)

Monitoring twitter and social media monitoring

LOOK AT WHAT MANCHESTER DID- POLICE, FIRE, AGMA, co-ordinated communications response, pointing people to accurate sources of info. Read '15 things Manchester did right about terrorist attack'. Twitter is now faster than emergency services, so there is an issue with people getting multiple sources of info.

Alerting system in Leeds- possibly vocal. Allows central information distribution.

[City] uses Pubwatch, Shopwatch, twitter account. Speed is still a question.

Community impacts- largely later on during the emergency.

Checking on tensions- increased risk of hate crime, for instance. Neil Denton focuses on this.

Much more based on the recovery of the community- looking at return to normal.

<u>If your work is related to emergencies, how do you incorporate behavioural responses with respect to this?</u>

<u>If you think that you could make use of an improved model of behavioural response to fear, What would be the most useful tool you would use?</u>

Case studies.

[Int 2] is not interested in a singular building, more in a group of heterogeneous buildings and communities.

A case study provides a way of processing data, an example of what to look out for. It should provide relevant experience. There are very few case studies for city centre evacuations, they include: Manchester (IRA bomb), London (power outage), London (7/7), Birmingham (post 7/7, at night). It would be worth looking at Belfast's evacuation procedures, as they are probably well defined.

[City] currently has a zoned evacuation plan- not full city centre. It is out of date.

Which has been seen for the university bomb evacuation.

City centre evacuations- there is usually a transient population that aren't local inhabitants. If there is, say, a 500m cordon, that potentially encompasses a lot of people. This is difficult to model and cannot be fully tested. NOTWITHSTANDING exercise strong tower- 2015. A terrorist attack exercise was carried out in London zone 1.

"never seen a model that will model a city centre".

“current evidence based planning is ‘finger in the air’ other than being based on police tactics.

Assumption of emergency services is that people will starburst- to account for the worst case scenario of people being contaminated or witnesses etc.

B.1.3. Int 3

What is your current main section of research/industry?

My work is looking at psycho social aspects of the management of emergencies and disasters in particular looking at human behaviour in mass emergencies and how the way in which an incident is managed can affect public experiences and behaviour and how that can affect the outcomes from the incidents.

We do the research and then we look for opportunities to use our findings to inform policy- the best example is our work into psycho social aspects of mass decontamination, which has been used to inform some guidance documents in the US, and we’re also looking to incorporate it into the training that is run at the fire service college for mass decontamination instructors.

How does your work relate to crowd/pedestrian movement and emergency evacuations?

How do you incorporate Human behaviour into your work?

More specifically, pedestrian dynamics?

More specifically, pedestrian group dynamics?

I suppose the only time we would look at actual movement of a crowd would be during mass decontamination. So we’ve done some research tracking people through the decontamination process to see where bottlenecks are, to identify why there might be bottlenecks in the process in certain parts of the process, and also how communications strategies can affect the flow of people through the mass decontamination system.

Our work is very much focused on understanding the factors that affect human behaviours during disasters. The behaviours we’re interested in would be compliance with recommendations, public anxiety, and what affects people’s willingness to comply with recommendations. So how different communications strategies can affect human behaviours in terms of compliance and cooperation. And also to an extent how communications strategies can affect crowd dynamics- so how different communications strategies can affect the relationships between members of the crowd and how that can affect their overall behaviour. It is based on Social Identity Theory. I reviewed various other theories, including attachment theories, and in terms of understanding the way in which crowd dynamics can shift very suddenly, when it becomes apparent that there is an emergency, I think the social identity theory covers that in ways that other crowd theories don't.

It's that recognition that it's not just that behaviour continues to be normative during a disaster. It's an understanding that the norms of a crowd can shift very suddenly when it's apparent that there's a shared fate involved.

I expect you've seen John Drury's work around the London Bombings? It's the 2009 one, it's basically showing how, although people on the underground would be very separate, very individual. As soon as the bombs went off, there was that sort of shared fate which brought everybody together and created a sense of unity, which prompted helping behaviour and those kind of things. And I think where previous theories of crowd behaviours have focused on either public tendencies to panic and that sort of contagiousness of panic that spreads through a crowd, or where existing norms of behaviour are maintained, so sort of bonds between families and friends are maintained. The social identity approach shows that new norms are created amongst people who didn't know each other before, but you still get those kinds of bonds and that norm of helping behaviour that comes out.

<u>How do you define fear?</u>

<u>How do you incorporate fear in your work?</u>

<u>If your work is related to emergencies, how do you incorporate behavioural responses with respect to this?</u>

My work is focussed more on anxiety than fear, and I see the distinction being that anxiety is a response to an unknown threat, and fear is a response to a known, identifiable threat.

<u>If you think that you could make use of an improved model of behavioural response to fear, What would be the most useful tool you would use?</u>

Ideally- we regularly test communications interventions and those kind of things in experiments and exercises. If there was a way we could take the findings that we got from that, and understand how the addition of fear into that scenario would affect the outcome, without actually making people afraid. Because that's a major issue that we have, obviously we're not allowed, and we wouldn't want to cause people stress. But there's always that thought that we might get different results in a situation where people are under stress.

So we need more data, but there's an ethical issue in obtaining data, but without it we can't be sure that what we're saying is correct?

Essentially yeah. We have some degree of confidence- we compare more qualitative research from actual incidents with our experimental studies and exercises and they broadly come out showing the same thing that people will come out and behave in a normative way, that people will be willing to help each other, those kinds of things. In an ideal world, we would want to be able to test a communication intervention during a real incident, if that makes sense.

B.1.4. Int 4

What is your current main section of research/industry?

I'm a behavioural psychologist. And my work primarily focuses on behavioural responses in extreme events such as terrorism. With a particular focus on the way risk perception can be used, with behavioural dynamics.

How does your work relate to crowd/pedestrian movement and emergency evacuations?

I've been working on projects since 2009, on a variety of projects. Not specifically evacuation. The first example is looking at CBRN. There is a specific risk perception with those types of threats, because you have something that has a fairly low impact in terms of casualties, for instance a 'so called dirty bomb', that isn't particularly effective in terms of the number of people.... That came out of the expectation that the public would be particularly scared by that type of event.

A lot of the early projects we were looking at were like for instance mass decontamination- going to monitoring centres, looking at bio threats looking at vaccination behaviours. We also looked at asking people to stay

We've most recently been looking at the run-hide-tell policy, obviously involves crowds and crowd behaviours because we're talking about people being caught up in an incident in crowded spaces.

Depending on how you're defining evacuation behaviours, run-hide-tell is probably the most relevant.

How do you incorporate Human behaviour into your work?

More specifically, pedestrian dynamics?

More specifically, pedestrian group dynamics?

It's a difficult one to answer from the point of view as an academic, because I study human behaviour, so when you say how do you incorporate it into your work, this is my work.

This is going to take me into the realms of theory:

In terms of people movement, we haven't been modelling it in the projects I've been involved it. I know that there's a whole host of psychological theory that we draw in terms of understanding how people are likely to behave. If you look at the crowd psych literature there has been a change from the 'irrational' model to understanding that there are a number of factors. There are slightly competing theories or different approaches, some look at emerging norms, some look at the norms that people bring with them, certainly there's quite a bit of work on attachment, and the fact that people want to reunite with their loved ones, so therefore that theoretical work I think should inform any model of crowd behaviour in recognising movement won't be unidirectional-people won't just focus on saving themselves.

So depending on all sorts of factors around what is happening- if someone has a loved one-child-pet-stuff that they care about- if they think they have time, they will go back to a site that they should be evacuating. That kind of theoretical recognition of human behaviour, but I don't know how you would go about modelling that.

How do you define fear?

How do you incorporate fear in your work?

If your work is related to emergencies, how do you incorporate behavioural responses with respect to this?

When we think about fear, we tend to conceptualise in terms of risk perceptions. We think about fear in the extent to which it is proportional to the event that it is happening. Risk perception literature looks at the factors that influence the extent to which people will be scared, or over respond, or underrespond to any particular event. There are a number of factors that are pretty well established around:

Is it something that is natural or technological, or is it freely chosen, or unfamiliar or not? Is it a dread risk (so therefore something that has a low probability of happening but when it does happen a lot of people are affected like a terrorist attack vs ongoing deaths in hospital where the numbers may be large, but actually people aren't really scared...).

We think about fear in those terms and we also think about it in relation to how it translates into action or behaviour. If you want people to take protective health behaviours, then they need to be a bit scared. If someone doesn't think smoking is bad for them and they enjoy smoking then they'll carry on- they need to be a bit scared of the consequences. So if you look at it in the context of emergency planning, if you're thinking about how people behave, if you're thinking like preparing for floods, what you might be concerned about is lack of action. Particularly if it is an area where people have past experience of flooding, they think it will be ok, they have past experience, they're not scared, and you need them to evacuate and leave the area... your personnel are being put at risk having to rescue people because they refused to evacuate, in those circumstances, lack of fear can be problematic.

If you think about it in the context of terrorism, one of the goals of terrorism is to scare people. So in those contexts there is a lot of work around reassurance messaging.

And saying if somebody's fear of a plane exploding when they get on it stops them from getting on planes, that's going to have negative economic impacts on the plane industry. It also has negative health effects. Again there was research after 9/11, that suggests, looking at correlational data, that people were diverting to drive rather than get in planes, because people were scared of flying, and there was an increase in road traffic accident deaths. So actually, in avoiding something that was perceived as a risky circumstance they put themselves in more danger. So there's those trade-offs... Fear isn't necessarily always something that is negative, but in the context of emergency planning, we have to think of fear as something that needs to be managed.

<u>If you think that you could make use of an improved model of behavioural response to fear, What would be the most useful tool you would use?</u>

I think that the kind of silver bullet would be to be able to produce a model of behaviour that incorporates the multiplicities of behaviour, but is somehow based on sufficient past case studies to have some confidence that the proportions of each type of behaviour are reasonable. Obviously you'll never be able to totally predict. I guess models need to be fairly site specific...

[This model would be able to tell us] where people would naturally move, where log-jams would occur, so I guess that would be the most useful thing.

You would have to think about the type of incident you are modelling for, and any public space, you're going to have one set of behaviours which is probably going to be different to other areas (eg Grenfell towers). Depending on who your audience is, I would have thought with the public spaces work, if you're having problems with movement, it's probably going to be building design or it's because of a lack of situational awareness.

[It depends on how people use the space and who they are there with. It's going to be full of idiosyncracies and is probably going to be context specific]

B.1.5. Int 5

What is your current main section of research/industry?

Senior consultant, in pedestrian and evacuation dynamics, so I run models of pedestrian and evacuation movement. As an academic role I research with various international institutions in this same area, my specialism is hospital evacuation, I'm also moving a bit more into stadia. And I've just become an adjunct professor at the [University].

Industry is a full time job. Academia is my secondary thing, although we do do research projects as part of work at the moment. In terms of working time its 100% industry, in terms of real life time, 80/20.

How does your work relate to crowd/pedestrian movement and emergency evacuations?

How do you incorporate Human behaviour into your work?

More specifically, pedestrian dynamics?

More specifically, pedestrian group dynamics?

My work relates to crowd pedestrian movement and emergency evacuation in that that's all I do. I'm a specialist in these things. So I run micro-sim models of crowd movement, so pedestrian movement, evacuation movement. So we model using micro sim models (Legion, Pedestrian Dynamics, BuildingExodus, MassMotion, STEPS, Pathfinder and others). We use micro sim models that are agent based to do dynamic simulations of what people would do in emergencies and in normal conditions. By normal conditions that could be looking at retail analysis (so footfall in a shopping centre), it could be a looking at an operational problem (so queuing system in a hospital, eg a desk in a and e- which is a pedestrian movement thing but not necessarily a crowd) and then there's large crowd analysis you have stadia stuff is the thing I've been doing most. But you know people have modelled the Hajj, and massive crowds of people moving around. So that's basically how it's used in work.

But also we do static calculations, so hand calculations, excel calculations of flow rates. So you might have the Green Guide, which is the stadia authority on safety- you can do a set of calculations for flow rates through a system. So just by hand you can work out the capacity of a stadium.

Will you use dynamic analysis as standard for every job?

No. So you've got micro and macro. Macro will be static- it's just like this is a large systematic review of a system. And it looks at it as not individual people but you just look at flow rates through a system. Then you've got microscopic which is every individual person. Then you've got mesoscopic, which is in between the two. You can say we don't care about the agent based level at this high level or at the start or the middle of the day because nothing significant is happening, but once you get to this level of congestion then we go into agent based, or the other way around. So there's no standard, and the thing about working in this industry is there's no real standard approach- it's not very regulated it means you approach each problem as new.

In some of these interviews we've talked about designing buildings- if there's no standard approach, is there any standard regulation?

There's Approved Document B, which is under a lot of fire at the moment, and in terms of the calculations in Approved Document B, you need to have this exit width per this number of people, and that's based on a 2.5 minute evacuation time. It's based on a 2.5 min evacuation time whether you're a stadium with 90k people or you're a house of 3 people. That is the time. So that's like a 100 year old piece of data, from 1911, so potentially not useful to apply to all situations. So you have to build a building to standards. But when looking at engineering a building for fire safety you can meet them with prescriptive standards (like this is the width of the doors, this is the exit route, this is level of protection) or you can run some analysis like we do to find out whether or not there is comparable level of fire safety.

How does human behaviour get incorporated, aside from just flow density characteristic?

All models, from a hand calculation to an agent based models have some assumption behaviour in it. So when you do a hand calculation of the evacuation of a theatre, you have to make an assumption of the proportion people going through each exit. Therefore you're making an assumption of the behaviour of the people in there. So whether or not you make that evenly spread or, you can do things like familiarity by exit availability in static calcs. You'd be like 'this exit isn't available because no-one knows it's there'. So you can replicate some levels of behaviour in terms of knowledge, communication, and sharing of information, and again exit use and familiarity of buildings. And again, even if you do a static calc, you're able to put in your flow rate.so what you could say is 'I think people are going to move in groups so there'd be group behaviour, or I think people are going to move independently because they're all commuters, and you could replicate that in your flow. So it's not that there's an absence of behaviour, it's that you only have a certain number of things you can control but you can still change them in order to, in some way, replicate human behaviour. Which is clearly not as intricate as putting in individual behaviours in agent based models.

<u>How intricate does that get? What is the highest level of replication?</u>

Different models work differently. The way in which you're able to put in characteristics is different across all models, because it depends on their underlying functionality. So you have models such as BuildingExodus or pedestrian dynamics, where you can set up conditions and so you can make it quite complicated. So what you can is you have agents, you give them itineraries, to replicate the behaviour.... So you can make people delay, you can make them go to certain areas. You can allocate those statistically... and that's quite powerful because you can replicate all sorts of different behaviours with it. You don't have models that have a default behavioural system that predicts what people will do. You as the user have [to decide] what proportion of things happen. So in that way they're quite powerful but the level to which you can put stuff in, it's like the flow, there's models on all of the scales. So you've got the flow thing where you can reduce your flow or increase your flow, saying I've got groups or not groups. Then right the way up to the agent based model that actually has groups in it. You have a model in all of these varieties. So some of them you might be able to replicate group behaviour by sending people to the same spot, but that's not actually specifically representing group behaviour.

<u>How do you define fear?</u>

<u>How do you incorporate fear in your work?</u>

<u>If your work is related to emergencies, how do you incorporate behavioural responses with respect to this?</u>

<u>If you think that you could make use of an improved model of behavioural response to fear, What would be the most useful tool you would use?</u>

All of these depend on the model. The one I have the most experience with in terms of development is BuildingExodus. There's a whole thesis on emotional modelling by Claudia Hollmann, and that's available from the Greenwich Website. What Claudia did was model itineraries, she modelled all the things people do before they get on a train. It wasn't fear necessarily, but it was all the needs that you have in that moment, and then you prioritise them. So it's like looking at a utility problem... the agents could then decide and drop tasks... so it's about moving around tasks on an itinerary, which presumably then could be used in terms of modelling fear, if we know if there's any behavioural theory of what people do, because if you're scared and you were originally buying a bottle of water, you'd stop buying a bottle of water, you'd stop engaging in the task you're doing, and engage in a new task that could be protecting yourself, protecting your belongings, getting out of there, fighting back, whatever. So in terms of modelling functionality, you could model fear based on those priorities, but I don't think there's any theory to my knowledge, yet, until you develop it, that says 'these are the things that people engage in. You've got things like Emergent Norm Theory, and Protective Action Decision Model, those you could put in, but you'd have to make big assumptions about what people would do to protect their property and everything else... and then put them in as some kind of priority and then in the model, when something happens... you have list of all these things you want to do, and then you choose one of them. So it has the ability to do that already. Exodus does in terms of dynamic itineraries, whether it does in the commercial version is a different story, it is in the research version. Pedestrian Dynamics certainly does. Legion, maybe. It might be worth looking at all the different versions. I think they're all working on being that open.

So there is 'here's the fear model, click on the button 'fear'', it's an open itinerary based, task based that you can use to input different stuff.

The theory. We need the theory. You're going to have to implement it in a model for proof on concept, but as soon as you have the theory, all the models would follow suit. Or at least, if the theory is there, I'll be able to go into my own models and use the theory. What we need is an idea of the distribution of behaviours: all the things you can expect people to do and the likelihood that they'll do it. Badly. If we were modelling a particular situation, can you characterise fear, like 'how bad is the fear', and how likely are you to do various things based upon that. Then we'll be able to incorporate that into models. But I guess it's hard to validate.

We're trying to get CCTV footage of individual incidents. We'd need a huge amount of data to calibrate the model, and that would only work for that specific scenario. Yeah because it depends so much on the position you're in, whether you're near the fire, you can't aggregate everyone's experience. We have quite a lot of data on that, and even then the theory is very sparse.

B.1.6. Int 6

What is your current main section of research/industry?

How does your work relate to crowd/pedestrian movement and emergency evacuations?

My job title is reader in social psychology. Behaviour in emergencies and disasters is one of my research areas and so is behaviour in pedestrian flow. I specialise in crowd behaviour.

How do you incorporate Human behaviour into your work?

More specifically, pedestrian dynamics?

More specifically, pedestrian group dynamics?

A key concept for understanding collective behaviour generally as well as in emergencies and disasters is the concept of identity. There's widespread agreement among social scientists is the concept of a norm. that behaviour is structured by norms that belong to groups. As psychologists we say why do people adopt one norm rather than another, a norm being a social rule and a definition of appropriate conduct. Because people have lots of rules because society gives us lots of rules.... Social identities have given norms ,different rules associated with them. People act in terms of their identities, identities include definitions of how to behave with other group members and with people outside [the group]....

That would tie in with your work on self categorisation theory?

That is the basis of many of the predictions I have made in the studies I do on collective behaviour generally and in emergencies and disasters.

Why is SCT better than ENT or SA theory, specifically in disasters?

ENT has value, firstly it goes beyond some of the earlier theory in that it says behaviour in emergencies and disasters is socially structured and social norms are ways to conceptualise that structured behaviour. That behaviour has upper limits if you like, that people don't just do anything. I think that part of the problem, of the critique with ENT started from other norm theories, before you got as far as SCT. Norris Johnson was another sociologist. His model is based on social norms. He points out that there isn't this requirement to reinvent norms in an emergency. That norms and rules that people use to structure their behaviour are sometimes the same as people use their day to day life anyway. As in, we help the more vulnerable, gender roles, social roles and so on. That was part of the critique, but that was within the norm approach.

For me, we have so many different norms, we need theory to explain why one norm becomes important to people rather than another. And that is not provided by those sociological theories. Unless you say, as ENT does, that we simply invent norms from a situation of normlessness. The other thing to say about these norm theories, not specifically ENT, but that family of sociological theories, is that they seem less good, at least the modern versions, at explaining collective co-ordinated behaviour amongst strangers. And this goes for Mawson's affiliation approach as well, which is based on the idea that we don't simply seek to get away from the emergency, we seek familiar others and this reassures us etc. and when those familiar other aren't there, this panics us, he says. There are studies showing, examples showing, that even when there aren't people that we know, people still behave in this socially structured way. So therefore we do need a concept of norm, it is true that our existing bonds with people are really important in telling us how to behave and our priorities. But more is required to explain those situations where there might be multiple norms as the first point, and we don't know people, which is the second point, there you still get structured behaviour.

<u>The only other model I've looked at is the PADM.</u>

Well one thing to note about computer simulations is they often try to include a psych element, but I suppose it's a consequence of the nature of academic careers and so on, but it's quite difficult to straddle more than one discipline and become experts in more than one discipline, so people become experts in computer simulation and then they approach the psychological literature, with less knowledge. So therefore there are a whole series of computer simulations that refer to LeBon or they refer to social identity, or they refer to other theories. And try to acknowledge or instantiate those theories in their models, often in very limited degrees, or in a kind of uncritical way, and that's been the problem up until now, that psychology is required, but it's not always included in the most well informed way.

I'm not familiar with the model you're talking about, but usually those kind of models that include some psychology are not based on the up to date research literature- what people actually do during emergencies.

What seems to be coming through is that to have a predictive model, you need some psychological expertise and behaviour considerations in the model.

I'd say so, yeah, you need some psychology. We did some work with some computer modellers in Munich, and they set aside some time to become familiar with the psychological literature. Because it is an effort and a time consuming effort to properly be interdisciplinary. So it wasn't just an add on, it was taking seriously the different theories and trying to get some expertise in them, so that they could properly represent what we know from the social science literature in the simulations they were developing

How do you define fear?

How do you incorporate fear in your work?

If your work is related to emergencies, how do you incorporate behavioural responses with respect to this?

When social scientists first started running experiments on escape behaviour, there was a debate around the concept of panic. One of the famous example is by Mintz, he had an analogue escape paradigm in which people dragged corks out of bottes. He pointed out that you could get blocked doors as when people tried to exit, and his study showed this in analogue form. People trying to exit in an uncoordinated way, without strong emotions.

You can turn that round, you can imagine an emotional crowd that nevertheless evacuates in a coordinated way. What he was trying to do was separate the analysis of escape behaviour in emergencies from the question of emotions. Then other sociologists like Kelley, criticised what Mintz was doing, because they said 'well his paradigm hasn't got threat in it'. Threat is a huge variable, people think they might do, be fatally injured, a paradigm that doesn't include that in a study on emergency egress has got a key feature missing. And that is absolutely true.

When you say fear, there is a debate, probably a helpful one, about to what extent that fear is reasonable or irrational. Usually people ask about panic, and then we say, what do you mean by panic, and it ends up that very often people mean a behaviour that is unreasonable or irrational, but then the problem is judging what is unreasonable or irrational. So fear is a better way of talking, I agree with you fear is important, that it is a better way of talking than panic, that doesn't include a judgement that is hard to measure on the extent to which that fear is reasonable or not, it is certainly its part of the experience for people in emergencies and disasters.

The first experiment we ran, ours was an analogue study as well, it was a role playing study using VR. We tried to get people to imagine fear, to imagine a threat that would get them fearful. Because that was a roleplay study, like Mintz, then that would only have a limited impact on people's behaviour. Soo it was then necessary to speak to people who had been in emergencies in an open ended way, hear about their experiences. Of course, on the one hand interviews are post hoc, you get post hoc reconstructions, you're going to get memory issues, you only speak to the people that come forward, all these things are true. On the other hand, if you do interviews with survivors, you get in an open ended way the issues that they think are important, that you as a researcher might not have prioritised. You start off with open ended questions, and you talk about emotions like fear.

When we did this with the London bombings of 2005, we interviewed survivors from there, and collected secondary data, we looked at the fear of death and even though the bombs had gone off by the time people were trying to escape, this wasn't an egress from an impending threat. Objectively it was a situation where the threat was over, but subjectively, as you can imagine, a bombing situation, after the bombing, the level of fear is still going to be very high. Because people don't know that they're objectively safe, that there isn't a secondary device, that the tunnel might not collapse, that they might not get electrocuted on the train lines, or another train might hit them. There are all sorts of reasons why people felt afraid, and in fact they felt in fear of their lives.

There was only one person I remember in our interviews who said 'I thought I might die, but I'm not afraid.' so you get sometimes people who are more serene about the threat of death. But most people talked about fear, most people thought they would die, and most people found this a frightening experience. So fear was a massive feature of the phenomenology if you like, of that emergency. Nevertheless, those same people who were in fear of their lives also talked about how much help there was, help they observed from others, help they gave to others. From an individualistic point of view the evacuation was not optimal, because people didn't act as rational maximisers, they behaved in a cooperative way, people would allow others to go first, they stayed behind and slow their own exit for others, and all these kind of things that problematized the usual starburst metaphor that is sometimes used to describe these things.

In answer to your question fear is important and when we research these kinds of event, at least when we do interviews, we try to be able to show the extent and importance of these emotions in people's experiences. And we try to show the extent to which they impact on behaviours. But in this case, and in others, the fear was very strong, we set aside questions of whether that fear is reasonable or unreasonable, you might say it is completely understandable when the bombings have already happened and finally it didn't prevent them it would seem, from acting in a cooperative, coordinated way with other survivors.

<u>I had an interview with [Int 3], and they make the distinction between fear and anxiety. Does your work specifically look at fear?</u>

Well it depends on the type of emergency, the reason that anxiety was important in [Int 3's] work, was that they looking at more the response to the emergency than the emergency itself. What she's measuring people's responses to is the decontamination process- if you can imagine the life history of an emergency like a CBRN incident, then to the extent that people realise they're in danger, you're going to get fear, then when the fire service turn up with their tents and their PPE, then you're going to get some anxiety. Because as well as threat from the emergency, you have social anxiety about the fact that people have to strip off in this rather unfamiliar situation. Part of the reason why [Int 3]'s work has that focus is because of the nature of the event that they are studying.

<u>Do you work with CBRN, or MTFA?</u>

I was [Int 3]'s supervisor, so I do that work, yeah.

The other situation you get anxiety is when people want information. Say there's an incident and it's not clear yet that it's a bombing or what kind of threat it is. Like something on the underground, when the announcement is not very informative, then that's when you get anxiety, not fear, because people don't know what it is, but you get anxiety. And that anxiety increases with the less information people get. Some people talk about good anxiety, bad anxiety, because you want a certain level of emotional response to act. If people are completely complacent, then that might be a reason that people don't move quickly enough. This is the case with fires, where people don't think they're in danger, don't feel any sense of threat, and therefore they're too slow, and that is one of the reasons that people are caught in these things. That might be a case of good anxiety. Bad anxiety is where people are waiting for information and not getting information, that can turn into a stress reaction, where it's not going to be helpful, because what is the action should they be taking anyway? So where there is a lack of information that anxiety is not good anxiety.

It seems to be coming up that fear/anxiety needs to be managed.

We're talking about different things here, because in those events where you've got the possibility where those responding or managing the situation can communicate, then that's true, but in some emergencies, you've got nobody there, you've got no responders , no authorities, in those situations, you haven't got that option.

If you think that you could make use of an improved model of behavioural response to fear, What would be the most useful tool you would use?

Linked to what you're working on, if we're going to do simulation type studies, something that makes the whole experience more realistic. For real world studies, field work, the problem is always access, people not wanting to talk, gatekeepers. If it's simulations, then something which is engaging, which ethics panels are not going to object to, which would allow the experimental study of these kind of things would be good.

B.1.7. Int 7

What is your current main section of research/industry?

It is crowd modelling and consultancy on industrial projects. We are also involved in research related to crowd modelling.

How does your work relate to crowd/pedestrian movement and emergency evacuations?

We're either involved in helping to progress a design for a new venue or event, looking at crowd movement in normal and emergency situations or we're involved with existing venues and events, helping to optimise their management procedures, to assess whether the event is safe and try to improve it.

Are there any particular types of software/tools that you use?

We use commercially available tools, such as Legion and MassMotion. We choose which one we use based on how suitable they are for the project. MassMotion for example can import certain file formats and make the whole process a lot easier. It also has a lot of features that can make certain processes easier. For example, if people are in a queue, and then they go to do something else. MasMotion models those logical processes quite well. Legion is required to be used right now by Network Rail, Transport for London and some others who request Legion specifically because they have a copy. But we also use Legion in certain situations where we feel it's easier to apply. We also have an in-house software, for a wide range of different types of modelling, that we'll often try and use, but they're mostly specific to a particular situation. So we'll decide to use them if they apply to that situation. One of those is an evacuation model that we created through the 'eVACUATE' research project. It's a real time modelling tool. The idea is that it gathers data in real time, then processes it much faster than real time to provide a prediction of crowd movement during evacuation in a matter of seconds. But we have also been using it for different projects where there are very large crowds. It runs really quickly, so we've used it with one million people moving in the model at one time.

What are the data inputs?

As part of the evacuate project, it took inputs from one source which was a data fusion engine. But that engine took in any form of data on the number of people. Either the number of people crossing a line in a flow, or the density of a particular area. It could also process Behavioural data, but in that model we weren't actually using behavioural data, and most of the sensors weren't at the stage where they could gather data that we could validate that is actually usable in any form. So in the end it was just densities and flows.

How do you incorporate Human behaviour into your work?

More specifically, pedestrian dynamics?

More specifically, pedestrian group dynamics?

At the moment, all the models that are commercially available, and the evacuate model I just described, is all focused around how crowds move, rather than how they behave. So any behavioural element that is put in is done by the modeller, who decides that this particular behaviour, based on experience, is more or less likely to occur. And quite often we then test different behaviours to understand what may happen in a range of scenarios. So in the example of an evacuation, we might say that 'there's really good of signs around the place, actually the behaviour is going to be that they follow the signs properly and hence go to the nearest exit. Or another behaviour that is very often witnessed is that people want to go out of the entrance that they came in. In which case we'll probably run a scenario in the model that tests a percentage of people wanting to head back to the entrance that they came in, and then see how that affects the evacuation. But none of the models truly incorporate behaviour i.e. taking a behaviour and forecasting how that behaviour may propagate through the crowd, for example with passing of information from one person to the other, or how one person's movement (or a groups movement), and their behaviour, would then influence another's behaviour.

So all of them are based on the model of judging what behaviours are likely to occur, and then running different scenarios to try and cover the things that will actually occur.

<u>How do you incorporate fear in your work?</u> . <u>If your work is related to emergencies, how do you incorporate behavioural responses with respect to this?</u>

It's really difficult. Usually, in terms of the evacuation, we don't know to what level fear plays a part in people's decision making. Fear is also a relatively individual trait, which we all feel in different ways and react to in different ways. So even if you could measure an 'average' level of fear across an entire crowd, it would then be very difficult to extract the properties of what the behavioural response for the crowd would be, and it is impossible to do for an individual. In practice, and if you look at video footage from different events, you see a lot of different things, either people seemingly moving very calmly, almost unaware of a threat, to videos of Manchester arena, of people climbing over barriers, and dropping down to try to speed up their exit, seemingly selfishly, but then you do see people helping them down. In terms of how we incorporate fear, really, we're looking at how the crowds would then move in an evacuation. There are some new theoretical models that try and incorporate panic or fear or other traits by having a parameter such that people follow certain people who seem to know where to go - I.e. the overall effect of people being scared is lower. Or the parameter makes people not follow anyone else trying mimic the perception of fear and panic whereby people run around a space not knowing what to do. However, these are yet to be sufficiently validated as to how people actually behave, meaning we can't yet incorporate these into our work.

In terms of our work, we're always trying to put management procedures in place, and strategies to provide pertinent information to the crowd to reduce any of that behaviour. It's all about reducing the risk that people are so scared that they don't know how to react.

If there were some research that showed for particular situations, the level of fear increased, and that had a particular effect on people's behaviour. For example, a particular speed distribution applied to the behaviour - i.e. people move slower or people start running; or people choose different routes when they are more scared, or less scared. As far as I'm aware there's very little research, if any. It certainly doesn't quantify it in any way, and even just subjective data doesn't really indicate there's a particular behavioural response, nor has anyone really been able to measure fear on a large scale.

I don't see any quantitative research out there.

If it was there, it could be incorporated into the model so long as it translates into some sort of movement of the crowds. If fear spreads through the crowd (person to person), that could also be modelled but would need to be validated in some way.

The big question is whether that is useful within the modelling. It would be useful in terms of researching which situations this happens in and how to limit it.

I would imagine in terms of the venue or event geometry, as we would be designing it, it may not have a great impact, it would more be to say 'ok this is the sort of information you need to provide the crowd, this is how you need to guide them, this is the management you need in the venue, these are the signs etc etc'.

To me it's something that is really worth looking into, but I would imagine once you've got to the stage of 'this is the reaction that occurs', you'd be looking at ways to mitigate that. And in which case it would be those things that would be implemented on industrial projects, rather than being implemented in the crowd modelling. We would just be using all the techniques to mitigate that within the building. It could be implemented for training purposes - I.e. the management strategy put in place has failed, and the models would test the different reactions to train staff in how to deal with the situation.

If you think that you could make use of an improved model of behavioural response to fear, What would be the most useful tool you would use?

I think one of the first things would be is: what effect does fear actually have? I presume there's a multitude of possibilities as to what individual people will do, but the real question is, when you've got a group of people, how do they react as a whole? It seems to me very likely, that if one person is incredibly scared and doesn't know what to do, the chances are that there would be another person there who has reacted in a better way, and is scared to a lesser extent, to the point where they then help the other person.

The next question would be - Is it different for different sizes of crowd? If you've got a huge concert and you've got a certain percentage of people who don't appear to be scared, and lots of people that do, does that fear spread through the crowd and how do people behave? If there is a smaller group, what percentages of fear or no fear would lead to particular behaviours.

What are the different levels that lead to people doing a particular behaviour? Which I'm assuming is speeding up, slowing down, not moving, or choosing different routes. But it'd be really interesting to know, if we have a few people who appear really scared, is that then automatically mitigated by other people around them who are less scared and help them out.

If you can actually measure the level of fear, somehow, then you'd have to attribute that to a description of behavioural response – for evacuation the important one being how people move. I presume that it would depend on the particular distribution of fear amongst the crowd, that would lead to different behavioural responses.

What would be really interesting would be to know what type of fear distribution throughout the crowd would skew the situation to being much more chaotic than it would be otherwise.

B.1.8. Int 8

What is your current main section of research/industry?

My field of research is fire safety engineering and I work as associate senior lecturer at [university], in which I teach the use of evacuation simulation software and human behaviour in fire. I work on both fire safety from the perspective of the building industry, so when it comes to building and transportation systems, as well as large scale fire disasters such as wildfire scenarios.

You also work in validation as well?

Yes, I work as team lead for a task group working on a document on the verification and validation of building fire evacuation models. It builds on the work I did with the National Institute of Standards and Technology (NIST) in the US, in which I was documenting the process of verification and validation for this type of evacuation simulation, this is also for fire safety engineering.

How does your work relate to crowd/pedestrian movement and emergency evacuations?

I've been working in and researching the development, use and validation for some models of crowd modelling for evacuation applications and then I teach this as well. So I teach both Swedish students in a Swedish program as well as international students, we have an international Master's program in the university, which includes the subject of evacuation simulations, and so pretty much what we teach is my main area of expertise, so extremely relevant to crowd dynamics and evacuation modelling. My title is lecturer in evacuation modelling, so extremely specific.

How do you incorporate Human behaviour into your work?

More specifically, pedestrian dynamics?

More specifically, pedestrian group dynamics?

Here in [University], we've been conducting several types of data collection efforts concerning different aspects of human behaviour during emergencies. Our main focus as I mentioned before, is fire safety engineering but we also looked at more basic, fundamental research let's say research concerning person dynamics. This all translates into modern development, so some of the most known commercial software for evacuation software make use of our datasets. In particular concern is people movement in smoke, because that's one the main areas of research that we have here. So we have collected some data, and some of those are actually implemented today in research and commercial software. When it comes to pedestrian group dynamics we have done different types of experiments to investigate different emergency conditions, different types of scenarios, which involve group of people, so not only individual experiments. This may range from a variety of problems, one example I can give you is we had a project on fatigue and how group dynamics may interact with the people behaviour when it comes to the motivation levels.

So when you get tired and when you are in a group, how you slow down your speed, what do you do, how close is the person overtaking people, what are your bonds with people that you know and you don't know.

This all takes place in an experimental setting. I don't want to call it an artificial setting

Sometimes what we do is research connected to drills which might be announced or unannounced, the key is of course the degree of ecological validity of the data depends on what people know about the drill, if they are aware of it being not real, or if they not aware of it. This is another stream of research that we do up here in [University] and is used to build up different type of models. My own research. . .

Everything has to do with understanding and modelling the steps that lead to influencing other people in the position to evacuate. So for instance how the first people decide to start moving, and looking at what the other people are doing in relation to what they see, and what other sensory variables are affecting them.

This is a bit of a summary of the type of research we are doing. So we are mostly a research and education institution, focusing on development, use and validation of the models, and teaching these type of models.

So data gathering is important- do any computer models specifically detail which behaviours they can model, are there better models of doing this or worse models?

Some of them they can actually be somehow modified to implement quite complex behaviours. Others are, let's say, less flexible. The most advanced ones are based on agent based modelling techniques, or even different techniques, as long as they implement flexibility, they can be used for implementing also quite complex scenarios, which include group dynamics. The problem is not to put this in the model, but to set an input that is matching what is matching the actual behaviour. The validation of the model itself is the tricky part. With a computer you can pretty much whatever you want nowadays with these models.

How do you define fear?

How do you incorporate fear in your work?

If your work is related to emergencies, how do you incorporate behavioural responses with respect to this?

These are quite tricky questions I can tell you, because nowadays many models do not incorporate explicitly fear. I know that there have been quite some discussions about this. I know from the psychology field, there has been some research into how fear can affect different aspects of pedestrian dynamics, in a workshop we did in [University] on evacuation modelling in

June within the Symposium of the International Association of Fire Safety Science, a researcher from the psychology field has been talking exactly about this. What came out of this we could potentially do things like implementing the consequences of fear in pedestrian dynamics models. For instance how fear can affect spatial navigation, for example. We know that one reaction to fear is avoidance of the area or whatever caused this stimuli, so spatial navigation could be potentially modified in order to account for this. Many other things could be implemented like for instance the way group dynamics change in a model in a fearful scenario.

So again looking and considering a definition that refers to the avoidance concept, something that you want to avoid, we can go and think about the things that can be implemented in behavioural models and models for pedestrian dynamics

I can tell you these things aren't so well developed, not because it can't be done, but because there is a limited understanding of the psychological basis of the consequences of fear in the context of pedestrian dynamics. The research that has been conducted in this area, it somehow needs to be more and more integrated in the engineering applications, because the engineering models they tend to be simplified, because of course they need to be simple in order to be used, but on the other hand they need somehow to have complex elements. You have models that attempt to include many complex behavioural variables.

Going down the line the more complex the model becomes, the more it looks like a research piece than an engineering applicable tool. So the trade-off between complexity of represented behaviour and the complexity of use has to be taken into account. You can have the most complex and advanced tool, but if people don't know how to calibrate it, it becomes quite useless. So you have to have a balance between these two things.

Some of the efforts of the researchers like me are in the middle, because I have an engineering background, but I work with psychologists and social scientists, trying to hear what they say and trying to put this into tools that are usable for engineers. Because from the point of view of fire safety engineering, the people that then apply these models are in most cases engineers, so these are the people that have to be able to understand and use these models.

I know some people have attempted to model some very complicated stuff with these evacuation models, but more than giving one model, I can tell you that some models allow enough complexity to try implement this, but it's always channelling the end use. So a good user, that knows what he wants to get out of these models, he can actually implement quite advanced behavioural scenarios, including fear. The problem is then that nowadays, there is the discussion of how much responsibility you should leave to the user, and how much this model should include complexity. Because if you put too much complexity, and you have users that don't understand it and just use defaults, then you completely miss the point. Again the trade-off between these two things is quite important.

From my perspective there are models that can implement complex behaviours, including reactions to fear. As I say, the way fear affects spatial navigation, that's something you could implement in today's models, or group dynamics. The problem is we first need a deep understanding of what actually happens, before we go and implement it in the model.

<u>If you think that you could make use of an improved model of behavioural response to fear, what would be the most useful tool you would use?</u>

The answer is very simple: data. We need basic research. We engineers, and also as researchers, we struggle to fund basic research. We have a much easier time to fund and get an understanding of engineering applications, and things that are related to real life applied problems, but we have a much harder time to get an understanding the fundamental basic principles related to human behaviour in fire. This is what we need. We do not need a model, we need data to calibrate a model. The type of data we need is fundamental basic research for the case of fear but I can make many other examples in similar fields, or behavioural responses in the same field. So we need to dig more into the understanding of the fundamental.

My answer is that what we really need to create this type of improved model is the basic fundamental research which aids us understanding what to put into this model, rather than having a computer simulation. Because this computer simulation itself is quite easier to set up once you know what to put in there. And the use of this well, when we come to the use of such a tool, I come from the fire safety engineering field, so it's bit of a different type of application. If we look at terrorism, and these kind of extreme happen more often, the use of evacuation models may be more complicated.

In cases in which we might reach an extreme scenario, but those are very rare cases, in the majority of those cases it's hard to have people understanding what is going on, understanding there is a threat, understanding there is a problem, rather than calming down. In the media when we always see words like 'panic' and so on. That's the one false myth that we try, as researchers to dismantle. The problem is to get people out of the building and understand that something is happening, and that they have to leave, rather than instead calming them down, and not having competitive behaviours, that's only in very extreme cases (in the fire safety field), in the majority of cases we have the opposite problem.

B.1.9. Int 9

NB: this is not a transcript, this is a summary of notes taken down during interview. Any quotations are identified in quote marks.

What is your current main section of research/industry?

Counter-terrorism, including "how to mitigate the effects of terrorism".

How does your work relate to crowd/pedestrian movement and emergency evacuations?

Making sure we understand how people, in order to reduce casualties. Also we need to ensure that any mitigation measures deployed do not introduce hazards during day-to-day life or during the evacuations themselves. An example of this is the use of bollards in London. Crowd movement analysis shows that they don't affect movement in day-to-day life, but this analysis still needs validation for crush injuries in serious emergencies,

This work involves mapping out scenarios in design stage, and considering 'have we created pinch points or hazards?'

How do you incorporate Human behaviour into your work?

More specifically, pedestrian dynamics?

More specifically, pedestrian group dynamics?

Human behaviour is a fundamental of the business. We need to consider it when we ask how to influence members of the public into buying into security considerations. Another factor is asking how can we influence potential terrorists and their behaviours, to potentially deter or mitigate an attack.

How do you define fear?

How do you incorporate fear in your work?

If your work is related to emergencies, how do you incorporate behavioural responses with respect to this?

We include the concept of fear in our work in several ways. We include it by sowing doubt into our adversaries' minds- a fear of detection, fear of disruption and fear of being caught and sentencing, and the social stigma that accompanies it.

We also use fear for in our work considering members of the public, but not in a scientific way. One major thing we consider is how do you design a building that you can use intuitively when you go about your business, but then also use intuitively when there is an emergency happening. We can alter the design process by talking to architects, trying to introduce subliminal influences to help movement.

<u>If you think that you could make use of an improved model of behavioural response to fear, What would be the most useful tool you would use?</u>

"A better understanding of fear". Not a tool necessarily, but having something that can test threat scenarios, and assess different designs. This might be a computer program, or a physical test that can accurately reproduce what occurs during these emergencies (in a controlled manner). This will help us improve in multiple areas, not just the design of buildings but also, how building runs and how the security of this building runs. An example is the design of the lobby area, where you need to consider the location of entrances, fire exits, and security stations. You also need to consider the training of security guards and receptionists.

We recently wrote a guidance document about the best way to consider approaching the problem of 'vehicle as a weapon' attacks. In it we made the point that we may be in a world where we can't protect everything at all times, but instead we can look at the deterrent value of various items, like trees, bollards, work signs, etc. However, this is not tested or validated, and therefore relies on experience.

B.2. Pilot study

B.2.1. Interview 1

AS: I'm recording, and can everyone please introduce themselves for the recording.

K: I'm Kirsty.

M: I'm Mina.

D: My name's Dan.

A: And my name's Alan.

AS: Brilliant, thank you. Ok, now what we're going to be doing today: I'm going to be talking through an experimental protocol, and what I'm going to do is I'm going to check in with you at various times during that protocol about what you think is happening, what you would imagine doing next, stuff like that. So this is going to start off with basically me asking the first question. What does everyone currently think that this experiment is about? Now, I've given you very few details, but I've included the participant recruitment pack. So can we start off, I think probably the easiest way to do this would be alphabetically, so can we start off with Alan. What do you think is currently that the experiment is about?

A: Right. I think the experiment is about engaging interpersonal reactions in relation to cognitive tasks.

AS: Ok, brilliant, thank you very much.

A: And the reason I say that is with reference to the flyer that you attached initially.

AS: Excellent, excellent. Thank you very much. Next would be Dan.

D: Ok. My view is that I've noted its PHE involvement, it's a human behaviour experiment scenario. Other than that I have no preconceived ideas, apart from the picture that I see which isn't on a file on my screen, it's actually a picture, it looks like a basketball court or a five-a-side court – it's quite a small one, I haven't tried to open it or anything like that – with people standing in it. So it's clearly some kind of sporting venue. Other than that, no preconceived ideas, as I said.

AS: Brilliant, thank you. And Kirsty?

K: Yeah, I didn't really have a clue at all, but again I've seen the picture and read something about you needed to be fit, you needed to be able to run 100 metres, so I'm assuming something sporty, and something about human behaviour, but I've not really got a clue.

AS: Ok, brilliant. Mina? Last one. But not least, of course.

M: Hi, I'm sort of similar in that I've not really got much of a clue, but looking at the recruitment page leaflet that you sent I wonder if it's maybe to do with kind of barriers to people, you know, joining experimental trials or kind of questions that people might raise, the worries they might have, but I don't really know.

3.10 AS: OK. Brilliant, brilliant, very good. So now what we'll do is, I'll start walking through the experimental procedure and just to set the scene, if you could open the picture file which, as we've said, is of a sports gym, and that's where the setting of this experiment will be. So people will turn up, it will be in late December, approximately 4 pm and there will be about ten people around you. And you will see in this sports hall, there will be some boards in the middle with some pieces of paper on them that you can't read just yet, and there will be a place where you can put your bags, there will be some seats, again against the wall. So I'm just setting the scene here, and if everyone has seen the picture then we can get started.

[General murmur of agreement]

4.05 AS: OK. Now, what I'm going to do is, I'm going to read out what, exactly what I'm going to say to the participants, so it might come across a bit stilted, but please just give me some rope on that one. Good morning everyone. Please feel free to put your bags and coats there; we'll get started when everyone has arrived. Thanks. Now that everyone has arrived, we can get started. We're here today to investigate the movement of people in groups, and specifically in relation to cognitive tasks. An example application of this might be the design of open-plan offices. Before we get started, we'll need to spend a few minutes filling in forms. Please take your time to read these forms: they say that we will be taking personal data and using it in the research, but that all this research will be anonymised and stored securely, as well as never being used for anything except this research. Ok, break in character here: could everyone please open document 1 and document 2. Has everyone done that?

[General assent]

5.17 AS: So, yep, fantastic. So, in document 1, you can see an emotional questionnaire where, I'm not suggesting that we do it now, but take ten seconds to look at what it's asking and imagine how you would go through it, it should take about 3 or 4 minutes to fill in the whole thing, and it's just asking for a number of what emotion you're feeling now on a scale of 1 to 7, 1 being 'not', 7 being 'an extreme amount'. This is for use of a baseline emotional level. And then in document 2 there will be some questions regarding how you know each other: how do you feel about the group, do you know any other members of the groups? Ok. So, back into character now. Ok, once that's done, can you please fill out these forms, using your ID numbers – I would hand out some

ID numbers. These forms tell us your name and several important pieces of information, including your bank details, so that we can pay you. In addition to the emotional questionnaire form, we will also need to take saliva samples. At this point, I would hand out essentially test tubes that you can provide a saliva sample within. Ok. Now we can start on the experiment. As I said, we'll be looking at how people move while they do cognitive tasks. So we need to track your movements and check saliva samples. During this cognitive task, there will be financial incentives to do well. Please take these stickers – I have each of you individually coloured stickers – as they will help you in the task. You can earn up to £40 in this task, which we expect to take around 25 minutes. But there is no time limit, so don't worry about taking longer if you need. Please keep in mind that if you are taken out of the game, or you fail the task, you will lose all of the earnings you have built up, and be paid a minimum of half an hour's minimum wage. You must follow any instructions that you received during the experiment, and you will be able to read these instructions once you start the task. But the stickers will be used to keep track of how much you earn, so keep them close. Ok. So now we walk into the experimental area. As you can see, the experiment is mainly on the boards in the middle of the area, and to the side there are 'finish areas'. There is a finish area – ah, break in character – there is a finish area next to the seats, which is marked out by cones. If you have finished or you want to leave the task at any point, just go to these areas and you will have completed the task and banked your money. On a health and safety note, all of us will be moving, so please be careful where you step, and don't run. There are quite a few of us, so please don't push each other or anything; as we said, there is no time limit. If you receive any instructions during this experiment, please follow them, or, if I intervene, please do as I ask. Ok, break in character. This is the second question: what does everyone think the experiment is about now? What are the key pieces of information that have stood out in the last five minutes or so of me talking? And if we go in reverse alphabetical order, so Mina, if you could go first.

8.40 M: I'm still quite confused about the task. The thing that did stand out to me is that you said there's no time limit on this, and I wonder if it's human behaviour around, you know, doing something under a time pressure, people will work quicker to get things done, you have more uncertainty if you're given a longer time scale to complete things. Um, yeah, I'm not sure at all.

9.08 AS: Brilliant, brilliant. Ok, thanks. Kirsty?

K: So you said something about 'open-plan office', so you're wanting to design an open-plan office and you're wanting to know how people move around each other when they're busy working, so you're gonna, you're, you're going to look at how people move, you're going to give them tasks to do that they're thinking about and you're going to look at how they move and how they respond to other people in the area, so that you can work out the best design for an open-plan office. That's what I think.

09.36 AS: Ok, thanks. Dan?

D: Ok, much the same as Kirsty, um, saliva testing tells me that, rightly or wrongly, we're looking at some kind of stress analysis here, potentially more. Filling in the form, the bit that you mentioned previously, worries me about bank details and such like, and also giving saliva samples. Aside from that, I'm intrigued, and wait to hear more. And that's all I can take from it.

10.05 AS: Thank you very much. And Alan?

A: Yes, I pick up on Dan's intrigue, um, obviously I, I reinforce pretty much everything that's been said before, but I want to add to that, I think if I were a participant, at this stage I would feel it was a very controlling experiment, and I think um, I, I, I think there could be a fair threat level in amongst this, particularly when I heard the mention of saliva samples. So I'm reckoning that this is going to be a fairly fine-grain assessment, and, I suspect, emotional, and I think, um, I think, I think a participant would probably... or many a participant would feel a bit threatened by it.

11.08 AS: Ok. Thank you very much. Um, what I'll do next is I'll continue talking through the experiment. Ok, let's start the experiment. The task will be for each participant to put a sticker on correct mathematical equations. If everyone could open document 3 please. [Alan groans] This is not meant to be off-putting in any way, these are just a small cognitive task, but if you could spend, um, you know, about 30 seconds looking through the equations that you see on your screen, some of them will be correct, some of them will be incorrect, and what we would ask experiment participants to do is to put a sticker on each correct mathematical equation. And on each correct mathematical equation, if they put a sticker on it, they put £1 in their bank. There will be approximately 40 of these sheets, um, across three boards, front and back. So we can spread out a little bit around these boards, um, but again there will be about 10 people involved in the experiment, so you will be close enough to people. Ok. Is everyone, has everyone looked at these questions? Had a bit of a brief, uh, kind of, immersion into what this cognitive task is about.

[General noises of assent]

12.57 AS: Fantastic. Right. Now, in the next ten seconds, what happens is, you're going to open the audio file and what happens is you have turned around and you have seen two people looking at each other. One of them is touching the other on the shoulder and is saying exactly what you hear in the audio file. So if everyone could open up the audio file and listen to it, please.

13.25 Audio: Stop. You have been taken out of the game. You have failed the task, and you lose all your money. Don't move.

13.35 Alan: Ok, I've listened to it.

K: Yep.

AS: Did everyone else hear that?

D: Yes

M/N: Yes, I've listened to it.

13.43 AS: Ok. So here's the next question. What do you think is happening right now? What would you do in this case? Does anything important come to mind? And, uh, we will start off this time with Kirsty.

K: Em, I would be wondering why this person has been taken out of the game, what they've done wrong or, and I'd be a bit scared that I might be next because I'm not brilliant at maths and maybe I've mucked up the questions and I haven't done very well and I think I'd be a bit anxious that I'm going to lose all my money.

14.16 AS: Ok. Dan?

D: I think I'd be feeling that I'd missed something here. I don't see it as threatening, although the tone of the, um, what I've just heard, appears to be threatening in a way that it is deliberately making a statement, 'you will obey' type of statement. Em, other than that, I'm still intrigued. I can't understand why that kind of action would result from what I have just been looking at on, in document 3, on the equations. That's my thought.

14.57 AS: Ok, thanks. Ah, Mina?

M: Um, so, I would agree that the tone is quite threatening; it's quite, quite scary to hear him say that and it's quite unexpected. But my thought was that the person doing the test who, whose been accosted, um, had been cheating, and they've, they've seen that he's cheating, I don't know, doing it on a calculator [laughs] or a phone or something, and that he'd immediately been removed from the game. So yeah, I think that would make me feel quite anxious that, am I being watched by somebody, um, yeah.

15.32 AS: Brilliant, thank you. And, uh, Alan.

A: Yes, um, I support all that's been said before. I continue with my Orwellian view of this, I think I would feel, I would be even more reinforced by feeling that this was a very controlled environment. I think depending on the age of the participant and their numeracy, their education attainment levels and so on, and ergo their levels of self-confidence to some extent, I think there would be a degree of threat.

16.08 AS: Ok. Thank you. Next I want you to imagine that what you've just seen and heard, the next thing that happens is that the person turns towards you and comes towards you. What would you think, feel, and do? [Laughter] And to start off with, please could we have Dan?

M: Is it the person in charge or the person that's been accosted that comes towards you?

16.36 AS: This is the person who you heard on the audio file who was touching the person on the shoulder.

M: Oh, the person talking, ok, all right, ok.

AS: Thanks.

D: Ok, I'd want to hear what this person had to say, I'd be... I keep using the word intrigued, I'd be further intrigued that he was advancing towards me, um, knowing full well I haven't used any calculators or whatever else and I don't really see the need to with this kind of equation, with these kind of figures. Em, so I'd be interested to see what happened, and, given the scenario to date, I'd be quite happy if they said to me, 'Stop, don't do anything, you're out of the picture', so to speak. That's my thoughts. But happy to continue, by way.

17.18 AS: Ok. And, uh, just on that point specifically, within the audio file you did hear the person say 'You lose all your money'.

D: Yes.

AS: Do you remember that?

D: Yes.

AS: And that is still, you're still happy if this person is approaching you?

D: Yes.

AS: Ok! Ok, thank you. Um, uh... ok, so we started with Dan – Alan?

A: Defensive. Um, I think I would be defensive. Because I've already seen this authoritative figure denude my colleague of any money they'd already earned on, on relatively straightforward arithmetic exercises, so I think defensive.

18.04 AS: Ok, thanks. Uh, Kirsty?

K: Yes, I agree, I'd be defensive, because I think I'm going to be immediately waiting for him to tell me that I'm out and I'm going to want to start arguing with him because, well, what have I done wrong? So, so I'm going to think it's most unfair that he's going to take my money away from me, if I've done ok with the task, so. And if he didn't, he didn't explain at the beginning that, you know, there were certain reasons that you could be taken out, then, then, that's not fair, because I didn't know that, so, yeah, I'd be very anxious and, uh, angry. If he was going to tell me that I've failed.

18.40 AS: Sure. And Mina?

M: Em, I'd be the same in that I'd be quite defensive, but my thought was, 'Oh, they're looking at me, maybe they think I've also cheated, so I'd be quite ready to saying, 'Well, well no, I didn't, you can check my bag, you can check whatever, your cameras or what not.' Um, and the other thought I thought at the same time was, 'Maybe they're going to ask me if I saw the person they were accosting', you know, 'Did you see them do this, did you see them take a calculator out of their bag, or do anything untoward?' Um, yeah, it kind of didn't strike me that I'd be booted out of the game as well, the same way the rest of the panel are assuming that might happen. I kind of thought maybe they're coming to me to get my take on what's happened over there. But now that everyone's mentioned it, I'm suddenly really worried about being kicked out of this game. [Laughs.]

19.30 AS: Sure, sure. Brilliant, thank you very much. Um, this also covers what the next question is, and the next question is, what do you remember from what I said as the instructions at the beginning of the talk, anything to do with the financial side of things? What do you remember about the physical arena? Now the thing we didn't discuss there, what do you remember about the physical arena? So, can I ask Kirsty to start – what comes to mind when I ask about the physical arena?

K: So you said that, you said that, you told us that we could put our bags and coats down over there, so I would have thought everybody would have put them down over there. You told us that there's boards, um, with the sums on, and um, you said there's about 40 of them, so it should, I was thinking when we said that somebody might have been cheating, I was thinking they might have been looking over at someone and looking at their answers. But if they're all spread out and there's only 10 people, you would have thought you would have spread yourselves out far enough that you weren't right next to someone to be able to cheat. Um, and there's some, there's some cones, I remember that there's some cones that you go to if you've finished. That's, yeah, that's what I can remember.

20.41 AS: And, so that was, there were 6 boards and 40 sheets of paper, so it's not so large an area, just to clarify that.

K: Oh, not so large an area, ok, well then maybe, yeah, then maybe people were looking at each other's, em... With regard to the financial thing, I remember you saying that you, you know, you're going to bank this money and you didn't say that, you didn't tell us how you would lose the money. You didn't mention how you would lose the money.

21.09 AS: Sure, sure. Thank you very much. Uh, Alan?

A: Uh, all the foregoing, in addition to that, um, key points that struck me were late December at 4 o'clock in the afternoon, don't know why but I just thought it might set the mood! You talked about open-plan office areas; you talked about saliva samples; you talked about earning up to 40 quid – but I think as I remember, coming away with a minimum of about half an hour's rate, as I recall? Something like that. Em, and I think you also mentioned the number of people? I think you mentioned there were 10 people, I think, later on? And that's all I remember.

21.54 AS: Brilliant, thank you very much. Mina?

M: Um, so, yeah, I remember what you said about there being somewhere to put your coats, um, so you put them all away in the corner. There are 6 or so panels and lots of these sheets on them. I didn't remember the thing Kirsty just mentioned about the cones being in the corner, but I did remember about earning... earning half a minimum wage, but actually I can't remember what the circumstances around that were now, what the reason would be that you'd get half the minimum wage, but that stood out.

22.27 AS: Ok, thank you very much. And Dan?

D: Ok, em... I don't know whether it's cheating, but I was writing these things down as I started. [D and AS speak over each other.] Sorry, Alastair, I didn't hear that.

AS: No, no, that's absolutely fine. What part of what you've written down and what you remember stands out?

D: Em... the fact that it was a sports hall scenario; the fact that it was late December, approximately 4 pm; the seating area; the coned area that was by the seating area, if you wish to, as I put it on my shorthand bang-out, to finish; there's no time limit; you can leave at any time; and what the task was. I actually hadn't recorded the ten people that was mentioned; the 6 boards with 40 sheets of paper, yup, and the up to £40 with a minimum half hour. Em, I have to say that my thoughts were that I'm not monetary influenced in this particular area; I'm just intrigued by being able to understand what we're trying to do within the scenario, and that's my feeling – that's all I've got as my memories.

23.45 AS: Brilliant. Thank you very much. Um, ok. Now I'm going to go back into character. Um, and what happens in the next 3 or 4 minutes of the experiment is the person that you saw touching the other participant on the shoulder and who was saying the things that you heard on the audio file continues to go around touching people on the shoulder until he touches all of them and everyone stays still, or, until everyone has moved into the finish area. Um, so what you can imagine is, if you see this person coming towards you, he will touch you on the shoulder and say exactly those same words, and you will lose all your money, or he'll move towards other people and you will see him doing exactly the same thing. going consecutively around all 10 people. He will avoid people who are in the finish area. Ok. Now, the experiment has now finished. And I'm going back into character. Ok, thank you everyone – the experiment is over. Can everyone move to the side please. I would like to start by saying that there will be no financial penalties on anyone who was touched during this experiment. I will now take you all through a debrief, after which we will take our final saliva sample. First of all, our experiment was looking at movement behaviour, but during stressful scenarios. We built up a cognitive test and financial incentives to make you think that you had something to lose if you were taken out of the experiment. And then an actor was used to imitate an aggressor. However, as I said, all of you will be paid the full amount and no penalties will be taken. The saliva samples are used to test for cortisol levels, which are strongly correlated with a 'fight or flight' response. We measured your position through this experiment with the tags, and these will provide us with quantitative measurements of how people move when they are in extreme emergencies. And we use our emotional questionnaire and the saliva samples to have a corresponding level of fear. I would ask that you now complete this handout – and, break in character, could everyone please open up document 4 and 5. So does everyone have that open?

[General assent]

26.19 AS: So, in document 5 you can see exactly the same questionnaire and what I would be asking the participants is to fill it in, um, fill in how they felt during the experiment, not now during the debrief. And in document 4 there are more, further questions related to how people feel as part of a group, uh, and how people feel towards the hostile member of the group. Um, there are also hypothetical questions of what they would have done, what they wanted to do, or if they wanted to do something differently, why did they not do it? Um, there's also a further question of, did they guess the real aim of the experiment, and, if so, at what point, because that will be a confounding variable. Ok, and I'm going back into character now. Uh, thanks. Now we're going to hand out some contact details for you if you need to contact any health services as a result of this experiment. If you experience any lasting effects, please either get in contact with your GP or any service on this sheet, or email myself and I'll get you in contact with the correct people. Finally, as I'm sure you can appreciate, this experiment requires that the participants be unaware of the true nature of what is being measured. Please could you refrain from talking about it for two days, as this is our data-gathering period. This is especially important if you know anyone doing the experiment. After the two-day period, you're free to talk about it as much as you like. And at this point I would hand out the second saliva sample. Ok. Ok. That's the end of my talk-through, and I have a couple of further questions, um, and first things first, uh, it's, the, the question is kind of summarised as, 'What do you think?' What emotions did you feel as the experiment progressed? And now... and we'll go on to the next question afterwards. So, what do you think, and what emotions did you feel as, as the experiment played out? And could we please start with, eh, Alan?

28.33 A: Just repeat the questions for me again, Alastair, will you?

AS: Sure, sure. So, the question is, what did you think during the experiment, and, what did your, how did your emotions, uh, progress throughout the experiment?

A: Right, thanks, I was just, uh, I was just double-checking on that, so that

I could structure an answer to you. But I'm not sure I can separate think and feel.

AS: Ok, yup.

A: So let me start with feel. I think, I, I think were I a participant in that, I would have been fairly finely tuned to what the experiment was about in terms of measuring my reactions to things. I think the saliva test was a good giveaway, as was the questionnaire about emotions. So I think I would have been pretty wise to what you were out to collect. Um, and how did I feel about it, going through it? Well, I don't know if it's a question about me, or a question about me as a potential participant... If it were me, I think I would have felt, um, quite defensive throughout, and I'd have been relying upon my... ability and... innate self-confidence to handle the situation. Now, when I think about it, [INAUDIBLE 30.05] earlier on and I think I stick with that. I'm slightly uncomfortable about this experimentation. I think it would work better with a younger group than an older group, um, I could explain that but I'm not going to take time, I've taken quite a long time, um, and I think, not least because of your trail [?] at the end about, you know, contacting, contact details for GPs or other medical or psychological sources in the event of [INAUDIBLE 30.40] [laughs], my thoughts are it's, it's quite a, it's quite a controlling experiment, and I would be cautious about using it. I think, ah I think, I think the populations should be chosen quite carefully for something like this.

30.58 AS: Ok, thank you. And Mina?

M: Um, I, so I think to start with, it's quite exciting when you first walk in, and I don't really know what this is about, either as a member of this panel or actually being somebody who participates. I think that was quite intriguing; I was quite keen to know what was going to happen next. I liked seeing all of the calculations come up because I thought, actually, I'd be able to do these quite easily, and maybe it's quite an easy way of spending the next half hour, I'll probably score quite highly, and I think I'd probably be a bit competitive, a bit, 'Oh, I think I'm going to score top marks, I'm going to catch every single one' [laughs]. So that was quite fun, quite exciting. When I heard about the person dropping out, I thought, oh, that's, I don't know, a bit of drama to the day, probably, didn't really think much of it, um, but then when you said they started tapping other people and kind of, I don't know the reason, the criteria for that sort of thing, you've lost and you're going to lose your money, you can't participate, that made me really, really quite anxious, um, I don't know what basis they're choosing people on, is it something they've done, is it, I don't know, a way that they look? I, I found that really quite worrying, um, and at the end when you explained that actually that was all part of it and, um, you know, you're going to get the reimbursed and treated equally, um, I felt quite a big sense of relief, actually, just being on the panel, and I thought, 'Oh, that's fine, everything's ok now', um, but I do feel like it's a really mean experiment to run, it made me quite anxious during it, um, and I think hearing a list of organisations at the end, I did think, oh, you know, don't be silly, you wouldn't need that degree of help with it, I don't think I'd need to phone somebody up and say, 'I feel really upset at what's just happened', but actually, um, maybe I would. It leaves quite a bitter taste in your mouth, I think. You went into this trusting the situation, and something quite horrible happened in the middle of that, um... but yeah, it's really interesting, and I think it does the job exactly. It would, it induces a high stress environment onto you, you wouldn't know what to do, but it ends quite quickly, so it's evolved quite quickly. Um, yeah.

33.25 AS: Brilliant, thank you very much. Uh, Kirsty?

K: Uh, yes, so I, again, yeah, similar to Mina, I was quite excited to start with, and I felt, you know, quite looking forward to it. And then when the, em, and I'm not brilliant at maths so I wouldn't have been quite as excited about that aspect of it [laughter], but I would have given it a good go. Em, but em, but when the first guy gets, when the first guy gets tapped on the shoulder, then I would, I'm going to start to get quite tense and, and I'm gonna, and I like everything to be fair, so if I can sense that there's a bit of injustice, there, I'm going to get quite angry, and I'm gonna, I think I'm going to be wondering at the time, well, 'Should I be saying something?' Should I be speaking up for them and saying, well, 'Hold on, what's going on?' I probably wouldn't have done because it's a controlled environment, they've told you you have to do what you're told, so I probably would keep quiet, but bubbling underneath me, I'd be going, I'd be thinking, 'Well, this isn't fair, why is, why has that person been picked, what's going on?' I'd be wanting to ask questions and I'd be, I'd be feeling that, I'd be in a bit of turmoil underneath. And I think when the next two or three, when the next one or two people got picked I'd still be feeling that, but then I think after a while, when everybody's getting picked, I think I'd think, 'Ok, ok, this is, there's something funny going on here', um, I don't think I'd get more and more stressed as everybody got picked, I think I would, I think I would think 'Ok, then, there's something. . . ' I don't think I'd have worked out what was going on, and I don't think the saliva and everything would have clued me in to what you were testing, stress, and it wouldn't have occurred to me at all. So I don't think I'd have though that, but I'd have, I would have thought, 'Well, this is something' [coughs] excuse me, 'Well, this is something to do with, this is part of it', and I don't think it would have made me more and more stressed as it went on; I think it would have, I would have calmed down a little bit. But I still would be wondering what's going on and is it fair or not, and then obviously at the end the relief. And also, like Mina said, I, when you said about the doctors, I was like, 'Oh, that's a bit silly', you're not going to want, you know, it was just [laughs] it was just a little bit of anxious, you know, anxiousness about whether I'm going to get my 40 quid or not but it's not the end of the world, is it? But then you never know how some people are going to react to things like that, I'm quite a calm person so, um, I don't think that would affect me, but maybe other people would be really, really, um, a bit traumatised by it, so, um, but yeah I'd be really relieved, and everything's fine in the end, so that's, yeah.

35.34 AS: Okey doke, brilliant. And Dan?

D: Yeah, um. . . In my particular case, I always had in the back of my mind – or at the fore of my mind, to be honest – the fact that I had volunteered for this, so I didn't at any time believe that I would be personally threatened, so I never felt anxiety, or never felt that there was the need to be anxious. I was very intrigued, and I had heightened awareness when we had the first part of the scenario with the tapping on the shoulder and the words that we heard. Em, that made me more defensive, defensive is probably the right word and has been mentioned a few times, em, and intrigued further, and that level of intrigue is the one that it comes down to in my thought process, em, as to what on earth is going on, and why one of the ten people would be taken out so early, and I couldn't imagine how he's cheating, etc etc, so that would have left me to think, 'Well, I don't know what's happening here, I have no idea, I have no clues, I have no evidence to either support or deny the fact that they've been taken out', so as long as I'm playing the game – I use 'the game' in inverted commas – the way I thought the rules told me it should be played, then I'd be quite happy. Other than that, I. . . do see the need for the doctors, I do recognise the, the stress and the anxiety levels and such like, um, it doesn't apply to me and I'd like to think that it wouldn't impact too much on other people. However, throughout the whole thing, I would have been very aware of the levels of the other participants in there with me, um, and how they were feeling, and the possibility that somebody might need to help somebody here, if this gets out of hand. And that was my thought process. The other side of it is, I would have gone, as soon as I realised that there was something happening, I would have gone to the finish area to retain whatever cash I had, and take myself out of the tap-on-the-shoulder situation. And that's about all I've got.

37.47 AS: Ok, thanks. Can I, can I pick up on two things you said there? You said, 'if things got out of hand'. Now, what specifically did you mean by that?

D: Well, there's not too many people in there, um, if I felt it was becoming… if the situation was becoming hostile, then, I'm afraid, I'm a volunteer, I'd terminate my voluntary contribution this and walk away.

38.13 AS: Ok, so that's not, um, you're still imagining this from a completely experimental point of view – if you're unhappy with the experiment, you're happy to leave, and you're aware of this.

D: Absolutely.

AS: Yeah, so it wasn't a case of, you think people are actually going to fight, uh, because…

D: Oh, no, I didn't think that.

38.35 AS: No, you were aware that it's an experiment? Ok. And secondly, in the finish area, um, one of the key thing we're looking at is, uh, how quickly people identify, um, that there is a threat, or in this case a financial threat, uh, but if, in the case of an emergency it might be a physical threat, how quickly they identify an area of safety, and how quickly they move towards that area. Um, you, you say that you were always aware that there was a finish area and that if the person was coming towards you, you would move towards that finish area?

D: No, no. The first time that it happened, uh, when the person was coming towards me after the, the little soundbite that we heard, the audio file, em, I recognised what was happening there. We were then asked to continue with the experiment and, em, carry on doing the same arithmetic problems. Em, if then I saw people, other people being tapped on the shoulder with rapid succession, I would then move to the finish area to protect what I had. Now in the end I understand that that wouldn't make any difference whatsoever, but that was my immediate reaction during the scenario.

39.50 AS: Ok, brilliant, thanks. So that was probably what I said then, so I'm going to change that for what I do next. The aim here was to not give you any prompt after you hear the audio scene, so, so what I'll do is change that in what I do next, with the aim being, yeah, if you recognise that there is a financial threat, do you move to the finish area? That is one of the questions. Um... ok, so yeah. Final two questions. Um, now that you know what's being measured, and why, would you sign up to this experiment in the first place? This is basically because we can't provide all of the information beforehand, we have to ask afterwards, would you be happy doing this experiment again, in the knowledge that you would be naïve? And... so let's start with that, and then we'll get on to the next question. Um, so, Mina first.

M: Um, that's a really hard question. I think the... I didn't enjoy the period of feeling anxious and thinking, 'Oh, is it going to be me next?' or 'Why are people being picked?' So I think knowing that that was going to happen, I kind of think maybe not, but if I had a bit more context about what you intend to do with this – I mean I know you talk about the open-plan offices, but it's not really emotive enough reason for me to think, 'Ok, I'll go ahead and do this'. Um, but if I thought it was for something that, I don't know, a bit more useful maybe, then, I would.

41.29 AS: Right, ok, so maybe I didn't explain very well. The aim here is to provide movement data on how people react during emergencies such as terrorist attacks. Um, so were you to be walking down the street and you see, uh, an emergency such as that, where do you move? Now the aim of this data is to inform emergency planning, so, talk to the police and tell them what to expect, or to inform building design, so, how can you change the layout of a building to improve, kind of, how we move in the case of an emergency. And as it stands, this data just does not exist. Um, so the, the aim would be completely exploratory, to say, 'Right, how fast do people move? Which direction do they go?', this kind of information. Um, and sorry, just to say, so you would say, Mina, you would say probably not, knowing what you know now?

M: No, actually, I think I just missed all of that, that context that you provided, maybe you said it earlier, but knowing that, yeah, I would, I wouldn't hesitate to do something like this if I knew that, if I knew this beforehand.

42.50 AS: Brilliant, sorry, yup. Thank you, thank you. Uh, same question, so, knowing what you know now, uh, would you, uh, consent to be part of this experiment, to Alan?

A: Um, well, I think, I think yes. And I think that's probably driven a bit by myself, in the sense that I'm quite a competitive and curious person, curious I think before competitive. And therefore, like Dan, I mean, I would have a heightened curiosity during all this, and, and because I've got a fairly competitive, driven nature, I, I'd be wanting to succeed in it. So yeah, I think against the background of all you've told me, you know, I would. One comment I would make, though, however: I was very struck on the flyer that went round earlier, a week or two ago, whenever it was, about the exclusion criteria. And that made me – you remember there was exclusion criteria relating to high blood pressure, coronary disease, and anxiety, depression, psychological issues?

43.53 AS: Yep.

A: So therefore, I was reckoning that this experiment was going to be, to some extent, stressful slash threatening slash, you know, a bit of a wind-up, so to speak. I was kind of anticipating that, and therefore, against that background, would have participated. Is that helpful?

44.16 AS: Yep, thank you very much. Um, ok, and Kirsty?

K: Ah, yes, I'd definitely would have found that fair, because it wasn't, the stressful situation wasn't so bad that, you know, you were not going to get over it, it wasn't that exciting, and I, I quite like doing this sort of thing, I like research stuff, so yeah, I would definitely, would definitely have done it. And also it would be really useful, so, yeah.

44.41 AS: Okey doke, thanks. And Dan?

D: My immediate thoughts, as I'm volunteering, um, yes, I would go for anything like this. The whole idea of being able to assist in the future, if it meant that, I don't know, lives would be saved or, working environments would be improved or whatever level between one and the other, from one extreme to the other, I'd be quite happy to, um, give my time happily to, yes.

45.13 AS: Ok. Thank you very much, guys. There is the final question, which is where we're going towards optimising the experiment. Is there anything you'd do differently? And more specifically, how scary can this actor be before you would say, 'Actually, that's too much, that's going into unethical'? Um, because, and this is, we, we've picked up on this quite a, quite a few times, there is only a certain level of fear that is produced in this experiment.

Now this will be dependent on each person, but at no point do people think, 'My life is in danger' – which they would in the emergencies we're trying to model. Um, so, the question is, how, how far would you be happy to go, um, in this experiment? Um, and would that, um, uh, Dan?

D: Ok, unless I saw physical violence, or the possibility of physical violence, I'd be quite happy to go ahead with it, again, knowing that it's an experiment and I'm a volunteer.

46.24 AS: Ok, thank you very much. Uh, Kirsty?

K: Em, the thing is, because I know that the person that's tapping you on the shoulder is part of the experiment, I'm going to trust – because I've signed up for this – I'm going to trust that it's, no matter how harsh his voice might seem, it's all in, it's all how it's supposed to be. What would be more scary to me would be if it was someone outside that, or if one of the group, if one of the group who you thought was one of the members, you know, one of the participants, was actually an actor, and if they started getting stroppy with the person, if they got angry, that would intense the stress more, because I'm not going to believe that the person running the group is going to do anything, you know, really stressful and really scary. Because they wouldn't, it's their experiment. But if it was somebody else, you know, somebody else that wasn't part of it – or that they were really but they weren't, you know, I didn't know – then I think I would be more on edge. And I'd be willing to be pushed to be a bit more scared, I wouldn't have a problem with that, or I probably would have at the time, but I think you could get away with it. Um, but for me it would be more believable if it wasn't the person running the group or whatever.

47.30 AS: Can I, can I cut in there? The person that you saw touching the person on the shoulder, he was a member of the group. Um, I didn't make that clear, um, and I will in future, but that...

K: But when you say a member of the group, what...?

AS: As in, he is a – you, at the beginning, think that he is a participant in the experiment, uh, and being paid much like yourself.

K: Oh! Well, then, that... that confuses things even more, because I would think, 'Why is it up to him who gets tapped on the shoulder?' That's really confusing me now.

48.00 AS: Right, ok, so what I'll do...

K: But why is it his choice? I thought it was the person in charge of the experiment.

AS: Right. Ok, no, I'll make that clear. Um, does that change anything – obviously, we can't go through all of the questions again, but does that change anything in how you feel about the experiment, any of the questions I asked?

K: Yeah, because I'd wanna know why is it his job? Why is it his... If he's just a member, why is he going round tapping people on the shoulder? It's nothing to do with him. That would be, I'd be wanting to question that.

AS: Ok. Ok.

K: It doesn't make sense.

48.33 AS: Um... right. Now you say that that would make it more scary. Would you be happy to do the experiment in this particular scenario knowing now that he was a participant?

K: Yeah? [Pause.]

AS: Yeah? Ok. Um, and just to clarify, uh...

K: But I don't know why people would take notice of him if he's not in charge. Why would people take notice of him?

AS: Mmm.

A: Mmm, mmm.

48.56 AS: Now, is everyone else... did everyone else think that this person was a, uh, an, an investigator rather than a member of the, the group?

M: Yes, I thought that he was an investigator.

AS: Ok.

A: And I also thought that he was not a member of the group.

M: Yes.

49.17 AS: Right, ok. Um, in which case, what I'm going to do, I'm going to slightly change this. Can I go through, uh, and ask one final question to all of you, and ask for overall impressions? Because, uh, I will need to make that clear in future. So, going back to it...

A: Alastair, Alastair, before you go there, Dan may [INAUDIBLE 49.40] I'm wondering if I could just comment on your previous question, and just throw one thing in. This question about, you know, would one participate or not against this [INAUDIBLE 49.49] of information. I'm very struck by make-up of the population of your group.

49.59 AS: Um-hum.

A: And I think, um... I think, I think reactions will be, are bound to be determined to some extent by the, um, socioeconomic band, probably, age make-up of the group. In other words, my, my, my hypothesis is that younger and fitter people – I know what the eligibility criteria were, but, younger and fitter people are probably keener to participate than more mature, less fit people, both physically and psychologically. There's a thought for you.

50.39 AS: Sure, thank you, thank you very much. Uh... Em... Right. How do this this? Um...ok, so before I ask that last question, actually, uh, Mina?

M: Yes?

AS: Would, knowing now... what you know in terms of the, the hostile member being a, an original member of the group, being a participant, uh... Would you, uh, sign up to the experiment in the first place, and also, is there anything that you would do differently in this experiment? Specifically to improve it?

M: I think I would sign up. I think that when you explain now that he's a participant, he's kind of, one of us, if you like, and he's not one of the organisers, that would worry me more in the experiment, because that's definitely, um not what I was expecting to happen, I would expect an outside party to be kind of managing the experiment. So if I saw a, another participant suddenly say, 'Hang on, you can't take part, and actually I've taken your money away' or whatever, that would ramp up the fear a bit more, probably, a bit more to the kind of levels that you're thinking. Because it's not just fear, it's then confusion as well, and 'Hang on, what am I really here for? Are these people really participants?' So it's probably a different kind of fear that you're, that you want to replicate, but it's still an elevated risk of fear, whereas it's an examiner or somebody who's part of it, you'd be sort of, you'd be thinking, 'Well, they've just broken the rules and that's that.' Um, so yeah, I would, I would participate. I can't remember the second part of your question? Would it change...?

52.15 AS: Would you, would you improve it in any way? Would you make the actor more scary, less scary, just as scary as he was?

M: I think you could probably get away with making him a little bit more scary, because, um, I agree with the others... I mean, you're a certain kind of person if you agree to enrol in an experiment where you don't really know what's going on, so I don't think you'd be kind of long-term affected by the anxiety that it causes, you know, I'd hope not. So I think you could get away with making him a bit more threatening

52.48 AS: Ok, thank you very much. Um… right, is there anyone I've missed on that particular question, sorry, the improvements to the experiment question?

D: Yeah, you've missed me, Dan.

53.02 AS: Dan, Dan, thank you.

D: Yeah, um, immediate thought are that I think now that I recognise that it's somebody within the group and I didn't know anybody in the group in the first place, I would want to know, um, without asking the question, I'd be very, very aware that somebody within our group has deemed to take control of what is happening. I would be watching that person very, very carefully, um, and I would be… um… yeah, I'd, I'd be wanting to make sure that I was ok, and I'd be worried for other people as well. Um, I would think that having the scenario given the way you've just explained it, that person is scary enough in themselves, um, I'm still a volunteer and I'm still going into an experiment, I would have that all in the back of my mind, um, I, I hear and agree totally with what Alastair [sic Alan] is saying, yeah. I'm the kind of person who would want to carry on and look at this – intrigue is the biggest thing I've got into it – but all of a sudden with somebody within our own little group that we've been brought together, suddenly taking a level of control that I hadn't anticipated, I'd be watching them very carefully. Second part of the question is, I wouldn't want to make it any more scary than that, I think it's scary enough that that is happening.

54.25 AS: Ok. Ok. And would you, would you sign up to this knowing now, knowing what you know now?

A: Yep, I'm, I'm always willing to try and improve things in some way or another. I recognise the fact that my participation has been, the opportunity to participate has been given to me, I would never stop participating in this kind of, uh, thing, particularly for the PHE.

54.53 AS: Ok, thank you very much. Um, now, Kirsty, is there anything else – because we kind of moved towards a different questioning branch.

K: Yeah, just that I wouldn't really believe that this person… I wouldn't believe them if they're a person, if they're a person in the group, one of us, and then they start tapping people on the shoulder and saying that they, they have to go out, I would be questioning that. Why is them? I'd be going and talking to the organiser and saying, 'Well, hold on a minute, is this part of it?', because I just, it would completely confuse me and I wouldn't believe it.

55.28 AS: Right.

K: So I know you're trying to present a, a, a, you know, a threat and it's gotta be in the way that it's doing it, but for me it doesn't make sense. It doesn't make sense. To me what would be a more scary situation would be if the organiser said, came and tapped to on the shoulder and said, 'Right, you're out', then I'd believe them, and then someone, and then one of the members started to get angry about it, that would be a more scary situation for me. I just think this one I wouldn't believe it. Because who, who, how can someone that's not, how can someone who's involved as a participant start telling people to come out of the game? That doesn't make sense.

56.01 AS: Um-hum. Um-hum.

K: That's my overriding, my overriding thought, knowing, knowing that it's, that that's what it is.

56.07 AS: Ok. And what could myself as the experimenter, or the, uh, the, uh the, the authority member, I suppose, what could I say to you that would, uh, mean that you continued with the experiment, uh, in the way that it's meant, now that you know it?

K: Uh, what, in the middle of it, once I'd got a tap on the shoulder?

AS: Yeah. So if you turned to me and said, 'What's happening?' What should I say at this point?

K: Um, well, if you just said, 'Oh, it's fine, oh it's fine, just carry on'. Yeah, because I just want to know that everything good, going correctly, so if it's fine, if that's part of it, then yeah, that's ok, fine.

56.49 AS: Ok, ok.

K: I'd believe you, because you're in charge.

56.52 AS: Right. Uh, thank you very much. Now, that's us come to the end of all the questions. Does anyone have anything particular that they want to say at this point? And, uh, if I've missed anything? Um, if so, please bring it up, I'm, I'm, I, as we say, we are developing this at the moment, so, anything at all.

D: Yeah, so, just if I can step in – this is Dan speaking, Alastair. Of course, when I was talking earlier on, I said I 'as Alastair said', I meant 'as Alan said'. Sorry, I mixed up the names. Um, the other side of it is that the question that you've just asked, I fully agree, but I think if it was somebody who was in a controlling, um, in a kind of a controlling environment, if you then said, 'Please carry on with the experiment' rather than say, 'It's ok' – it's got a little bit, yeah, just making it clear as to the fact that this is an experiment. Other than that, I will shut up now.

57.55 AS: Ok, thank you very much. Anyone else? Any, any comments at all?

A: Uh, I think, uh, I think, just a couple. As to what Dan said, I think we all realised that he got the names mixed up. Two things, I think, um... I think it will be important to strike the right balance between surprise and a bit of pre-briefing. That will be a matter, that will be your judgement call. It did strike me going through this that, um... I think I could have taken a little bit more information at the outset without it necessarily reducing my fear [INAUDIBLE 58.41] about the surprises that are following. And I just thought I would ask you to question that a little bit. It's Dan's point about... a lot of people who volunteer – and that probably includes the four of us who are here today – are happy to do things for public good, if you follow me, so I think there's a balance to be struck there. The second thing I just wanted to say, in case I forget before we finish this, is that you might think about whether this thing works, um, best as a phone call or, or, or best in a face-to-face situation, um, just reflect on that. Uh, I, I suspect that there's a discussion to be had around that. About, you know, your work with us today, um... Just a thought for you.

59.28 AS: And just on that, of all the participants, you were the only one who mentioned the word 'threat', at the very beginning in question number one. Um... so, so do you still kind of think that, that you could have had more information and that wouldn't have pushed you towards more of the threat kind of priming.

A: Yes. Yes, yes, yes, I think, yes, I think so. And the other thing I would say there, uh, was, I completely support all that's been said about confusion arising from the identifier of the toucher, because I think that also affects, um, one's aggressive, defensive, um, response or, or one's feeling of threat. So, um, yes, is the answer. Sorry, a long way of saying, yes, a little bit more information would have changed that, could have, effect.

1.00.00 AS: Brilliant. Thank you very much. Um, ok, Mina and Kirsty, anything at all? Now worries if not. [Pause] Hello? [Pause.] I... Is everyone still here? [Pause] I think my computer's stopped. No, I can hear something. [Pause] Hello? [Pause] Oh no, my computer's giving up a little bit. Um. [Some scrambled noises; laughter.] Hello? Hello? [Computer pings.] Ah no. [Some blurred background speech; Alan seems to still be talking.] Hello, hello, hello – sorry, I'm, I'm back now.

A: Yeah, yeah, I thought the same thing.

1.01.28 AS: Hello. Are you all here?

A: Oh, he's back!

AS: Sorry, my computer died a little bit. Um, so, unless I missed anything, I was asking if Mina and Kirsty had any further, um, further comments or thoughts?

M: Eh, no, no, all good, but it sounds very useful and yeah, would definitely take part.

AS: OK, thank you.

K: Yeah. I think, um, I don't know how much context you could provide, but providing that bit of context so we know how useful it's going to be would make you be able to manage this, um, as a participant you'd be able to the sort of manage the scary situation a bit better if you knew actually this is exactly, you know, this is, this is what we're looking for, we're interested in something that is going to cause some fear and some anxiety. Um, but this is the endgame, this is what we're looking at. I don't know how much of that you'd be able to give away without kind of, you know, limiting the usefulness of the experiment, but a bit more context probably would be, would be helpful.

1.02.35 AS: Probably would be helpful. Ok. Ok, uh, thank you very much everyone – you've been absolutely amazing, um, both in developing the experiment and developing the protocol for the pilot study of the experiment, so thank you very much

[Chorus of 'thank you', 'good luck' and 'goodbye'.]

B.2.2. Interview 2

AS: So the recording has started. Could everyone please introduce themselves for this?

[Overlapping conversation]

AS: Sorry, that's my fault, that's my fault. I should, uh, what I'll do is, I'll call out names. So, uh, Liz, could you introduce yourself?

L: What do you want me to say, a little about myself?

AS: No, just name.

L: Hi, I'm Liz.

AS: Scott?

S: Scott.

AS: Jenna?

J: I'm Jenna.

AS: And Derek.

D: Derek Clarke.

0.34 AS: Brilliant, thank you very much. Um, ok. Now, uh, you've all been sent a, uh, recruitment document which, um, was sent earlier, uh, detailing some of the things about this, uh, experiment, that we will be using to recruit people when we actually do the experiment. You should have also been sent, um, several documents that, uh, I'm going to ask you to not open, but they should be documents 1, 2, 3, 4, and 5, as well as an audio document and a picture file. Does everyone have these?

[General assent]

1.14 AS: Brilliant. [Jenna's speech overlaps.] Oh?

J: Um, yeah, I'm just going to, yeah, I think so, yep, at the bottom, yep. So I've got, yeah, I think, yeah, document 1, document 2, document 3, picture, document 4, document 5. I've got those.

1.30 AS: And an audio file.

J: Yes, yes, yes, yes, that's beside, a little audio file, yes, cool.

AS: Fantastic. Ok, right. So what, um, let's get into to it. What we'll be doing is, I'm going to be reading through a script that I've written. So some of it might come across a bit stilted, but at certain points I'm going to break character and I'll ask you various questions such as, uh, 'What do you think about the experiment?' etc. etc. etc. Uh, and at that point I'll ask you individually what you think, um, and ask you to potentially imagine some areas. Um, so, without any further ado, could I please ask the first question? The first question is, um, for, one second, I've lost my order… The first question is, what does everyone think this experiment is about at the moment? You've been given very few details, but including the participant recruitment pack. Um, so what do you think the experiment is about, and that please is for Jenna.

J: Em, sport? [Laughs] That's really… and that's just by looking at the picture, I haven't really looked into it at all.

2.52 AS: Mm-hmm? And, and what impression did you get from the recruitment sheet. Did you, did you get any impressions at all or no?

J: I didn't look at it, to be honest with you. Sorry.

3.05 AS: Ok, no, that's absolutely fine. Um, and next would be Scott, please. What do you think the experiment is about?

S: I would imagine some sort of maybe physical activity? Uh, based on what I recall from the recruitment sheet, which talked about needing to be fit and, ah, able to do various things which relate to physical activity, it seemed to me.

3.34 AS: Absolutely, yep, sure. And Derek?

D: Just for, um, [INAUDIBLE] the recruitment procedure, in this case for the example for some sort of physical activity.

3.50 AS: Ok. Fantastic, thank you very much, guys. So, um, I'm going to get into character now. The, uh, just to set the scene, this is a late December Friday, it's about 4 pm, and you, if you could imagine yourself going into, uh, the location that, if you open up the picture file, uh, will show sports hall, so as you, several of you commented on, ah, it's a sports hall, so you can imagine, like, the sports equipment and the sports people around you. There are approximately ten people that are going to be doing this, uh, experiment with you, um, so if you could imagine yourself going in as part of a group, um, and just standing in the corner of the sports hall. In the middle of the sports hall, if you could imagine several boards with pieces of paper on that you can't read just at the moment. Um, and, uh, I'm going to now get into character as the, the, instructor. Ok. Good morning everyone, thanks, now that everyone's arrived we can get started. Please feel free to put your bags and coats over to the side, uh, and, what we're here to do today is to investigate the movement of people in groups, specifically in relationship to cognitive tasks. An example application might be the design of open-plan offices, looking at how people move around each other while they're doing stuff. Before we get started, we will need to spend a few minutes filling in forms. Please take your time to read them. They'll say that we will be taking personal data and using it in the research, but this research will be anonymised and stored securely, as well as never being used for anything except this research. Ok, I'm going to break character, guys: could everyone please open up document 1 and document 2. And in document 1 you should see an emotional questionnaire where you would be asked to put in a number corresponding to each emotion as to how much you feel it at the moment. So if you're feeling particularly, uh, happy, then you'd put, um, a 7 next to a happy. If you're feeling not grossed out, you'd put a 1 next to the grossed out. And then in document 2 there's some further questions about the members of the group. Um, has everyone seen these, happy with what they are?

[General assent]

J: And do we answer every single one?

6.35 AS: Ah, so in the experiment, absolutely you would, but I'm, uh, don't do it because because it's, um, it's just for information at the moment

J: Oh, ok.

6.45 AS: Thanks. Ok, I'm going back into character. Ok, once that's done, can you please fill out these forms, using these ID numbers, and also wear these tags. These forms tell us several important pieces of information, including your bank details, so that we can pay you. In addition to the emotional questionnaire form, we will also need to take saliva samples now.. Ok. Now we can start on the experiment. As I said, we'll be looking at how people move while they do cognitive tasks. So we need to track your movements and check saliva samples. During this cognitive task, there will be financial incentives to do well. Uh, please take these stickers, as they will help you in the task. And now I'm going to hand out individually coloured stickers to all of you. You can earn up to £40 in this task, which we expect to take around 25 minutes. But there is no time limit or anything, so don't worry about taking longer if you need. Please keep in mind that if you are taken out of the game, or you fail the task, you will lose all of the earnings that you have built up, and be paid a minimum of half an hour's worth of minimum wage. You must follow any instructions that you receive during the experiment, and you will be able to read some instructions on the boards for the task once we start. But for the moment, keep the stickers, because they'll be used to keep track of how much you earn. Ok. So now we're moving into the experimental area, again within the sports hall. As you can see, the experiment is mainly on the boards in the middle of the area, and to the side there are the 'finish areas', surrounded by cones, next to where you put your bags earlier. If you have finished or you want to leave the task at any point, just go to these areas and you will have completed the task and banked your money. On a health and safety note, as all of us will be moving, please be careful where you step, and don't run. Uh, there will be quite a few of us, so please don't push each other or anything; as we said, there is no time limit. If you receive any instructions at all, please follow them, and if I intervene, please do as I ask. Ok, I'm going to break character again, and I'm going to ask the second question. So now that I've had a bit of an introduction, and you've imagined yourself in the experimental set up, um, can I ask: what does everyone think the experiment is about now? What are the key pieces of information that have stood out from the last few minutes? And can I ask that to Derek first, please?

D: Uh, it's an experiment about movement and interacting with other people, uh, but also with quite a lot of detail with what to do if you've completed the task, uh, and that sort of thing, so I suppose that brings in some sort of cognitive behaviour and test.

9.55 AS: Sure, brilliant. And um… is there anything else that sticks out from the, em, from the instructions in terms of the physical arena or any of the instructions that stick out other than that?

D: Uh, not that I can think of, other than the tags that they're given to identify them, that's, well that's needed for identification when everybody's moving round, I suppose.

10.21 AS: Okey doke, brilliant. Thank you. Same question to Liz.

L: Um, so, yes, I, can, I'm thinking it's about people moving around and how they feel about other people and movement going on and tags and also the, um, incentive, you've the monetary incentive and would anyone be really, you know, um, uh, angry if they did something and then lost it all and did this or something like that. Um, so, um, and I'm, it's kind of… Is the picture I have supposed to have cones in it?

10.55 AS: No, no, this is just a stock photo.

L: Just imagining it. So the cones, so I can see a bit of confusion, well, maybe a bit of confusion and bumping into each other and if, how people feel about that and… tags and how they feel, um, maybe, I don't know, being labelled and, um… there's a lot going on. Busy.

11.20 AS: Um-hum. Brilliant, thank you. Scott?

D: I missed the financial aspect of it – it's Derek.

AS: You missed the financial aspect of it? Ok, do you want me to re-go over that?

D: Um, just, an important point is that you included an incentive to do the task well, and a disincentive to fail. So that's important.

11.43 AS: Yes, yes absolutely. So you can earn up to £40 in this task, uh, but if you are taken out of the game or you fail the task, uh, you will lose all of the money that you've built up.

D: Yeah, that's it.

AS: Um, okey doke. And next, Scott?

S: Uh, yeah, that final point about the financial side certainly stuck in my memory. It seems to me that there's there's a, potentially a strong element of, um, an ability to really concentrate in this, in whatever this task is, in that you are going to be, in theory, uh, milling around with a number of different people all doing a different things potentially, and therefore concentration seems to be a key aspect to me. Um... uh, that's about it really, yes. Yeah, I kind of recall you talking about an open-plan office and, and moving around and how one interacts in those circumstances.

12.45 AS: Mm-hmm? Brilliant, ok. And, uh, last but not least, Jenna?

J: Em, yeah, I don't know, just, is it just to see how far people will go for money? I don't know. The money thing kind of stuck in my head and, I don't know, cos they've got a safe place to go in the cones, is it like, will they keep going for the money or will they go in a safe place? I don't know.

13.10 AS: So, the cones represent the finish area where you go into if you want to leave the task at any point, or if you've finished. So once you go into the cones, you can't come back out again.

J: Yeah.

13.22 AS: Um, ok, brilliant. Thank you very much. Um, I'm going to go, ah, back into character now. Uh, actually, it's going to come in and out. So let's start the experiment. The task will be for each participant to pit a sticker on correct mathematical equations. Now if everyone could open document 3 please. And what you'll see is quiet a few maths equations, and some of these will be right, and some of these will be wrong. And so the task is for you to put one of your individually coloured stickers on each correct maths equation. And if you put it on a correct maths equation, you get £1 in the bank. And, like I say, you get up to £40 for the experiment, so you go through each particular sheet, and you try to put as many correct stickers on, and then you move to the nest sheet. Uh, there will be 6 boards with 10 people or so, so you can imagine it's going to be not exactly tripping over each other, but there'll be people quite close to you, but more than enough boards and more than enough equations. Um... Is everyone happy with what the task is?

[General noises of assent]

L: Um, more than enough boards and more than enough equations for what?

14.55 AS: For each of you to get potentially up to £40.

L: Ok, for each of us to get the maximum amount? Ok, cool.

15.04 AS: Um, brilliant, brilliant, ok. So, now, what'd do is, we'd spend about a minute or so getting into this, where you'd put your stickers onto each, uh, uh, correct equation and you'd slowly build up the amount of money that you have in the bank. And every time you finish a sheet you would move onto the next one, or move around someone… Uh, but then after about a minute or two, uh, what you see is, you see is, you turn to your right or to your left and you see one participant touching the shoulder of another participant and they're saying something. Now if everybody could please play audio file 1.

[Long pause. Faint overlapping sounds of audio playing in the background.]

16.21 AS: Has everyone heard the, heard audio file 1?

[General assent]

AS: Brilliant, ok. So, you see um, one participant touching the shoulder of another participant, saying those words. They then start moving towards another of the participants. And my question is, what do you think is happening? What do you think you would do in this case? And does anything important come to mind? And that is to Scott first.

S: Um… Well, can, sorry, can I just clarify whether you said it was a participant talking to another participant?

17.00 AS: Yes, it is a participant talking to another participant, yep.

S: Ok. Um… what is happening? Uh… I, I, would I be correct in thinking perhaps that the participant doing the talking, um, was, uh, in a sense a plant, um, not a true participant? Um, perhaps put there as an observer, and has observed something amiss with the original participant? Um…

17.34 AS: Ok, I should say, it's not a test, there are literally no wrong answers – just say what you think, that's all. Anything more come to mind? [Pause] Is there anything else, or…?

S: Oh, sorry, um… No, not really, that was my thought, really, that that's potentially what it was. If they were, yeah, yeah.

17.54 AS: Ok. And what do you think you would do in this case?

S: As a participant myself?

AS: Yep.

S: Um, I would probably wonder what the heck's going on, but probably just carry on doing whatever my task was, or whichever board I was on, and just carry on.

18.12 AS: Ok, thank you very much. Eh, Jenna, same question.

J: Em… I don't, I don't really know, I just, I would stop, because I've been told to.

18.26 AS: You haven't been told to stop, sorry, the participant who said…

J: Oh, they have?

AS: Yep, they've been told to stop, yep.

J: Then I would expect them to stop.

18.35 AS: Ok, and what do you think is happening in the experiment right now?

J: Someone's just told them to stop! I wouldn't, I wouldn't overthink it, I would just… think it was part of the game.

18.50 AS: Ok, ok. And what do you think you would do in this case? Would anything change in your behaviour?

J: No, I'd maybe try to think what they'd done, cos they must have done something wrong, so I'd maybe look to see, so that I didn't do the same thing?

19.07 AS: Ok? Ok, brilliant, thanks. Uh, Derek?

D: Yeah, I think it, it sounds like a plant, as the other people have said. I think that's designed to distract the participants from their tasks to test how, uh, how positive they are at continuing to do the task. So the, the aim is that the participants would be expected to continue. Anybody that didn't continue or that was distracted would be penalised.

19.42 AS: Ok, ok, thanks. And Liz?

L: Sorry, me?

AS: Yep.

L: Oh, um, I… It, it didn't occur to me until I heard someone mention that it could be a plant, it would just, it would, um, I would probably think, 'Gosh, what on earth are you doing, um, bossing people around when you're a participant yourself, and you don't have the authority to do that?' Um, um, I'd just probably find it a little irritating and then, um, I would [laughs] probably sort of say to another participant, 'Oh gosh, did he just tell him to do that, why is he doing that when he's a participant?' and um, sort of… um, and… um, think he was just being really bossy and taking things into his own hands.

20.33 AS: Ok, ok. Um, and does everyone, just in terms of how we've interpreted the audio file, does everyone think that the person who was, uh, touched on the shoulder and spoken to, do you think that they've lost their money?

J: Yeah.

D: Not at that stage, I don't think, because the instructions were that you work to the instructions of the leader, not somebody that's in the team. So I would carry on and, uh, on that sort of basis, because that hadn't been, that information hadn't been given earlier.

21.17 AS: Ok.

L: I wouldn't think they'd lost the money either.

AS: Ok.

L: Because...

S: And I would – oh, sorry – I was just going to say, I would add that, um, they probably haven't lost their money, but are they a true participant themselves, as well? Or is there two of them sort of acting together as a, as I think Derek sort of suggested, as some sort of distraction?

21.41 AS: Right. Ok, ok, thanks. Um, what I'm going to do is I'm going to make an intervention here, uh, and inform you that the experiment is meant to make you believe that that person has lost their money. So the instruction that that participant gave, that the audio file participant gave to the other one is meant to be believed. Now, I'm going to have to make some changes on that front, but, uh, if we could move forward and we could believe that, uh, there is a potential loss of money. Um... ok, so the next question is, uh, the, the participant who said the words is now, uh, walking towards you. What would you think or feel or do? And, uh, we're going to go for the same order as well. Scott first?

S: I would, uh, be a bit concerned that he was potentially going to say the same thing to me as he had just said to the other person, um... As he's approaching me I'd probably be quickly checking to see whether what I'd done so far looked correct or not, um... and, uh, yeah, just be a bit concerned as to why he was, eh, approaching me and whether he was likely therefore to be saying the same thing to me.

23.03 AS: Ok. And, Jenna?

J: I would just try and dodge them, like, literally stay away in case it was a, the first person he'd seen, and hopefully he'd go with someone else. Just to try and stay in the game, just move.

23.21 AS: Ok, ok, brilliant. Um, and Liz.

L: Uh yess, I would sort of look and see was he definitely coming over to me and think, 'Oh no, I'm going to lose the money, um myself, now', and um...

perhaps I'd look up to see if anyone else is doing anything wrong so you can kind of make him go to them instead of to you. Just go, 'Oh, look, he's doing something wrong', so that maybe the other guy, the chap who, um, uh, told it to the one, the first one, that he's going to be [INAUDIBLE 23.59] will go to someone else and not to me. So shift it, try and shift it onto someone else. Maybe! I don't know if that sounds quite bad.

24.01 AS: Sure, sure, sure. No, no, that, again, no wrong answers, so thank you very much. Um, ok. And I'm going to go back into character. Uh, this, uh, movement of the participant continues until either everyone has been touched and stands still, or everyone has moved into the finish areas. And at that point, I intervene, ah, I come into the experiment and I say 'hank you very much, the experiment is over. Can everyone move over to the side please? I would like to start by saying that there will be no financial penalties on anyone who was touched during this experiment. I'll now take you all through a debrief, after which we will take our final saliva sample. First of all, our experiment was looking at movement behaviour, but during stressful scenarios, for example during uh, a terrorist attack. Specifically, we are looking at how people move when they're confronted with a bladed attacker. Here, our hostile member represented a person with a knife, and the fact that they had to physically touch you, uh, corresponded with the knife element of that. We built up a cognitive test and financial incentives to make you think that you had something to lose if you were taken out of the experiment. And then an actor was used to imitate an aggressor. However, as I said, all of you will be paid the full amount and there will be no penalties taken. The saliva samples are used to test for cortisol levels, which is strongly correlated with a 'fight or flight' response. We measured your position through this experiment with the tags on your t-shirts; this will provide us with quantitative measurements of how people move when they are in extreme emergencies, and the corresponding levels of fear. We would ask you to complete this handout, eh, so if everyone could open documents 4 and 5. And in document 4 you should see some further questions, and in document 5, eh, it's, eh, a redo of the emotional questionnaire that you did at the very beginning of the experiment. So in these questions you are asked about – in document 4, that is – you are asked about how you feel about the group, how you feel about the hostile member, um, and also, eh, some further questions about if you wanted to react differently, so if you felt like you were, eh, not able to react in a way you wanted. Ok. Ah, and

I'm going to go back into character now. Unless there are any questions at this point? No? No? Ok. Um. Ok, thanks. Back into character. Now we are going to hand out some contact details for you if you need to contact any health services as a result of this experiment. If you experience any lasting effects, please either get in contact with your GP, any service on this sheet, or email myself and I'll get you in contact with the correct people. Finally, as I'm sure you can appreciate, this experiment requires that the participants be unaware of the true nature of what is being measured. Please could you refrain from talking about it for two days, as this is our data-gathering period. This is especially important if you know anyone doing the experiment. After the two-day period, feel free to talk about it as much as you like. Ok, so, this is the end of the experimental walk-through, and I have a few further question. Um, the first question is, what did you think during this experimental, uh, walk-through? What emotions did you feel as it progressed? What did you think was going on? And that goes to Liz first, please.

27.58 L: Um, so right at the beginning, I said that I didn't [INAUDIBLE], I was probably a little confused and I vaguely could see what you were going – the probably main emotion I felt was, um, really was irritation with that man who went up and, a participant who told another participant that he, he was finished, you know, you just think, 'Gosh', you know, 'what right do you have to do that, it's not, it's not your, your business to do that.' I just, and then, I can see it's a knife so gosh, it's quite clever. And then I was thinking, 'Oh, I sound quite bad', because I said that I would try and um, not let him come to me but go to someone else, which looks really bad, so it looks like I've tried to get him to knife someone else.

28.48 AS: Um, just again, there are no correct or incorrect answers. What we're trying to do is gather data and, um, further to that, you definitely didn't know that it was a knife, so...

L: I know, but the thing is, that's the scenario, it actually looks... I can see what I've done, I've made, put it off or tried to get him to go, to think – God, it's absolutely awful, but – 'Don't come to me, go to someone else' – that's what I'd go: 'God, I sound really bad doing that.' That's what I'm just thinking – um, you get what I mean?

29.17 AS: Sure, sure sure sure.

L: Yeah yeah yeah. Um, but maybe that's just a sort of, um, a, a, a reaction. But if my children were there, I wouldn't have done that, I would have taken it for me and not them, probably. Like, oh, make sure they're all right and not, don't let other people... I don't know. I just – when you said about the knife, I thought, 'Oh god, it looks like I've, um, pushed the knife onto someone else.'

29.41 AS: Um-hum. Ok, thank you very much. And the same question to Scott?

S: Ah, yeah, um... I suppose at the beginning I was sort of a combination of intrigued and confused as to what I thought it was going to be all about. Uh, I... when I saw the mathematical equations, uh, I think I felt quite relaxed, because they looked reasonably straight forward. I didn't have a chance to read every single one or look at every single one, but they looked reasonably easy to, to work out whether they were correct or not, so I guess I was in a relaxed frame of mind at that stage, um, but, I guess, quickly was less relaxed and more concerned when the um, the scenario of the chap putting his, or the person putting his hand on his shoulder and saying what he said on the audio tape. Um... and, uh... yeah, and then just concerned about whether, the fact that he appeared to be, or he or she appeared to be heading towards me, uh, raised concerns in my mind again as to what was going on and whether I was likely to – I think I said earlier, um, I was worrying about whether I'd done anything incorrectly that might prompt him to say the same thing to me.

31.06 AS: Ok, thank you very much. Yup. Um... and the same question to Derek.

D: Yeah, same sort of thing. Little bit confused as to where we were going with the, um, the talk-through. The only other point I'd make is that, if it's supposed to represent, uh, a serious knife attack, the destructiveness of the plant didn't seem to me to be the sort of disruption that would create the sort of effect that you might want.

J: That's exactly the same as I feel, by the way.

31.44 AS: Um-hmm. Ok, sorry – and thank you, and Jenna, yourself now, is that, um, is it...

J: Yeah, like, 40 quid is just 40 quid, like, that would not be getting any fear in my saliva the same as like somebody with a knife or something, like. I just thought, seems a bit... And I'm not really that competitive, so I really wouldn't care if someone won more money than me, or... So if someone was walking towards me, well, I'm not really all that bothered, like, I don't, I didn't really get it. It certainly wouldn't drive in me the reactions of like, a terrorist. And that bit at the end, about if you've got any lasting effects, make sure you see a doctor – seriously, what if the thought of losing 40 quid, like – bit over the top?

32.32 AS: Brilliant, brilliant, thank you very much. Um, ok, so – is there any, any other further comments? If not, I'll move on to the next stage. Um, any further comments? [Pause.] No? Ok. So, this is an experiment, as I said, that will measure movement direction and speed in an emergency such as a terrorist attack. Um, the issue being at the moment we don't have any of this data. Um, we have some form of indication as to what people might or might not do, but there's no quantitative data. The aim of getting this data is to improve how we talk to the police, how we, how we plan emergency service management, um, or to improve the architecture of buildings so that they can be designed to accommodate how people move during these emergencies. And the question is, knowing what you know now, and, and why, uh, what was being measured and why, would you sign up to the experiment in the first place? And that question goes to Jenna first.

J: Em, yeah, I'd sign up to it, but I don't think I'd give you true readers of what I'd probably actually do in that scenario. If it was for a lot more than 40 quid, then there'd be a bit more at stake and it probably would get me a bit pumping, but that, I don't think you'd get a true reading from me.

34.06 AS: Ok, thank you very much, absolutely. Eh, next to Scott.

S: Em... yeah, I think I'd probably sign up for it, yes. Uh... yeah. I don't know if you want more than that, but yes.

34.23 AS: Sure! No no no, that's – if it's, if it's, uh, an easy answer then absolutely, that, that's perfect, thank you. Uh, Liz?

L: Do you want me to ring in on a different line or shall I just, just keep going? Can you hear that noise, or is it just me that can hear it?

34.37 AS: Uh, we, I can't hear it.

L: Oh good, it's a background, it's a beeping, it's just – um, would I sign up? Yes, um, if what you mean – if I knew it was a knife experiment, would I be more likely to, to take part in it?

34.47 AS: Not more likely, it's just that now you know what the experiment was about, would you sign up to it again, knowing you have to be naïve about the, the purpose of the experiment?

L: I... I would, I probably... Yes, I would, I would, because I think it's, um, it's, um it's, it's a very important area of, um, of saving lives, and if you can, you know, we are in a time of various attacks on anyone's life, can be fatal, someone not hurt so much, um, as a result of your work, I think it's really, really worthwhile, so I would definitely sign up for it.

35.27 AS: Brilliant, thank you very much. And Derek?

D: Yeah, I'd sign up for it, but I don't think it would give you the results that you wanted, that's the only problem. I think someone should have come in with a starting pistol or something and really give everybody a shock, so you could test their reactions to that sort of a situation, but I'd sign up for it. Presumably you'd modify the thing based on feedback.

35.55 AS: Okey doke. Thank you very much. Um, right, ok, so, final question. Um, and I'll, I'll, I'll have a bit of interaction on this one. Is there anything you would do differently in this experiment? We've heard quite a lot that we wouldn't get the right type of reactions; we've heard that, um, if the participant was saying this, you might ignore them because it's a participant saying it, not an instructor, so they don't really have the, the right to be giving instructions. Um, so that, uh, again I had to intervene and say, this is what's meant to be, uh, meant to be interpreted. If I can suggest, eh, if there was a rigged raffle at the very beginning, and the actor in this case was told very publicly, 'You're in charge, eh, and you can give people instructions', would that potentially, eh, eh, make you believe them when they say what they are going to say?

J: Yeah. Sorry. [Laughs]

36.59 AS: Yeah, no no, perfect, brilliant – if, uh, so we've got, uh, was that Jenna?

J: Yes, yeah, sorry.

AS: Um, and, and, so, Jenna, also, is there anything else you would do in this experiment to make it better at, uh, mimicking reactions in a terrorist attack?

J: Um… I don't know… Is it… It needs to be more than 40 quid. I don't mean you need to pay those people afterwards, because once you realise what it is, they'll probably be like, aw, I'm happy to do it for free, but if they thought there was a lot of money at stake, it might get that kind of reaction… But for just, like, adding up to £40, I just don't think people would be… that bothered about dodging that person.

37.48 AS: What if the actor was – because he will be a professional actor, he probably could come across a lot more scary – would, would that be ok with you, or no?

J: Yeah, like, I think they should run round with like a knife or something, really scare people, get true results, but like, obviously they wouldn't use it. But, yeah, just give people a fright. There's no harm in it.

38.12 AS: Uh, uh, ok – brilliant, thank you very much. And same question to Liz?

L: Um, the only thing I can think of is perhaps maybe being… uh, playing the game with the numbers in groups? Um, so that you can protect, um save your, like, with a friend or family or something in a terrorist attack, do you, are you going to protect them as well, or just, is it just, are you just out for yourself, or have you got someone else or someone in your team that you're also, you know, um, playing with, and or protecting the terrorist, um… a scenario.

38.53 AS: Mm-hmm. So, and, so, yeah, we have absolutely considered groups – this is meant to be a sort of 'morning commute' type incident, so you don't necessarily know the people around you. But…

L: Oh right, ok.

AS: No, but I do take your point, and it's a very good point, thank you. Um, um, but in terms of anything else, is there any other element of the experiment that you think, uh, would be an improvement? Uh, and an example here would be if there was a publicly rigged raffle, or if there was a rigged raffle at the beginning, telling the actor publicly that he's in charge and that he can give people instructions, would that be an improvement?

L: If the actor … so we all knew he was in charge?

AS: Yes, if you were to believe that he was in charge, would that make things slightly more believable?

L: Yes, and it would also probably change how I would feel, because I wouldn't be so angry and think, 'What on earth are you doing, taking matters into your own hands?', if I. um, knew he had the power to do that.

39.54 AS: Right, so would you, would you have less of a fear response, then?

L: Oh gosh, because I'm doing this on my phone, it's beeping as I'm getting low on battery! Oh, sorry, if you could just quickly repeat that – would I have a less…?

40.04 AS: Of a fear response – would you be less scared?

L: If I knew he… Yes, I would be less scared, because I, yes, because it's, it's not, it's not the unknown, I, I'd think, 'Oh, he's got the power so he's using it' rather than 'What on earth are you doing?', um, doing that when he's got no right to do it. Yes, it would make me think, 'Oh gosh, um, he's got the power so he's obviously feeling very bossy' rather than…

40.29 AS: Ok. Uh, and would you be ok with the actor being more scary?

L: Yeah. I would be ok with the actor being more scary.

AS: Ok, thank you very much. Uh, Scott.

S: Uh, yeah, I think, on those last couple of points, I think it would be much less effective if the participants did know that, that there was a particular individual in charge. Um… and I think it would be… eh… yeah, much less effective. And also, if, if the actor was more scary, then certainly, um, it would probably come across as more, or I would probably feel more unnerved bit it, if he were to, um… act or speak in a perhaps more aggressive way, um, without going any further than just putting a hand on the shoulder, that is. Um, I'm intrigued as to whether this whole scenario would be filmed, and if so, when the participants would be told it was being filmed, and whether or not the cameras would be hidden or not, if it was being filmed. Presumably it would need to be filmed in order for you to really analyse people's reactions and, uh, how they, uh, how they responded to the actor.

41.48 AS: It will be filmed, um, we can either inform [?] them or not; the cameras won't be hidden. What are you thinking in that respect?

S: I was thinking if, uh, if you know it's being filmed, are your reactions likely to be as true as they would be if would be if you just thought you were in a room with sort of fellow participants and were, uh, not necessarily being observed on film?

J: I agree with that.

L: Yes.

42.20 AS: Ok, thank you very much. Um… and Derek.

D: Yeah, similar sort of things to the, the other participants, but it's still sticking in my mind that tapping somebody on the shoulder and saying 'You're, you're out' when the, the only thing at stake is £40 won't stimulate the right sort of response, I wouldn't have thought. It's got to be somebody that, something that happens suddenly without warning and probably with a lot of noise and something that would create a, a, a, the right reaction from the participants.

43.02 AS: Ok, um…

J: And maybe can I just say, see if you got people like beforehand in a vulnerable kind of place, like, I dunno, maybe spend five minutes talking about your family, you know, so they're in that kind of vulnerable place, and they're not in a, I don't know, maybe that would make a difference.

43.22 AS: Ok. Absolutely, thank you very much for that. Um, and uh, uh, Derek, how, in terms of the actor being more scary, um, you, you also mentioned earlier about bringing in a starting gun or something. How, uh, aggressive could the actor go and you still be ok with, uh, signing up for the experiment?

D: Um, as aggressive as they want to be! [Laughs] Apart from physical violence, but, uh, it's more about... when people get frightened is when they don't know what's going to happen, and if there's a lot of noise and a lot of things happening that they didn't expect, and things happening suddenly. So, I, I, I... it's really a matter of under them circumstances being able to keep calm, and analyse what's happening.

44.18 AS: Mm-hmm, ok. Thank you very much. And, um, just so, on the experiment itself, what – one of the things we're looking to measure is, uh, how quickly people can realise what's happening and, uh, identify a location of safety and then move towards that location of safety. So in this case, the finish area surrounded by cones was the, the supposed area of safety where the participant who is, well, the hostile participant would not move towards. And now, did that come across at all? Because we didn't mention it very much at all.

D: I don't think in a... If you're trying to, um, generate a fright response, standing in against a cone [laughter] wouldn't really replicate that situation of safety! I think it would have to be something more physical than that.

45.15 AS: Ok, so behind a table, maybe, or behind another board?

D: Or, or in a side room or something like that, yeah.

AS: Ok, ok, thank you. Does everyone else feel the same way?

J: Yeah, like, I don't really think anybody's going to go to a 'safe place' because someone might tap them on the shoulder and say they're out. I don't think anyone would be that threatened by that. Put them in a room and then, like, put the fire alarm on or something, and then just say, 'Oh god, this isn't a drill!' and then see what people do.

D: Yeah, that's a good idea.

S: I guess if, if, if the actor is particularly aggressive, um, there's maybe a, a natural inclination for people to gather together, um, in a, in a group situation if you like, it does, you know, sticking together, got a better chance of surviving whatever the scenario is, than, than they have individually, so potentially I might be looking to see where the hell everybody else is going and just head for the same place.

46.18 AS: Mmm, ok. Absolutely.

L: In a knife scenario, I think that's the case, but if you've got [INAUDIBLE] there's going to be a mass shooting, obviously then would be able to take out a lot more people, [INAUDIBLE].

D: Mmm, true.

L: But actually, on another note, I actually don't think it's all right for you, um, I don't think people would, I think some people would not be happy for it to do as, as scary as possible in an experiment. I think you probably might get a lot of complaints. You know, if you really terrify people, I think, I don't think… I don't know, I think some people would, would not find that ok.

46.53 AS: No, sure. What do you think is the, the limit that the actor can go to?

L: For being, for being really scared? Um, for an experiment like that, I think it depends very much on different people [INAUDBLE] that you're trying to get, and… ah… I think a real, a real shock probably isn't ok. Um, you just never know, I mean… someone could be [INAUDIBLE 47.21] and in real shock, or, I don't know, a bit more [INAUDIBLE 47.27] or something like that, and, um… I don't know, it, it, a real shock can do a lot of damage to some people.

47.35 AS: Mm-hmm, ok.

L: I mean, that's what I was thinking. I mean, some people would be happy with it, I'm sure, but I think some people wouldn't – but just wanted to make that, that's what I feel. You know, it may be, may be really scary.

47.46 AS: Mm-hmm. Sure, ok. Thank you very much. Um, right. So that's the, that's the end of what I wanted to do, uh, today. Thank you very much for, um, for, for, uh, being here and, and helping me with this. Is there anything that anybody wants to say, anything that they, uh, feel that they didn't get a chance to say in the last, uh, fifty minutes or so?

D: It's Derek. I think, the only thing on the last participant, about the, uh, people might be damaged by the stress, if you use something like the, uh, the uh, the criteria required, the one, the example we've been supplied, it clearly says that you're excluded if you've got anxiety, depression, PTSD and all the rest of it, and high blood pressure, coronary disease, so you're trying to cover the, the bases as far as anybody that might be vulnerable to sudden shock.

L: I thought that was a joke! I thought that was to kind of get us all into the, into the swing of things.

48.59 AS: No, so that would be a recruitment page that we would actually send out and, is being used to filter people out.

L: Oh, I see! Oh, ok, no, I thought, 'Oh, gosh, it's just all part of it', you know. 'Oh, wow, this is going…', something. I didn't really, um, ok. Well, yes, I suppose you can say – yeah, but I mean… Yeah, I think some people have, you can't sort of say that everyone who reads that is going to be fine, for some people with an underlying, I don't know, disease or… A lot of people with depression or mental health issues don't admit they have it or whatever, so I, I don't think it rules out everyone who wouldn't, um, who would not be affected by severe stress, by [INAUDIBLE 49.40] going 'Oh, I'm fine' when maybe their mental health or whatever rules them out, when actually, you know, I don't know, really, that kind of thing.

49.50 AS: Brilliant, thank you, thank you both for that, absolutely. Anything else?

S: No, not from me.

AS: No? No? Brilliant, right, well –

L: Oh, do we, do we get to see your PhD when it's finished? [Laughter]

AS: Yes! I mean, it'll be in, ah, in, ah, two years or so, but… [laughs]

L: Oh, brilliant! Oh, good luck!

AS Thank you very much.

L: Good luck! My, my, yes, oh, good luck with it, it's tough. But I mean, it's fantastic, and your work, it's really, I know, it's really, really hard work, you're really brilliant doing it, and, no, good luck to you.

50.24 AS: Thank you very much. Uh, brilliant – right, well, I'm going to stay on the call and, uh, Amy, if I could – Amy, are you still there?

Amy: Yeah, I'm still here, yeah.

AS: Brilliant, if I could talk to you afterwards for a sec? But thank you very much everyone else, um, and, uh, go back to your evenings!

[Laughter, general thanks]

L: Are we supposed to do the questionnaires now?

AS: No, no, so they're not meant to be, um, completed at all. You're absolutely fine.

L: We don't have to do anything to do with it? We just leave it, we just leave it now?

AS: Yep, you're absolutely fine.

L: Cool.

AS: Thank you.

[General goodbyes]

B.2.3. Interview 3

AS: Malcolm, can you introduce yourself?

M: Hi, I'm Malcolm. Do you want more than that? [Laughs]

0.08 AS: No no no, just a name is fine. Eh, Georgie?

G: I just say Georgie?

AS: Yes please.

G: Yeah, hi, I'm Georgie.

0.18 AS: Rosie?

R: Hello, I'm Rosie.

0.20 AS: And John?

J: I'm John.

0.23 AS: Brilliant. Thank you very much for being part of this, um, and what I'll do is, uh, I'll kick the whole thing off. What I'm going to be doing is, is talking through a script that I've written, so some of it might come across a little bit stilted. Um, I'll, I'll point out when I'm going into character and out of character. Ah, and this script is going to be exactly the same thing as I say to participants, uh, when we do our actual experiment, which will be in December. I'll be asking, uh, seven or so questions throughout, where I'll ask your opinion or your response to something that I've said. Um, and so to begin with, you've all been given several documents, and I'll ask you not to open them, but could you confirm that you have documents one, two, three, four and five, an audio file, and a picture file?

[General assent]

01.29 AS: Brilliant. And also you've all been sent a recruitment pack, which is again something that we would send to participants, um, in order to recruit them into the experiment itself. Has everyone read this?

[General assent]

01.47 AS: So here's the first question: what does everyone think this experiment is about, at the moment? That first goes to Rosie, please.

R: Oh, it's about how people are going to interact with each other when they're in, uh, well, it's a lot of open plans, isn't it? But in… in a confined space as well, I think.

02.12 AS: Ok, brilliant, thank. Uh, Georgie?

G: Uh, yes, it seems as if you're going to be observing people, uh, when they're in an open-plan space, and looking at their body language and looking at how well they're able to complete tasks – understand what they're told to do.

02.26 AS: Ok, thanks. Eh, John?

J: I was, um, I felt it was in how to cope with working in an open environment. Um, I was somewhat a bit, um concerned about the relevance of having to run a hundred metres [laughs] um, but I got the impression that it, you know, could be quite a large environment that we're having to work in. And, um, probably a stressful environment as well.

03.08 AS: Ok, thanks. And Malcolm?

M: Uh, presumably something along the lines of sleep deprivation.

AS: Ok. Uh, brilliant, brilliant – thank you very much. Um, so what I'm going to do is I'm going to get into character now and I'm just going to set the scene. And the aim is for this to be as immersive as possible, so if you could, eh, imagine yourself in the scenario and try to really feel like you're, you're going through what I'm, ah, detailing. So to set the scene, it's late December, about 4 pm, um, there's about ten people around you, and all ten of you are part of this experiment. You're walking into the location of the experiment. If you could open up the picture file, you'll see where this experiment will be hosted, it's going to be in a sports hall. Now, this is a stock picture file, but if you can also imagine, um, there are, uh, some boards with white paper on them that you can't read at the moment, uh, in the middle. So there's six boards, all arranged in a line. Ok, I'm getting into character now. Good morning everyone. Please feel free to put your bags and coats over there, and now that everyone has arrived, we can get started. We're here today to investigate the movement of people in groups, and specifically in relation to cognitive tasks. An example application might be the design of open-plan offices. Before we get started, we'll need to spend a few minutes filling in forms. Please take your time to read them, and they say that we will be taking personal data and using it in the research, but that all this research will be anonymised and stored securely, as well as never being used for anything except this research. Uh, break in character here: could everyone please open document 1 and document 2. So, in document 1, you should see an emotional questionnaire where you indicate your level of emotion on a scale of 1 to 7, and you would apply one number to each one of these boxes. Don't do it now; this is just to get us into the, uh, mindset of what we're doing. In document, we'll be asking some questions about ho you know each other in the group. Has everyone seen those, happy with those?

[General assent]

05.48 AS: Great, thanks. Back into character now. Ok, once that's done, can you please fill out these forms, using these ID numbers. These forms tell us your name and several important pieces of information, including your bank details, so that we can pay you. In addition to the emotional questionnaire form, we will also now need to take saliva samples. At I would now hand out some essentially test tubes for you to provide a saliva sample. Ok. Now we can start the experiment. As I said, we will be looking at how people move while they do cognitive tasks. So we'll need to track your movements and check your saliva samples. During this cognitive task, there will be financial incentives to do well. Please take these stickers as they will help you in the task. And I would hand each one of you individually coloured stickers. You can earn up to £40 in this task, which we expect to take around 25 minutes. But there is no time limit or anything, so don't worry about taking longer if you need. Please keep in mind that if you are taken out of the game, or you fail the task, you will lose all the earnings you have built up, and be paid a minimum of half an hour's minimum wage. You must follow any instructions that you receive during the experiment, and you will be able to read these instructions for the task once you start. But the stickers will be used to keep track of how much you earn, so keep them close. As you can see, the experiment is mainly on the boards in the middle of the area. To the side, you can see there are finish areas. And here I would point out there are some cones surrounding a finish area, where there are also some seats where you can sit down if you want. Ok, back into character. If you have finished or you want to leave the task at any point, just go to these areas and you will have completed the task and banked your money. On a health and safety note, all of us will be moving, so please be careful where you step, and don't run. There are quite a few of us, so please don't push each other or anything; as we said, there is no time limit. If you receive any instructions during this experiment, please follow them, and if I intervene, please do as I ask. Break in character. Ok, so here's the second question: what does everyone think this experiment is about now? What are the key pieces of information that have stood out in the few minutes? What does everyone think this experiment's about? And that, please, goes to John first.

J: Um... I... I feel quite challenged on this. One, I felt it was rather prompt in asking for bank details, um, very early on, and my ID. Um, and I... I'm struggling as to the task would actually be for.

09.05 AS: Ok? Ok, brilliant, thanks. Anything, um, anything else?

J: Um, no, I can't think of anything at the moment.

AS: No? Ok, brilliant, thank you. And Malcolm?

M: Um, yeah, you've set down some ground rules. I'm slightly confused as to what you're trying to glean from this task. Obviously, you've attempted to incentivise people by suggesting that the way they do things can earn you money, but I find it still relatively vague, as I say, as to what you're actually trying to find out.

09.45 AS: Ok, brilliant. Uh, Georgie?

G: Yes, I'm not sure if I'm any the wiser. I think the list of questions is quite interesting, whether you're angry or scared. I think you're probably setting people up to... be angry, especially with, um, regards to taking away financial benefits, um, so I think it's, I think it's quite a lot of trickery going on here, so, um, you're wondering if people are worried or scared. There's obviously um, some, um, physical signs, so you're taking saliva and looking at, or letting them know that you're looking at some physical signs that might be in their saliva, and emotional responses. Um, so, I think it's just a human behavioural type of question, I think it's probably nothing much to do with open-plan spaces. It's more to do with how people react, people's emotional levels, but I'm not, not sure.

10.42 AS: Ok, thank you. And Rosie?

R: Well, you said at the beginning it's how, how people move while performing the cognitive tasks. But the bit I, I didn't like, that I think if I was really there that would make me uncomfortable, would be the saliva. I guess that's if people were getting more anxious they would produce more. So...

don't know if that really answers your question.

11.06 AS: No, sure. Um, what would you think the experiment is about? So the saliva has definitely stood out, um – what do you think the experiment is about? Or do you not know?

R: Well, well, you said how people move while performing cognitive tasks. So you'll ask people to do something and see how they... how physically they would react to what you're asking, I think. And you're probably going to put people under some kind of stress, I think.

11.36 AS: Ok, brilliant – thank you very much.

J: Alastair, can I ask, is this sort of entirely open, do you want any sort of feedback at any point? Do you want us to interrupt you, or do you just want to go through it and perhaps comment at the end?

11.50 AS: Ah, a good question. I think for points of clarification, feel free to interrupt me, um, but any comments, um, definitely wait until I've asked a question? Other than if you've not understood something.

J: Yeah, sure.

AS: Thanks. Um, ok, now I'm going to get back into character. Ok, let's start the experiment. The task will be for each participant to put a sticker on correct mathematical equations. Now of everyone could open document 3. You can see in document 3 that there are plenty of mathematical equations. Some of them are correct, some of them are not correct. And the task would be to put a sticker next to a correct mathematical equation. There are six boards, and on each board, there are, uh, approximately ten sheets. So there are plenty of pieces of paper in front of you. So you can spread around the six boards – again, there are ten of you. Um, and each time you would put a sticker on the board, you would put £1 in the bank. So if you put twenty correct stickers on correct equations, you would have £20 in the bank. Each time that a, a, a piece of paper is filled up with stickers or there are no more correct equations, you would move to the next piece of paper – or you would move around each other to do this?

J: So is this done competitively?

13.33 AS: Uh, there are more than enough sheets, um, for everyone to have, um, a complete set of correct answers. So it's not really... it's not competitive, it's more the movement around people.

J: Ok, thank you.

G: And it's timed?

13.49 AS: Ah, it's not timed, there is no time limit.

G: Ok.

13.55 AS: Um, so, you would spend about two minutes, uh, or so going through this, reading the instructions, and starting to put the stickers on the board. And each time you, like I say, each time you finish a sheet, you can move on to the next one. So after about, uh, two, two and a half minutes of this, um, uh, you, you start, um, earning money in the bank. Now, the next thing that you see, is you hear someone walking into the room, and you turn to look at them. This is a person that you haven't seen before. This person walks into the room, walks into the experimental arena, touches a participant on the shoulder, and says something. Could everyone please listen to audio file number one?

[Pause]

15.41 AS: Ok, has everyone heard audio file one?

[General assent]

AS: Brilliant, thanks. So, you see the person speaking, touching a participant on the shoulder, and saying that audio. You then see the speaker move the next person, the adjacent person, the adjacent participant, touching them on the shoulder, and saying the same thing. So the next question is: what do you think is happening now? What would you do in this case? Does anything important come to mind? And first of all, Georgie.

G: Um... It all seems a little bit strange. Em... and I think it's probably a bit of a red herring, I suspect these people already know that they've been taken out. So I think there are some people there who are, um, sort of extras, who have been told that they'll be tapped on the shoulder. So I think I would... just keep my head down and carry on with the task. And try not to be too distracted. Um, but in answer to your question, I don't think I'm much the wiser about what's happening. And I think you're setting people up to be distracted.

16.54 AS: Thanks. Next to Rosie.

R: Not a clue! I think I would be probably be quite anxious and hope that they don't come to me next. I'd be quite disappointed, so. But I wouldn't have a clue why. I'd probably be, like, wondering why, but not really a, a clue.

17.17 AS: Ok. Ok, thank you very much. Uh, John.

J: I would quickly doublecheck the, uh, stickers that I'd put on, that what I'd done at that point was correct, um, because I would be going into, 'Hang on, somebody's going to be checking my answers'. Um, and then, press on and get as many completed as I could.

17.51 AS: Ok. Um... sorry?

J: I'd be quite suspicious, but as I say, just make sure that I'm, uh, what I've done is correct and carry on with the task.

18.07 AS: Ok, thank you very much. And Malcolm?

M: Um, on the assumption that your, other than the, the statement contained in the audio file, um your additional participant isn't allowed to interact with any of the participants, um, I'm a little bit like John, I'd be inclined to, um, resent, because looking at the maths it's relatively straightforward, you wouldn't expect to get any wrong if you were being careful. Um, and like John I would have assumed that the person who's asked me to stop had either made an error or was therefore in some way making a separation between one participant and another by suggesting that somebody was doing something wrong, whereas other people weren't.

19.03 AS: Ok, ok – thank you very much. Now, the next stage is, you see the person that has said this audio file, that has taken another participant out of the, uh, experiment, coming towards you and trying to touch your shoulder. What would you think or feel or do? [John laughs.] So again, same order. To Georgie, please.

G: Oh, I'd be pretty cross actually, because I, I think my maths isn't great but it's good enough for that task. So I think I'd feel well and truly set up there! Also I don't think you mentioned to the participants that there would be somebody coming in, um, so I'd feel a little bit tricked by it all.

19.51 AS: Ok. What do you think you would do?

G: Um... I think I'd, well, I'd have to do what I was... Mmm, I don't know? I think I'd have to do what I was told, but maybe I wouldn't, maybe I'd keep going. Maybe it's a trick question to see if I could be distracted, and I'd think, 'Well, actually, I've got a task to do and I'm going to finish it! Because I said I would'.

20.19 AS: Ok, ok, thanks. Eh, Rosie?

R: Um, well, I would stop, but I'd be really angry, and em, disappointed that I'd been stopped in the task right near the beginning.

AS: So you haven't been stopped yet, but the, the new participant is coming towards you.

R: Well, well, I would just carry on doing what I was doing and hope that they're walking past me to the next person. But I don't think I would... do anything different, I would just... I'd carry on, try not to be too distracted, but be aware that they were coming towards me, and... but I wouldn't, like, say, 'Oh no, don't come to me' or anything, but I would be thinking that.

21.02 AS: Ok. John?

J: Eh, again, if I saw him walking to me, I would quickly scan what I'd done, um, and press on and, you know, fingers crossed they'll walk past me. [Rosie and John both laugh.]

21.27 AS: Ok. Ok, thank you, and Malcolm?

M: Yeah, there's one or two assumptions here, aren't there? There's, what, ten people and six boards, so people are sort of fairly well spread around the six boards, um, do the participants that haven't been told they've done something wrong and they need to stop, been able to hear the conversation from the interloper and the existing players? Are they absolutely aware that this person coming is telling people to stop?

22.03 AS: Everyone's heard. Yeah, everyone has heard everything. Not necessarily seen.

M: So it's quite loud?

AS: It's quite loud, yeah.

M: So the next thing then is after they've, they've been to one, you might perhaps... ignore that. If they've then been to a second, you're sort of attention becomes focused on it, and then, the, the only point I would make is, if you are looking at boards, and, and are fixing stickers, then your eye may not be drawn to the person that's, that's taking people out. They would need to probably be making eye contact with you for you to be aware that they were heading towards you and not somebody else. Um, and then, it's an artificial environment, so whether you would be constrained to just stay where you were and sort of accept your fate as it were, I'm not sure, or whether you would attempt to move away from them, and sort of run round the other side of the boards, I don't know – it's really hard to say.

23.06 AS: Ok, ok, thank you very much. Um... right. So, the next question is, it's quite related to what we've just done: what do you remember from the instructions at the very beginning of this, uh, conversation? Anything to do with the financial side of things, or the physical set-up of the arena? So the key thing is here, what do you remember from the instructions at the very beginning of this, uh, this talk-through? And that goes to Rosie first, please.

R: I think if you get stopped, you… you lose all your money, don't you? And you get half an hour on the minimum wage. I think that's what you said. So if I've been stopped, I imagine that's what would happen to my finances. Em, and… what was the other half of the question?

24.06 AS: The physical layout of the experiment.

R: Uh… Oh, can I go and sit on a chair, maybe, you said, if you've been stopped, can't I? I can't remember. I think I'd probably stand there and look at what the other people who had been tapped have done.

24.20 AS: Ok, ok. Thank you very much. Eh, and same question to Malcolm.

M: Um, from memory I think you said there were six boards arranged down the centre of the hall, um, unclear as to whether they're sort of double-sided, like, A-frames with figures on each side, so effectively you've got twelve surfaces on which to work, um, or whether everybody's on one side of the six boards. Um, yeah, you did mention chairs. I must admit… again I think the difference between being actually in the hall and being involved in the experiment, seeing the seating area, and it's more easy to relate to the instructions when you're physically in the space. Yeah, I noted the constraints on your earnings. If you're sort of pulled out, you stop earning and you get half an hour's minimum wage.

25.22 AS: Ok. John?

J: [coughs] Mine was, one, the, um, getting wrong, I would, you know, have quite a reduction, um, to the minimum wage, um, that there were the six boards and ten sheets that we had to complete, um, and… again it was, eh, the forty pounds for 25 minutes of time.

26.06 AS: Okey-doke.

M: Can I just ask one question, Alastair, while you're going through? If I remember rightly, you said that you issued people with instructions. Did they have them in writing, or did you just give them out verbally?

26.20 AS: Technically, I would be reading them out loud. Would written instructions work better, do you think?

M: I think if people read them themselves, then they're more familiar. Because again, if you're in a strange – this is an experiment you're doing, people are invited into the hall, yes, it's a hall but it's a strange environment for them, um, and so you're, I wouldn't say over-awed but you're slightly out of your comfort zone, and I think if you give people the instructions to read, it'll probably make more sense than if they're just read out to them.

27.00 AS: Ok, thank you very much. And finally Georgie, same question.

G: Yes, so, just in answer to that question, you mentioned about, yes, reading them out, and I would say that you didn't, you don't have anything on the initial sheet about whether you've got good hearing, so you mentioned about physical fitness, um, people may have hearing loss, or they may not have hearing loss but just can't hear in that sort of echo-y acoustic, so I think you should add 'good hearing' and 'able to hear in that sort of acoustic', and I would think that probably you should have written instructions as well. Um, and then people could go back to it. So what I remember is about the financial disincentives of stopping or getting things wrong, um, you mentioned about the boards being in the middle of the room, and you did mention that you have plenty of time, so it wasn't time-restricted, um, and... there was plenty of room for everybody. So it should have made people feel easier about the task, that they didn't have to do it within a certain time frame, um, and, and the fact that there was plenty of space. But you, people did have to walk round each other, and I'm not, it's quite difficult to picture how that would be apart from ten people and six boards. So probably offer a bit more clarity about, yes, whether they were A-boards, people could be either side, so exactly how that would work. Um, the financial disincentives would have probably added an extra layer of stress, I think, um... and... I don't think you mentioned anything about somebody coming in.

28.40 AS: OK. Ok, thank you very much. So I'm going to get back into character now. This, uh, this experiment, this participant that has come in and touched people on the shoulder, he continues to go around touching people on the shoulder until he touches all of them and everyone stays still, or, until people have moved into the finish area. Um, and once this has been completed, I step in and I start talking. Ok, thank you everyone – the experiment is over. Can everyone please move over here, and we'll begin the debrief. I would just like to start by saying that there will be no financial penalties on anyone who was touched during this experiment. I will now take you all through the debrief, after which we will take our final saliva sample. First of all, our experiment was looking at movement behaviour, but during stressful scenarios. Specifically, we are looking at how people move when confronted with a bladed attacker. Here, our hostile member represented a person with a knife. We built up a cognitive test and financial incentives to make you think that you had something to lose if you were taken out of the experiment. And then an actor was used to imitate an aggressor. However, all of you will be paid the full amount as there will be no penalties taken. The saliva samples are used to test for cortisol levels, which are strongly correlated with a 'fight or flight' response. We measured your position in this experiment with the tags in your t-shirts these will provide us with quantitative measurements of how people move when they are in extreme emergencies, and the corresponding levels of fear. We would ask that you complete this handout to let us know what you were thinking in this experiment and answer a couple of questions on your emotional state. So could everyone please open documents 4 and 5? On document 4 you should be able to see, uh, further questions, very similar to what happened before but basically looking at how you feel as a group, how you feel towards the hostile member of the group, and whether or not you wanted to react differently, um, than the way you actually did. Document 5 is a repeat of the emotional questionnaire that you took at the very beginning of the experiment, but again trying to imagine how you felt during the experiment as opposed to after I've done this debrief. And I'm going back into character now. Thanks. Now we're going to hand out some contact details for you if you need to contact any health services as a result of this experiment. If you experience any lasting effects, please either get in contact with your GP or any service on this sheet, or email myself and I'll get you in contact with the correct people. Finally, as I'm sure you can appreciate, this experiment requires that the participants be unaware of the true nature of what is being measured. Please could you refrain from talking about it for two days, as this is our data-gathering period. This is especially important if you know anyone doing the experiment. After the two-day period, please feel free to talk about it as much as you like. Ok. Ok. So that's the end, that would be the end of the experiment, um, in practical terms, but what I'm going to do now if I'm going to ask a couple more questions, um, and the first question is, what did you think during the experiment? What emotions did you feel as it progressed? Uh, and that first goes to Malcolm, please.

M: Um, what, imagining as if we'd done the experiment?

32.32 A: Yep.

M: Um… I'm sort of a relatively equitable person, so I, I, and I'm intrigued by these sorts of things, so I'd have been quite enjoying it. I think at the point at which I was stopped, um, or was under threat of being stopped, I'd, I mean, you've got it on your form, I'd have become pissed off. [Laughter] I'm being, I'm being prevented from, from doing something I've come here and been tasked to do. And yeah, you have to appreciate that that's clearly part of the experiment, but nonetheless I'm happily putting my little stickers on against the, the right answers on the sheet, and I think with ten people and six people and, I forget, twelve sheets, did you say? There's plenty of sheets, so you know, I don't think it's going to be that competitive. So if that's acting as a distraction, um, to your potential, um, person with a knife, um... I'm not sure, I'm, I'm struggling to, um, match the validity of the task with, um, the data you're trying to acquire in terms of an, an assailant with a knife. It seems a bit extreme in the circumstances.

34.02 AS: Ok, thank you very much. And Georgie?

G: Um, yes, I'm not altogether surprised, and there's clearly the task wasn't difficult, the situation was putting people in a difficult situation, you'd given them some pre-reading to do, so they knew that they were going to be watched for their... their, em, body language and how they could cope with understanding things in a certain type of environment. Em, so it makes a lot of sense to me that actually you were recording cortisol levels. I just didn't have, I didn't know that, that, that was what you would be recording, but that can tell you how that sort of fight or flight mode that humans can go into, um, when something unusual happens. But I think clearly you'd go into that knowing, you know, we're already asked whether we have any problems with depression or anxiety and if we did then that wasn't, you weren't relevant to the, we weren't relevant to the task, and high blood pressure and that type of thing. So you want a sort of fairly average, healthy individuals. The running was not sort of relevant, but you just wanted to look at fit, healthy individuals in mind and body. So clearly you're going to be doing something that was going to ruffle people's feathers. So, I think I'd probably feel a bit, perhaps a bit cross that I'd got myself into that situation, but not altogether surprised.

35.29 AS: Ok, ok. Thank you very much. Eh, John?

J: Em, so, having been on the receiving end of a knife attack... [laughs]

AS: Oh no! I'm sorry.

J: ... a few years ago, I, I struggle to see how this experiment... what your objective is with this experiment. Are you trying to, um, ascertain how the person, you know, is acting with the knife or the one being attacked? And as I say, having been the recipient of one, I, I, I can understand how an exercise like this can be, you know, needs to be looked at. But I wouldn't have... the experiment, going... Sorry, the experiment, I don't think would come anywhere near the anxiety that I have experienced in the past. But, em, I would certainly be very suspicious, and it's a trait that isn't normally in me, to be honest, em, but certainly I would be getting quite challenging with the six boards and ten sheets. I'd feel quite competitive in that, whilst I could walk around I'd be trying to stick as many yellow stickers on [laughter] to those sheets as possible, and I would be getting a bit frustrated with somebody, um, you know, like I'd be almost trying to over-stick my stickers on theirs and realising 'Hang on, you've got it wrong but I've got it right'. And I'm a bit of a gambler myself, so I would be really chomping at the bit to get these stickers down. Um, but interesting exercise.

38.02 AS: Okey-doke, thank you. And Rosie?

R: Can I have the original question again? [Laughter]

38.07 AS: Sure, it's what did you think during the experiment and what emotions did you feel as it progressed?

R: Em, so I, I think similar to what Malcolm said at the beginning. I probably would have been enjoying it at the beginning, and as I saw people being tapped on the shoulder and told to stop I would have been getting more... um... anxious, because I wouldn't want to stop, if it were numbers, I quite like numbers, so I'd probably be getting quite into it and want to do as many as I could, and would feel like I was cheated if I was stopped early. But if it was a real person with a knife then I think there would be, there would be fear, but in this it wouldn't be like a fear. I think you'd just be more, a little a bit angry that somebody might be trying to stop me from doing what I'm quite happy doing. And if I did get stopped, I would be pissed off. And disappointed.

39.05 AS: Ok. Ok. Um, thank you very much for your, eh, your feedback, guys. What I'll do now is tell you exactly what we're thinking, exactly how we're trying to measure it, um, and I'm going to ask basically if you can imagine, um, uh, doing it in the first place, and then also whether or not you'd be, uh, whether or not there should be any improvements. So, what we're aiming to do is measure how people move when they see a threat. Now, here, the financial incentives are to get you involved in the experiment, and to make you feel like you have something to lose. And the person coming in touching you, telling you that you've lost money, acts as a potential threat.

Now, um, this, what we're looking to see is, for one example, how quickly people recognise that there's a threat, ah, identify a place of safety, which in this case was the cones area where you could go and, and remove yourself from the experiment, um, and then start moving towards that area, um, and, and, depending on how people move around each other, around boards, which would represent, uh, walls, for instance. Um, this, the aim of this is to provide information, quantitative information, which just doesn't exist at the moment. And what we would do is we would provide that to, say, the police, to let them, eh, understand better what to expect in these kinds of scenarios, or to improve architecture, so that you can design buildings, eh, to accommodate people and how they move during these scenarios. So now that you know what we're aiming to do and why, the question is, would you sign up to this experiment in the first place? So, knowing that you'd have to be naïve about the experiment, would you sign up in the first place? And that question goes to Rosie first.

R: Probably, because I'm… if I didn't know anything about it, because I'm curious and interested in all human behaviour kind of stuff, so, if got the invitation, I would have said yeah, I'm sure I would.

41.31 AS: Ok, thanks. Um, to Georgie?

G: Um…I think… I think probably not, actually. I think there might be other ways of looking at that sort of behaviour without tricking people. Em… and, and, I think people can be very easily swayed by, em, financial incentives, and I know, I understand what you're doing here, um, and so I understand how you can then perhaps pretend that you're taking away those incentives, um, and I know that you're actually not allowed to do that, so I would be a bit suspicious. So I think it's a little… if it's just about looking at behaviours of people in spaces, then it's a fairly drastic way of dealing with that. I think if you were looking at something more urgent, like how people respond in an urgent situation, yes, that's fine, but I think in a way perhaps that, that should be warned. So, on balance, I think I wouldn't sign up for it.

42.30 AS: Ok, ok, thanks. Malcolm?

M: Um… Yeah, I mean, given that one is signing up without knowing what it's about, and it's an opportunity to, I'm guessing you're probably talking about students that provide a regular audience for this sort of thing, and a reservoir of participants, and the opportunity to grab some cash is probably welcome. Um… so yeah, I'd probably sign up for it. As I said earlier, I'm quite up for this sort of thing, it's quite good fun usually, quite interesting. Um... I'm not sure... My only issue, and you may be coming on to this, and I apologise if I'm pre-empting it, um... trying to draw a distinction, or trying to equate somebody who's just telling you to stop putting stickers on a board with somebody appearing with a knife. I would, I would question whether that's a valid comparison. The difference between... and I'm not an endocrinologist, so I don't know whether you would experience an adrenaline rush with somebody coming to tell you to stop. You definitely experience an adrenaline rush with somebody appearing with a knife in their hand that was actually attacking or threatening people. Um, and maybe there, there's perhaps a better way of designing the experiment to introduce, not necessarily somebody with a knife, because that's probably quite terrifying, but perhaps somebody that might be perceived as a more real threat. I don't know.

44.19 AS: Ok, thank you very much. And John?

J: Yes, I, um, think I would sign up to this type of experiment. As I say, it's an area that I feel that's, you know, putting something into society rather than taking, and is certainly something that I feel quite strongly about. The initial doing it and monetary gain doesn't really, you know, have that much interest for me; it's more for what I would get out of it, you know, personally, and I do this type of thing, you know, where it would help, you know, local community or society, um, it's something that I'm in a fortunate position at the moment where I have time on my hands, um, and I'll do whatever I can to, you know, make a difference. The... Knowing a bit more about the type of experiment, I think Malcolm questioned as to whether the putting the stickers on the board would give the impact that you're, you know, or the response, um, that you'd be looking for? You know, would it be very, putting a sticker on and the monitoring that you were doing through the saliva testing etc, I don't know if that would go high enough on the Richter scale [laughs]. It's such a serious issue that you're trying to addresses, and, you know, I certainly wouldn't have considered at any stage of this experiment of going to a coned area, um, being the safe zone of not having my shoulder being touched. But no, I, I would certainly, you know, be for it.

47.10 AS: Ok, thank you very much. Um, right, yes – and yes, you did pre-empt it in your answers – the next question is, what can we do to improve this experiment? And we've touched on the idea that, uh, this is not the same thing as an actual terrorist attack, it's not the same thing as an actual bladed attack, and there are definitely limits to what we can do in a controlled laboratory or an experimental environment while being ethical at the same time. Um, so the question is, what could we do, uh, differently in this experiment to improve it, and you still sign up in the first place? Um...

G: Can I just ask what the objective of it was again? You just revealed the premise of the experiment – can you just repeat that?

AS: Sure, sure. So the objective of the experiment is to obtain quantitative movement data, so, how fast do people move, which direction do they go in, how quickly do they react, information that just doesn't really exist at the moment, um, in terms of, responses to terrorist attacks. And this information, once we've got it, would be used to provide the emergency services with, ah, an improved planning capability, so that they know what to expect, ah, or to improve the design of buildings, to kind of accommodate how people move during these types of incidences. Does that answer your question?

G: It does, yes, thank you.

48.58 AS: Um, okey-doke. So, some things that have come up are, whether or not we could get the actor to be a little bit more scary. Um... now there's, it will be a professional actor, [laughter] so we could go probably to an arbitrary level of scariness, and then there's further things like, do we get the actor to say particular things like, 'I'm an instructor, I can take your money away and I'm now taking your money away', to make it more realistic. Um, or... what kind of things would instigate in you a kind of fearful reaction? So that's, it's kind of a very broad question, but anything at all, any comments at all? And that first goes to Malcom.

M: Um... I think, personally I think that the level of threat needs to be much higher much more real, because the scenario you've described, the experiment itself as it currently stands, seems to be relatively benign, even with someone – given that you know you're in an experiment, and people are, have a level of expectation that something may happen, somebody coming in and saying stop is not very threatening and, as John said, losing twenty quid is not like a big threat, really. It doesn't, it's not going to spark... and frankly I'd be surprised if it would even spark the sort of movement that would be anywhere near would what happen in the event of a knife attack or a terrorist attack.

50.47 AS: Mm-hmm. Ok, thank you very much.

M: Certainly, from point one, your interloper needs to be much, much scarier, much more aggressive. So that in and of himself, he's frightening, which will make people respond to him in a way that him just approaching you and tapping you on the shoulder is unlikely to elicit, I would suggest.

51.15 AS: Ok, thank you very much. Um, same question to Rosie.

R: I, I don't know how you would make him more scary, but I think without – my, when you were going through it, my reaction was, well, I would just carry on what I was doing and hope he doesn't come near me, but I think maybe actually being, like, if you were actually doing the experiment it might be more clear that you could... Well, you'd probably run away, wouldn't you? If he was really threatening and I, I didn't feel when we were talking through it that it was an option to run away from the man who was tapping the shoulders. Whereas if I was really there, if, if I felt like that was an option, I would probably, if he was really scary, then I would forego the money and I would run to the coned bit, the safe bit. But I didn't feel from talking through that that was an option. So maybe when you're there, that would just need to be... maybe that would be clearer. Because that, I don't think you would, you wouldn't really just carry on, like, you know, I'm enjoying doing what I'm doing [laughter] if there was, there was a real, if there was a real threat, you'd run away, wouldn't you? Or I would run away.

52.29 AS: And with respect to the actor being more scary or saying different things, or...

R: Yeah, I mean, the voice, the voice was, the audio wasn't nice, you know, I didn't like the audio when he spoke, just, my thing was like – you know when you were watching Doctor Who as a kid? You think, not nice, don't like it. But I don't know how you would make an actor more scary. Sorry!

52.54 AS: I think they have a couple of ways of doing it, I mean, there would be volume, there would be diction, tone, stuff like that. So if, if we could get the actor to become more scary, would that be acceptable to you as a participant?

R: Well, it's difficult, because if you want to see how people react to it, you can't tell them that they're going to be scared. So I'd probably be quite upset afterwards if I'd done something and I felt threatened and frightened. But you have to make people feel threatened and frightened to get the results that you're trying to get, so, I don't really know how you would...

M: Sorry, sorry, going back to the make-up of the experiment: in a terrorist attack, you've got a period during which people are, um... discombobulated, they're unsure, they don't know what's going on. And so there's a realisation, a process that takes place. Perhaps with the experiment, if you said, if you primed the participants that, right, this is your task, you have to put the stickers on the board, and at some point during you're, um, stickering, somebody will come in and they will endeavour to stop you stickering. And it's your task to try to avoid them. So that then people are primed, in, in real life, if you translate that to real life, they know that there's a terrorist attack, so that they now want to escape. Whereas there's no hint in the experiment, with the relative lack of aggression from your, um, extra person, that, that I would relate to something as extreme and severe as a terrorist attack. So you're not going to get a realistic response from your participants for what would be something far more extreme in real life.

54.54 Ok, ok, thank you very much. Um… we were on Rosie, right

R: [laughs] Yeah.

M: Sorry, I apologise.

AS: No, no, no, thank you very much, it's a really, really good addition – I just wanted to make sure that we're still in the same order. Rosie, would you have anything else to say, and, eh, would you agree with that? Any, any comments at all?

R: Yeah, no, I think, I think, I tink I said all I need to. It's just that he, your person, your distraction, would need to be more threatening. The financial incentive probably wouldn't be enough, you know, you would forego it, wouldn't you, in a, in a really threatening situation, you'd… But it doesn't quite correlate properly, I think. But I'm not sure how to fix it.

55.46 AS: Okey-doke. Thank you very much. Um… Georgie next.

G: Em, yeah, I think I agree with what everyone's said, so rather than repeat it, um… I just wonder whether the experiment is really a good way of finding out how people would react in a terrorist attack and just from a little bit of personal experience, fortunately I haven't been involved in a terrorist attack but I've had to think about it quite a lot because I worked as an usher at quite a well-known opera house last summer and we had a very brief day's training and we watched some films and things, and one of them was about having a terrorist, you know, this is what would happen if there were a terrorist attack, you'd, is it, run something and hide… and it was a very, the, the film was very powerful. Um, I think what would have been much better is if they'd had dedicated an extra day, and I think it should have been, you know, a whole day on here's what you would do in the event of a fire, talk people through it, ok, this is what's going to happen, you know, we did a fire drill, and we had to evacuate the opera house, and we did that quite quickly, and I think we should have done that for a terrorist attack as well, really looked at how that particular building, what you could do, what doors you could lock, you know, in fact, there didn't seem to be any way you could lock somebody out if there was a terrorist attack. And actually, it was quite likely, you know, last summer everyone was on high alert and I think it made us think a lot about what we would do. I don't feel that we were properly assured that we'd had, or reassured, that we'd had enough training. And I think maybe just doing something upfront, and saying, this is, you know, 'We want to look and see what you would do if there's a terrorist attack, so we want to go through a drill and work out how, how this building, you know, how would you work the building.' Um, and, you know, they didn't make it clear whether we were protecting ourselves or whether as ushers our primary task was to protect the, um, the audiences. So I think training and actually talking people through those buildings, so you could understand if you were building a new office block or something, what you, how you could make it fairly watertight, with a little bit of, um, you know, if you knew somebody had invaded the building, what you would do to stop that person spreading to other parts of the building.

58.06 AS: Ok.

G: That's probably not really what you wanted to hear, but I agree with everybody else, I don't think I would have been scared enough by somebody tapping me on the shoulder. It didn't really equate to, to being attacked or to the possibility of... you know, a terrorist attack.

58.27 AS: Ok. Ok. And just to bring that back to the earlier question where you said you wouldn't sign up in the first place – can I, can I clarify that? Because what this question was asking before was essentially presumptive consent. So, can I clarify – when you said that you wouldn't sign up, is that because you don't think the experiment is getting the correct results or because you wouldn't be happy doing the experiment?

G: No, I'd be quite happy doing the experiment. And I'm always open to things like that, you know, hence being on this, this panel now – I think it's really interesting, and I think it's a good thing to do to, um, to test the water with people. But I just... I don't think you're going to get the results that way, and I think it's a hugely important thing. So I think there are other ways of doing it. And it doesn't actually mean I wouldn't sign up for it; I probably would sign up for it, but I would be a bit disappointed that I don't think you'd be getting the results you were setting out to get.

59.38 AS: Ok, ok. Yep, no, absolutely. And finally, to John.

J: Um, I, I think the experiment, em, would need to deal with the type of threat. You know, my issue that I had when I was attacked, I'm a Chelsea fan and I actually went to Millwall [laughter] many years ago.

M: Sorry to laugh.

J: But I think in that type of environment there was, you're more conscious of almost a sort of expectation... so there was almost like a preparedness, you know, in the environment that I was in. But if... you could be attacked where some people are actually in isolation, you know, walking down, you know, a country lane and being attacked, and I think when you're coming in on what the exercise is that you're trying to do is measure people's movement and how they react, is if it's, you know, somebody being knife-attacked down a country lane, I would suspect there'll be, em, quite a period of time where the person being attacked is literally frozen, because of the surprise. But, like, at a football game, I wasn't necessarily expecting, but at that time, you know, physical violence was an expectation almost, and I think you prepared to some extent. And similarly, when I go to, em, London, you know, with everything that's been going on in the last ten years or so with terrorists, I think generally we are more conscious of our surroundings. And so I think the experiment, you would need to cover different types of environment that the experiment is being done in. Um... because, you know, people initially react, you know, differently in their surroundings. It's quite a challenge, and I can understand what you're trying to achieve, and, but, uh, that would be my, you know, personal response to your question. And I quite agree with the, the others that contributed and I agree with them, you know, on quite a lot of points. But I think the surroundings of where an attack could be taking place, um, would make quite a difference. And similarly if it was a knife attack in a lane or in a road somewhere, is trying to see how people would move differently, um, if it was an attack on you personally and nobody else, whereas some of these terrorism attacks are at random, and at the same time, you know, their achievement is, is to probably, you know, injure, um, or if not, you know, maim people or kill them en masse. So that's, you know, it's trying to get an experiment that covers those points, I think.

1.04.11 AS: Ok. Thank you, thank you. Um. Right, so we've definitely covered all the design, um, points of this experiment and gone through them in detail. Um, we're running out of time, but if I could suggest, um, what if the recruitment page covered the, uh, the idea that the experiment is looking at movement in stressful or emergency situations? So you are priming the participants a little bit in saying 'stressful or emergency'. And then you get exactly the same experiment, but when the actor comes in, the actor is, uh, aggressive and, not violent, but loud and, uh, there is no kind of question about what the actor's aims are. Would that cover some of people's comments about the, the experiment? Would that be an acceptable way to go forward?

G: Yeah, I think so.

AS: Sorry, was that Georgie?

G: Sorry, it was, yes.

AS: Um, so, sorry, that would cover some of your questions about whether we're getting the correct data and whether or not it would be applicable?

G: I think it would go some way towards it. I'm still not sure if that's the whole story though, whether you should be setting the scene a little bit more beforehand. I know that you need to do it so that people don't, I guess that's the whole thing about a terrorist attack, it's got to come out of the blue, and you do want to see people, so you don't want to prime them too much, but yeah, I think just a little bit more information would be good.

1.06.03 AS: Ok, thanks. Anyone else?

M: As, as, as I said before, there's obviously a lag between the appearance of an aggressor and, and the understanding on the part of the person that, that, that the person is an aggressor and that they are under threat. So, um... in a real-life situation, there's this time delay. In your experiment at the moment... [cuts out]

1.06.45 AS: Hello?... Hello? [blurred feedback, scrambled speech]

1.07.00 M:... people in the experiment, the perceived level of threat for your people in the experiment, it has to be higher, it has to be... not necessarily generally, generally scary, but they have to be aware that this person is posing a threat and there are consequences to that threat, and... I mean, obviously you can't just throw money at it, but if you said you were going to get a thousand pounds for every number you get right, you know, and people are sort of racking up thousands and then somebody comes and says, 'Stop now', I know that's unrealistic, but there needs to be, for those in the experiment, a, a genuine perceived threat to them, whether it's, obviously with a knife attack it's going to be a physical threat, but whether you can engender some other type of thing where they feel really threatened, I think is the key to this. Because it's obviously [coughs], excuse me, it's obviously the, the level of threat and the movement that that prompts that you're trying to identify.

1.08.15 AS: Right, ok. Thank you. Uh, anyone else?

G: Can I just add something? Sorry, it's Georgie.

AS: Mm-hmm. Yeah, that was Georgie, right?

G: Yes. I just – it was, sorry – I just wanted to add that, um, I'm just wondering whether it would be better to get people to work in teams of three or four, um, and maybe that would help to understand how people react. So you could, you could say that the task is about teamwork, but also about individual, um... you could make it be a teamwork, um, thing but also that individuals, you know, would be watched to see how they react individually. Um, and I wonder whether that would be more useful, because then people would not be quite sure whether they need to look after themselves or to help their teammates. Maybe that might give you some more interesting results. It was just a thought.

01.09.20 AS: Ok, thank you very much. No, it's a very interesting point, absolutely, absolutely thank you. Um, Rosie, did you, ah, do you have anything else?

R: No, I was going to say I don't think I do have anything else, I think everything's, everything's been covered that I might have wanted to add.

AS: Okey-doke, thank you very much. Um... right, ok. John, did you have anything else? I mean, I think you, you, you covered quite a lot in what you thought in the last question, but if you have anything else to say, absolutely.

J: Yeah, the, the only other thought was with, you know, all the CCTv that must be available to people, you know, like yourselves and experiments like this, um, you know, I would have thought, you know, with the trap and trace systems that are available with digital, um...

AS: I mean, in terms of real data, you're absolutely right, eh, but we're focusing on the experiment at the moment. But thank you, yes, absolutely, you're right, yes. Um, but anything to do with the experiment?

J: No, I, I can't think of anything else to add, really. It's trying to come up with the, with the surprise [laughs] um, impact on people is I think the challenge for you – and at the same time I think I can see why, uh, in the explosions you wouldn't necessarily want anyone with coronary diseases. [Laughs]

1.11.13 AS: No! Ok, ok, you're absolutely right, and thank you very much. Um… Brilliant, thank you very much – that brings us to the end, guys. Unless anyone has any further questions, I'll say thank you very much, and, eh, I'm going to stay on the line and have a bit of a chat with Karen, um, and, and let you guys get back to your days. Cheers!

[General thanks and goodbyes]

C. Experimental paperwork

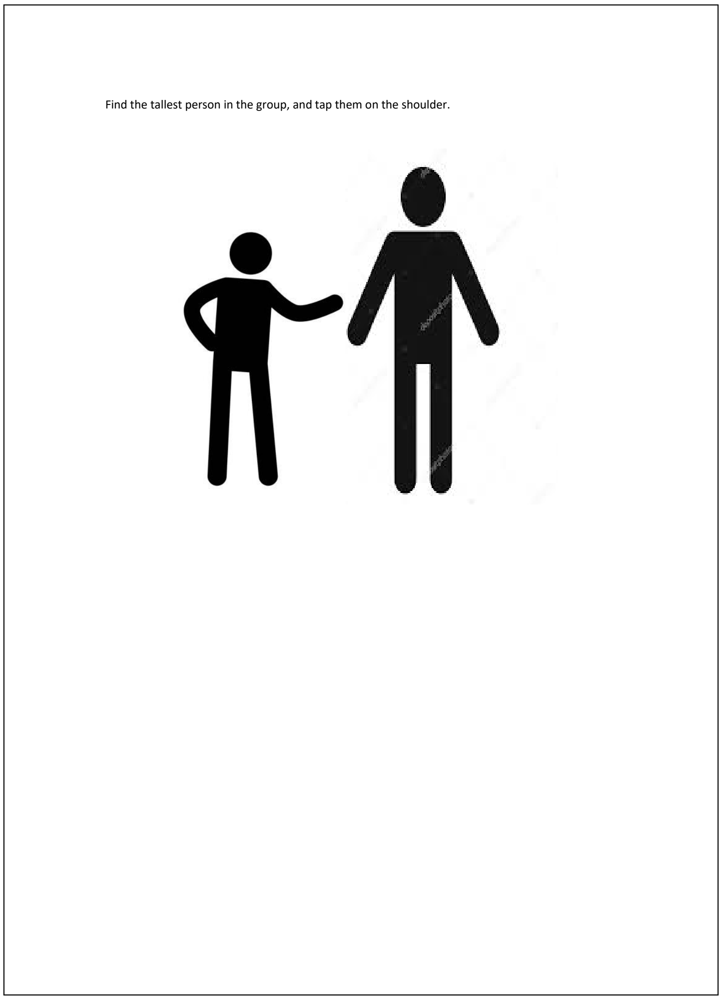

Find the tallest person in the group, and tap them on the shoulder.

Figure C.1.: PR Example distractor tasks

STAI-T A number of statements which people have used to describe themselves are given below. Read each statement and then choose the most appropriate statement to indicate how you generally feel. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your feelings best.

	Almost never (1)	Sometimes (2)	Often (3)	Almost always (4)
I feel calm (1)	❍	❍	❍	❍
I am tense (2)	❍	❍	❍	❍
I feel upset (3)	❍	❍	❍	❍
I am relaxed (4)	❍	❍	❍	❍
I feel content (5)	❍	❍	❍	❍
I am worried (6)	❍	❍	❍	❍

STAI-S A number of statements which people have used to describe themselves are given below. Read each statement and then choose the most appropriate statement to indicate how you felt during the experiment. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your feelings best.

	Not at all (1)	Somewhat (2)	Moderately (3)	Very much (4)
I feel calm (1)	❍	❍	❍	❍
I am tense (2)	❍	❍	❍	❍
I feel upset (3)	❍	❍	❍	❍
I am relaxed (4)	❍	❍	❍	❍
I feel content (5)	❍	❍	❍	❍
I am worried (6)	❍	❍	❍	❍

Version 2.0 13/6/2018

Figure C.2.: PR Questionnaires page 1

PANAS This scale consists of a number of words that describe different feelings and emotions. Read each item and then list the number from the scale below next to each word. Indicate the extent you have felt this way during the experiment.

	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
Interested	❍	❍	❍	❍	❍
Distressed	❍	❍	❍	❍	❍
Excited	❍	❍	❍	❍	❍
Upset	❍	❍	❍	❍	❍
Strong	❍	❍	❍	❍	❍
Guilty	❍	❍	❍	❍	❍
Scared	❍	❍	❍	❍	❍
Hostile	❍	❍	❍	❍	❍
Enthusiastic	❍	❍	❍	❍	❍
Proud	❍	❍	❍	❍	❍
Irritable	❍	❍	❍	❍	❍
Alert	❍	❍	❍	❍	❍
Ashamed	❍	❍	❍	❍	❍
Inspired	❍	❍	❍	❍	❍
Nervous	❍	❍	❍	❍	❍
Determined	❍	❍	❍	❍	❍
Attentive	❍	❍	❍	❍	❍
Jittery	❍	❍	❍	❍	❍
Active	❍	❍	❍	❍	❍
Afraid	❍	❍	❍	❍	❍

Version 2.0 13/6/2018

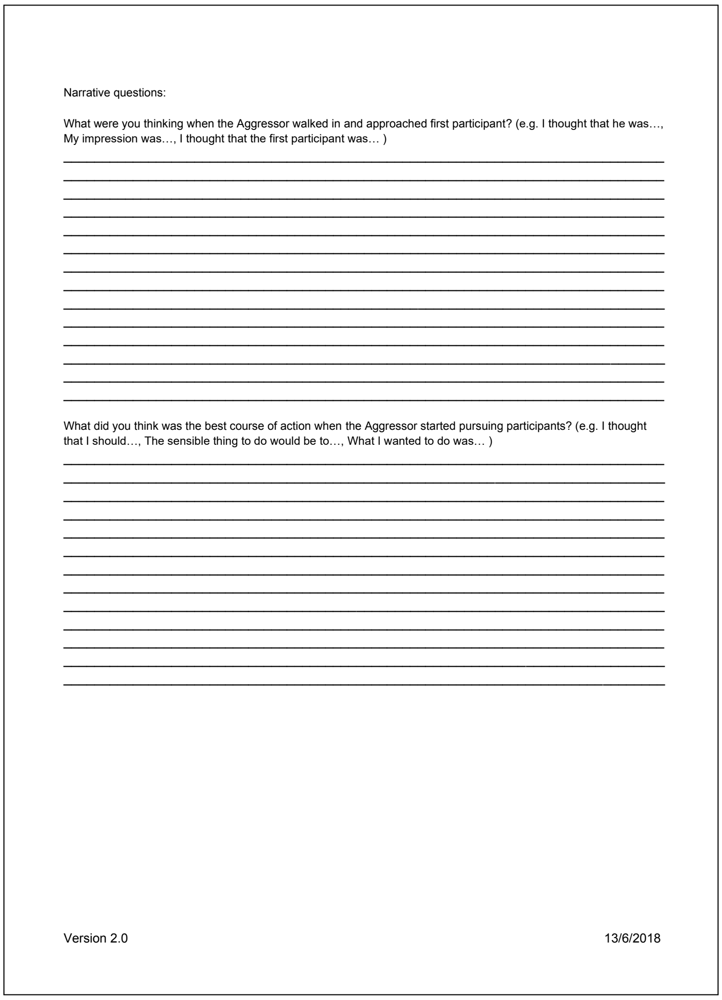
Figure C.3.: PR Questionnaires page 2

Group Identity
This table shows a number of statements that correspond to how you feel about the members of the group. Read each item and then indicate your level of agreement.

	Disagree Strongly (1)	Disagree (2)	Slightly disagree (3)	Neutral (4)	Slightly agree (5)	Agree (6)	Agree strongly (7)
I feel a bond with the group	○	○	○	○	○	○	○
I fell solidarity with the group	○	○	○	○	○	○	○
I feel committed to the group	○	○	○	○	○	○	○
I am glad to be in this group	○	○	○	○	○	○	○
I think that this group have a lot to be proud of.	○	○	○	○	○	○	○
It is pleasant to be in this group	○	○	○	○	○	○	○
Being in this group gives me a good feeling.	○	○	○	○	○	○	○
I often think about the fact that I am in this group.	○	○	○	○	○	○	○
The fact that I am in this group is an important part of my identity.	○	○	○	○	○	○	○
Being in this group is an important part of how I see myself.	○	○	○	○	○	○	○
I have a lot in common with the average member of this group.	○	○	○	○	○	○	○
I am similar to the average member of this group	○	○	○	○	○	○	○
In this group, people have a lot in common with each other.	○	○	○	○	○	○	○
In this group, people are very similar to each other.	○	○	○	○	○	○	○
I shared an identity with the other participants in this study	○	○	○	○	○	○	○
I think of myself and the other participants in this study as part of a single group	○	○	○	○	○	○	○
I think of myself and the other participants in this study as part of one large family.	○	○	○	○	○	○	○
I had a sense of 'we'-ness with the other participants during this study.	○	○	○	○	○	○	○
Besides the differences I have with the other participants in this study, we share the same identity	○	○	○	○	○	○	○
I had a feeling of unity with the other participants in this study	○	○	○	○	○	○	○

Version 2.0 13/6/2018

Figure C.4.: PR Questionnaires page 3



Narrative questions:

What were you thinking when the Aggressor walked in and approached first participant? (e.g. I thought that he was…, My impression was…, I thought that the first participant was…)

What did you think was the best course of action when the Aggressor started pursuing participants? (e.g. I thought that I should…, The sensible thing to do would be to…, What I wanted to do was…)

Version 2.0 13/6/2018

Figure C.5.: PR Questionnaires page 4

Pre experiment questionnaire

STAI-T A number of statements which people have used to describe themselves are given below. Read each statement and then choose the most appropriate statement to indicate how you generally feel. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your feelings best.

	Almost never (1)	Sometimes (2)	Often (3)	Almost always (4)
I feel calm (1)	❍	❍	❍	❍
I am tense (2)	❍	❍	❍	❍
I feel upset (3)	❍	❍	❍	❍
I am relaxed (4)	❍	❍	❍	❍
I feel content (5)	❍	❍	❍	❍
I am worried (6)	❍	❍	❍	❍

VR/gaming related questions

How much have you played video games? (1=never, 7=daily) ………..

How much have you played in virtual reality? (1=never, 7=daily) ………..

Version 2.0 13/6/2018

Figure C.6.: VR Questionnaires page 1

Narrative questions:

What were you thinking when the Aggressor walked in and approached first participant? (e.g. I thought that he was…, My impression was…, I thought that the first participant was…)

What did you think was the best course of action when the Aggressor started pursuing participants? (e.g. I thought that I should…, The sensible thing to do would be to…, What I wanted to do was…)

How long did you think the experiment lasted? Please give two answers: for the non-stressful part, and the stressful part:

Non stressful part: ____________ minutes _______ seconds
Stressful part: ____________ minutes _______ seconds

Do you have any experience that is relevant to this type of experiment?

Version 2.0 13/6/2018

Figure C.7.: VR Questionnaires page 4

VR questions:

For these questions, please answer with a number (1=awful, 7=realistic)

How would you rate the visual component of this experiment? ……….
How would you rate the audio component of this experiment? ……….
How would you rate the realism of the other participants' avatars in this experiment? ……….

For these questions, please answer fully

What do you think the other participants will say about the experiment?

How much do you think you acted like you would in real life? How would you act differently in the real world (if at all)?

What do you think were the emotional responses of the other participants in this experiment?

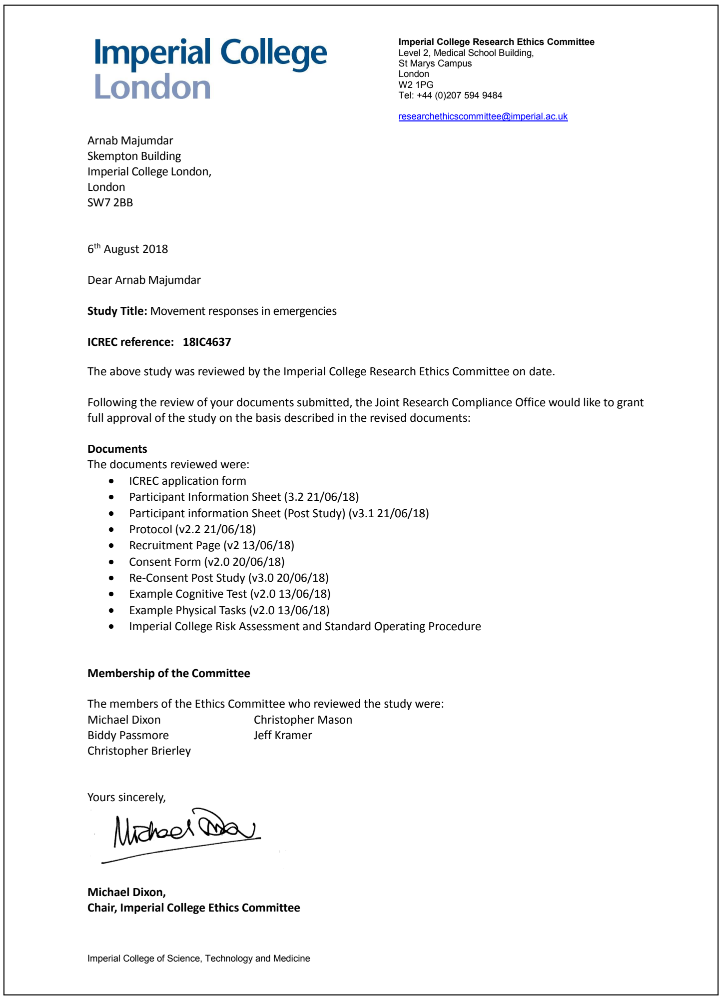
Do you think your responses were affected by the other participants?

How much did you think that the other participants were controlled by people? ……….
(1=not at all, 7=completely)

Version 2.0 13/6/2018

Figure C.8.: VR Questionnaires page 5

D. Ethics Committee Certificates



Imperial College London

Imperial College Research Ethics Committee
Level 2, Medical School Building,
St Marys Campus
London
W2 1PG
Tel: +44 (0)207 594 9484

researchethicscommittee@imperial.ac.uk

Arnab Majumdar
Skempton Building
Imperial College London,
London
SW7 2BB

6th August 2018

Dear Arnab Majumdar

Study Title: Movement responses in emergencies

ICREC reference: 18IC4637

The above study was reviewed by the Imperial College Research Ethics Committee on date.

Following the review of your documents submitted, the Joint Research Compliance Office would like to grant full approval of the study on the basis described in the revised documents:

Documents
The documents reviewed were:

- ICREC application form
- Participant Information Sheet (3.2 21/06/18)
- Participant information Sheet (Post Study) (v3.1 21/06/18)
- Protocol (v2.2 21/06/18)
- Recruitment Page (v2 13/06/18)
- Consent Form (v2.0 20/06/18)
- Re-Consent Post Study (v3.0 20/06/18)
- Example Cognitive Test (v2.0 13/06/18)
- Example Physical Tasks (v2.0 13/06/18)
- Imperial College Risk Assessment and Standard Operating Procedure

Membership of the Committee

The members of the Ethics Committee who reviewed the study were:

Michael Dixon	Christopher Mason
Biddy Passmore	Jeff Kramer
Christopher Brierley	

Yours sincerely,

Michael Dixon,
Chair, Imperial College Ethics Committee

Imperial College of Science, Technology and Medicine

Figure D.1.: PR Experiment ethical approval

Imperial College London

Science Engineering Technology Research Ethics Committee
Level 2, Medical School Building,
St Marys Campus
London
W2 1PG
Tel: +44 (0)207 594 1872

researchethicscommittee@imperial.ac.uk

05 December 2019

Dear Arnab Majumdar

Study Title: Group Movement responses in virtual reality

SETREC reference: 19IC5216 and Notice of Amendment 1

The above study was reviewed by the Science Engineering Technology Research Ethics Committee on 14/05/19 and approved 22/09/19.

Following the review of your amended documents submitted on the 02/12/19, the Joint Research Compliance Office would like to grant full approval of the study on the basis described in the revised documents:

Documents
The documents reviewed were:

- Application form (v1.2 12/06/19)
- Protocol (v1.1 1/05/19)
- Email Invite (v1.0 1/05/19)
- Participant Information Sheet 1 v1.1 1/05/19)
- Consent Form pre (v1.0 1/05/19)
- Participant Information Sheet 2 v1.1 05/12/19)
- Consent Form post (v1.1 05/12/19)
- Debrief (v1.0 1/05/19)
- Questionnaires (v2.0 13/06/18)
- Massey Approval Letter

Notice of amendment Documents

- Notice of Amendment 1 (v1.1 05/12/19)
- Protocol (v1.2 05/12/19)

Membership of the Committee

The members of the Ethics Committee who reviewed the study were:

Chris Hughes	Stephen Webster
Vivienne Kuh	Chris Watkins

Yours sincerely,

Chris Hughes

Chris Hughes,
Chair, Science Engineering Technology Research Ethics Committee

Ruth Nicholson,
Head of Research Governance and Integrity

Imperial College of Science, Technology and Medicine

Figure D.2.: VR Experiment ethical approval

E. External analysis models

E.1. Log models and Individual standardised models

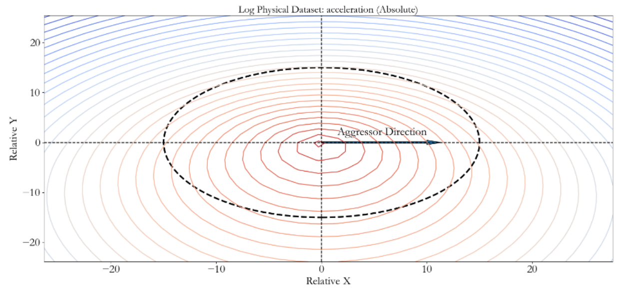
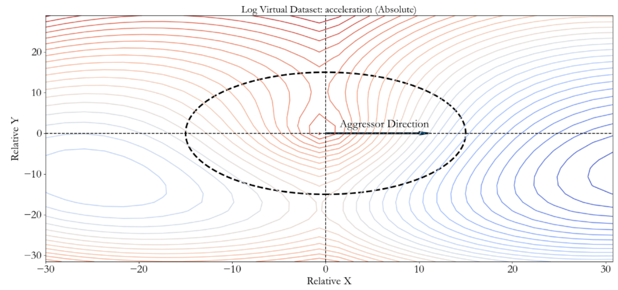
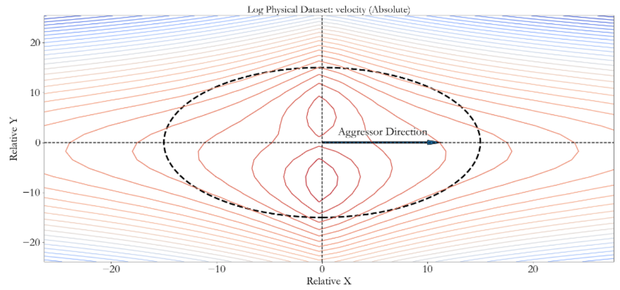
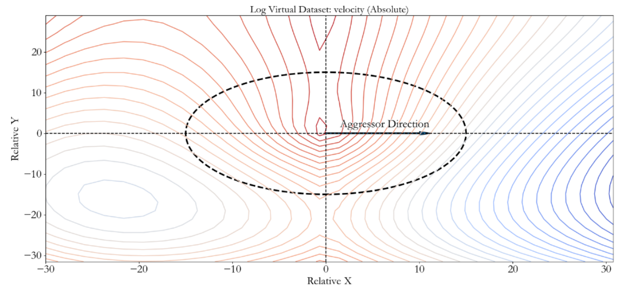
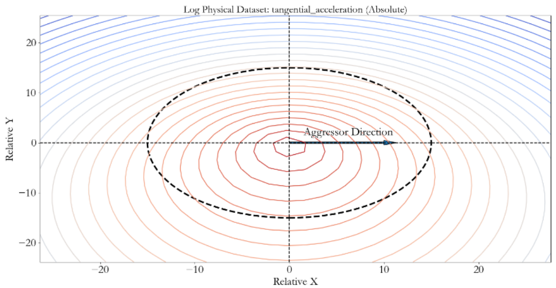
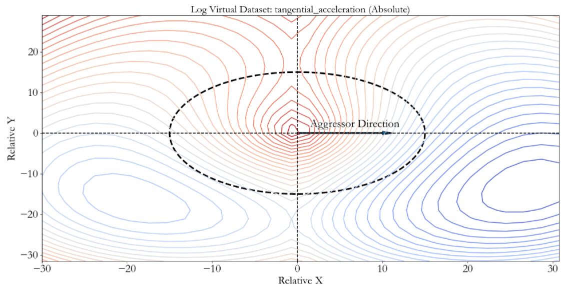
PR Stress: Log Acceleration (Log Absolute, ms^{-2})						VR: Log Acceleration (Log Absolute, ms^{-2})				
										
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	1.6861	0.0823	20.4801	6.33e-91***	Constant	2.3140	0.0616	37.5469	2.70e-301***
Positional	Distance	-0.0240	0.0026	-9.1935	4.76e-20***	Distance	-0.0947	0.0104	-9.0709	1.25e-19***
	Distance^2					Distance ^{2}	0.0028	0.0003	10.7350	7.83e-27***
	X					X	-0.0220	0.0009	-25.4028	8.22e-141***
	X^2					X^2	0.0010	0.0003	3.4762	0.0005***
	X^3					X^3				
	Abs(X)					Abs(X)	-0.0628	0.0062	-10.1945	2.31e-24***
	Y	-0.0092	0.0019	-4.9589	7.23e-07***	Y	0.0382	0.0018	20.8610	6.16e-96***
	Y^2					Y^2	0.0018	0.0003	5.8580	4.73e-09***
	Y^3					Y^3	-4.93e-05	1.21e-05	-4.0808	4.50e-05***
	Abs(Y)					Abs(Y)	-0.0254	0.0074	-3.4137	0.0006***
Demog.	Age	-0.0150	0.0032	-4.6740	3.00e-06***	Age	-0.0057	0.0020	-2.9063	0.0037**
	Exercise					Exercise	0.0228	0.0052	4.3727	1.23e-05***
	F	0.0877	0.0270	3.2479	0.001167**	F	-0.0809	0.0134	-6.0405	1.56e-09***
Adjusted r^2				0.017		**Adjusted r^2**			0.190	
F-statistic (p-value)				36.3892 (3.4176e-30***)		**F-statistic (p-value)**			617.1712 (0.0***)	
Omnibus test (p-value)				428.1166 (1.0856e-93***)		**Omnibus test (p-value)**			2226.1368(0.0***)	

Table E.1.: PR Stress and VR Log Acceleration models

PR Stress: Log Velocity (Log Absolute, ms^{-1})	VR: Log Velocity (Log Absolute, ms^{-1})
	

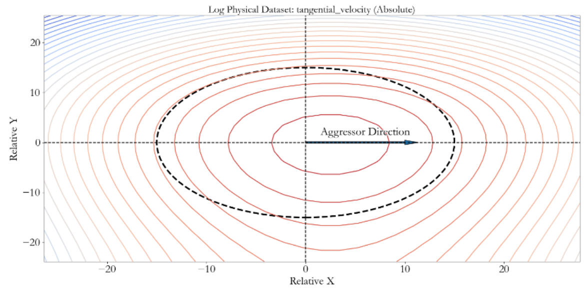
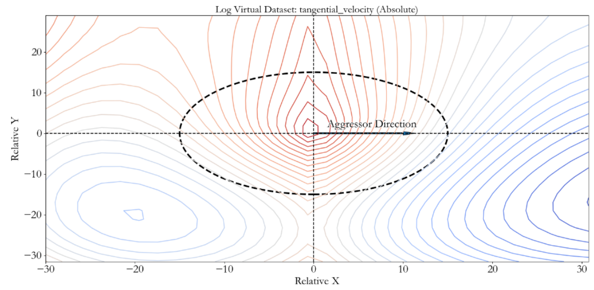
Sub.	Variable	Coeff.	Std. error	t-value	p-value	Variable	Coeff.	Std. error	t-value	p-value
C	Constant	1.6296	0.0464	35.1528	4.55e-252***	Constant	1.3538	0.0296	45.7934	0.000 ***
Positional	Distance		0.0294 0.0072	4.0826	4.50e-05***	Distance	-0.0429	0.0050	-8.5717	1.07e-17***
	Distance^2					Distance ^{2}	0.0012	0.0001	9.1736	4.87e-20***
	X					X	-0.0067	0.0008	-7.9682	1.67e-15***
	X^2					X^2	0.0006	0.0001	4.4928	7.05e-06***
	X^3					X^3	-2.63e-05	5.75e-06	-4.5813	4.64e-06***
	Abs(X)	-0.0388	0.0059	-6.5697	5.35e-11***	Abs(X)	-0.0433	0.0030	-14.6573	1.80e-48***
	Y	-0.0051	0.0011	-4.7090	2.53e-06***	Y	0.0249	0.0009	28.3276	3.77e-174***
	Y^2	-0.0034	0.0005	-7.5859	3.67e-14***	Y^2	0.0005	0.0001	3.5016	0.0005***
	Y^3					Y^3	-3.92e-05	5.79e-06	-6.7675	1.34e-11***
	Abs(Y)					Abs(Y)	-0.0103	0.0036	-2.9000	0.0037**
Demog.	Age	-0.0256	0.0018	-14.2728	1.15e-45***	Age	-0.0101	0.0009	-10.7256	8.67e-27***
	Exercise					Exercise	0.0191	0.0025	7.6470	2.12e-14***
	F					F	-0.0533	0.0064	-8.3178	9.35e-17***
				Adjusted r^2	0.054				**Adjusted r^2**	0.268
				F-statistic (p-value)	92.6956 (3.3314e-95***)				**F-statistic (p-value)**	886.0630(0.0***)
				Omnibus test (p-value)	299.4171 (9.6027e-66***)				**Omnibus test (p-value)**	2209.6022(0.0***)

Table E.2.: PR Stress and VR Log Velocity models

PR Stress: Log Tangential Acceleration (Log Absolute, ms^{-2})	VR: Log Tangential Acceleration (Log Absolute, ms^{-2})
	

Sub.	Variable	Coeff.	Std. error	t-value	p-value	Variable	Coeff.	Std. error	t-value	p-value
C	Constant	1.3390	0.0745	17.9667	8.32e-71***	Constant	1.9798	0.0541	36.59072	1.37e-286***
Positional	Distance	-0.0169	0.0024	-7.1376 1.03e-12***	Distance	-0.1053	0.0079	-13.2543	5.58e-40***	
	Distance^2					Distance ^{2}	0.0029	0.0002	12.4378	2.01e-35***
	X					X	-0.0174	0.0008	-22.9753	8.95e-116***
	X^2					X^2	0.0018	0.0002	8.1967	2.57e-16***
	X^3					X^3				
	Abs(X)					Abs(X)	-0.0577	0.0053 -10.9715	5.95e-28***	
	Y	-0.0067	0.0017	-3.9992	6.41e-05***	Y	0.0336	0.0015	21.9831	3.10e-106***
	Y^2					Y^2	0.0010	0.0002	4.9064	9.33e-07***
	Y^3					Y^3	-4.41e-05	1.00e-05	-4.3953	1.11e-05***
	Abs(Y)					Abs(Y)				
Demog.	Age	-0.0130	0.0029	-4.4543	8.53e-06***	Age	-0.0054	0.0017	-3.1123	0.0019**
	Exercise					Exercise	0.0271	0.0046	5.9204	3.25e-09***
	F	0.1033	0.0245	4.2228	2.44e-05***	F	-0.0573	0.0118	-4.8673	1.14e-06***
Adjusted r^2				0.013		**Adjusted r^2**			0.182	
F-statistic (p-value)				28.4303 (1.7140e-23***)		**F-statistic (p-value)**			646.1253 (0.000***)	
Omnibus test (p-value)				818.8949 (1.5109e-178***)		**Omnibus test (p-value)**			2404.1393 (0.000***)	

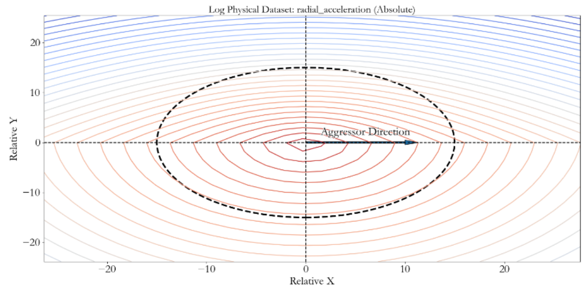
Table E.3.: PR Stress and VR Log Tangential Acceleration models

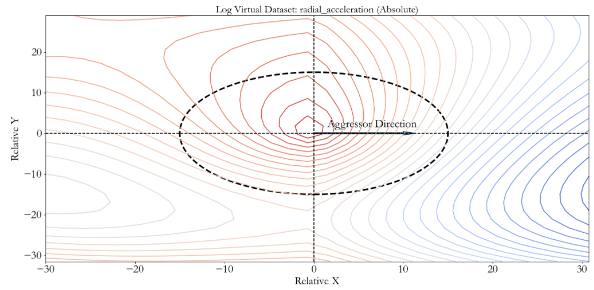
PR Stress: Log Tangential Velocity (Log Absolute, ms^{-1})	**VR: Log Tangential Velocity (Log Absolute, ms^{-1})**
	

Sub.	Variable	Coeff.	Std. error	t-value	p-value	Variable	Coeff.	Std. error	t-value	p-value
C	Constant	1.0927	0.0613	17.8267	9.34e-70***	Constant	1.1067	0.0257	42.9918	0.000***
Positional	Distance					Distance	-0.0427	0.0038	-11.2839	1.82e-29***
	Distance^2	-0.0010	8.85e-05	-10.8269	3.94e-27***	Distance ^{2}	0.0013	0.0001	7.0633	1.66e-12***
	X	0.0032	0.0010	3.2602	0.0011**	X	-0.0056	0.0007	-7.6401	2.24e-14***
	X^2					X^2	0.0009	0.0002	5.40134	6.67e-08***
	X^3					X^3	-1.45e-05	5.01e-06	-2.8995	0.0037**
	Abs(X)					Abs(X)	-0.0425	0.0025	-17.0268	1.07e-64***
	Y					Y	0.0194	0.0007	26.7596	7.54e-156***
	Y^2					Y^2				
	Y^3	-3.13e-05	9.20e-06	-3.4003	0.0007***	Y^3	-2.91e-05	4.76e-06	-6.1185	9.57e-10***
	Abs(Y)					Abs(Y)				
Demog.	Age	-0.01719	0.0019	-9.0184	2.35e-19***	Age	-0.0087	0.0008	-10.5963	3.46e-26***
	Exercise	0.0238	0.0059	4.0479	5.22e-05***	Exercise	0.0213	0.0022	9.7840	1.43e-22***
	F					F	-0.0361	0.0056	-6.4624	1.05e-10***
				Adjusted r^2	0.035				**Adjusted r^2**	0.245
				F-statistic (p-value)	59.3830 (6.2320e-61***)				**F-statistic (p-value)**	856.9583 (0.000***)
				Omnibus test (p-value)	407.7180 (2.9185e-89***)				**Omnibus test (p-value)**	3620.9801 (0.000***)

Table E.4.: PR Stress and VR Log Tangential Velocity models

PR Stress: Log Radial Acceleration (Log Absolute, ms^{-2})						VR: Log Radial Acceleration (Log Absolute, ms^{-2})				





Sub.	Variable	Coeff.	Std. error	t-value	p-value	Variable	Coeff.	Std. error	t-value	p-value
C	Constant	1.321	0.0692	19.0952	1.52e-79***	Constant	1.5269	0.0513	29.7748	6.51e-192***
Positional	Distance	-0.0160	0.0033	-4.8736	1.12e-06***	Distance	-0.0605	0.0070	-8.6345	6.19e-18***
	Distance^2					Distance ^{2}	0.0017	0.0003	5.3474	8.99e-08***
	X					X	-0.0173	0.0007	-24.2155	2.89e-128***
	X^2					X^2				
	X^3					X^3				
	Abs(X)					Abs(X)	-0.0277	0.0018	-15.7215	1.83e-55***
	Y	-0.0089	0.0017	-5.3727	7.97e-08***	Y	0.0300	0.0014	21.1329	2.20e-98***
	Y^2					Y^2				
	Y^3					Y^3	-6.73e-05	9.05e-06	-7.4367	1.06e-13***
	Abs(Y)	-0.0108	0.0037	-2.8941	0.0038**	Abs(Y)				
Demog.	Age	-0.0139	0.0027	-5.0646	4.18e-07***	Age	-0.0051	0.0016	-3.1137	0.0018**
	Exercise					Exercise	0.0138	0.0043	3.1750	0.0015**
	F					F	-0.0617	0.0112	-5.5347	3.14e-08***
			Adjusted r^2	0.017				**Adjusted r^2**	0.131	
			F-statistic (p-value)	36.7809 (1.5998e-30***)				**F-statistic (p-value)**	436.2170(0.000***)	
			Omnibus test (p-value)	1274.4458 (1.8097e-277***)				**Omnibus test (p-value)**	3823.8776 (0.000***)	

Table E.5.: PR Stress and VR Log Radial Acceleration models

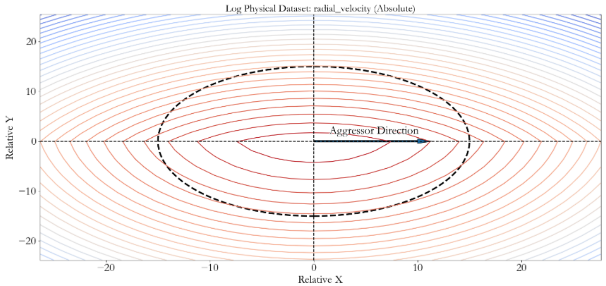
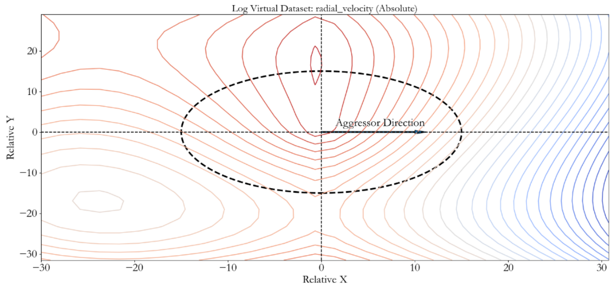
	PR Stress: Log Radial Velocity (Log Absolute, ms^{-1})					VR: Log Radial Velocity (Log Absolute, ms^{-1})				
										
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	1.1696	0.0425	27.5196	1.87e-159***	Constant	0.8358	0.0229	36.5103	2.28e-285***
Positional	Distance					Distance	-0.0189	0.0039	-4.8688	1.13e-06***
	Distance^2	-0.0008	0.0001	-6.4772	9.89e-11***	Distance ^{2}	0.0003	9.72e-05	3.2577	0.0011**
	X					X	-0.0038	0.0006	-5.8624	4.61e-09***
	X^2					X^2	0.0173	0.0007	25.4923	8.83e-142***
	X^3					X^3	-2.65e-05	4.45e-06	-5.9570	2.60e-09***
	Abs(X)					Abs(X)	-0.0184	0.0022	-8.0510	8.53e-16***
	Y	-0.0058	0.0010	-5.6641	1.53e-08***	Y				
	Y^2					Y^2				
	Y^3					Y^3				
	Abs(Y)	-0.0105	0.0022	-4.6833	2.87e-06***	Abs(Y)	-0.0073	0.0028	-2.6575	0.0079**
Demog.	Age	-0.0187	0.0017	-10.9532	1.00e-27***	Age	-0.0078	0.0007	-10.7449	7.04e-27***
	Exercise					Exercise	0.0099	0.0019	5.1039	3.35e-07***
	F					F	-0.0389	0.0050	-7.8410	4.62e-15***
				Adjusted r^2	0.038				**Adjusted r^2**	0.191
				F-statistic (p-value)	80.5986 (3.4974-67***)				**F-statistic (p-value)**	568.9188 (0.000***)
				Omnibus test (p-value)	1321.6868 (9.9854-288***)				**Omnibus test (p-value)**	6070.0603 (0.0***)

Table E.6.: PR Stress and VR Log Radial Velocity models

PR Stress: Standardised Acceleration (Absolute, ms^{-2})						**VR: Standardised Acceleration (Absolute, ms^{-2})**				
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	0.3258	0.1270	2.5655	0.0103*	Constant	1.2099	0.0596	20.3131	4.24e-91***
Positional	Distance	0.0159	0.0177	0.9002	0.3680	Distance	-0.0697	0.0100	-6.9384	4.05e-12***
	Distance^2	-0.0002	0.0005	-0.4258	0.6703	Distance ^{2}	0.0023	0.0002	9.2187	3.21e-20***
	X	0.0007	0.0034	0.2033	0.8389	X	-0.0179	0.0017	-10.3435	4.95e-25***
	X^2	-0.0002	0.0006	-0.2677	0.7889	X^2	0.0007	0.0003	2.3141	0.0207*
	X^3	-2.90e-07	2.80e-05	-0.0103	0.9918	X^3	-8.43e-08	1.19e-05	-0.0071	0.9943
	Abs(X)	-0.0180	0.0137	-1.3159	0.1883	Abs(X)	-0.0509	0.0061	-8.3642	6.32e-17***
	Y	-0.0020	0.0037	-0.5462	0.5850	Y	0.0320	0.0018	17.7262	6.17e-70***
	Y^2	-4.00e-05	0.0006	-0.0620	0.9505	Y^2	0.0016	0.0003	5.4563	4.90e-08***
	Y^3	-5.06e-05	3.39e-05	-1.4923	0.1356	Y^3	-3.79e-05	1.19e-05	-3.1832	0.0015**
	Abs(Y)	-0.0280	0.0132	-2.1288	0.0333*	Abs(Y)	-0.0264	0.0073	-3.5894	0.0003***
Demog.	Age	-0.0112	0.0039	-2.8680	0.0041**	Age	-0.0067	0.0019	-3.4452	0.0006 ***
	Exercise	0.0094	0.0118	0.7939	0.4273	Exercise	0.0144	0.0051	2.8115	0.0049**
	F	0.0667	0.0295	2.2589	0.0239*	F	-0.0568	0.0132	-4.3014	1.70e-05***
				Adjusted r^2	0.009				**Adjusted r^2**	0.135
				F-statistic (p-value)	6.8780 (1.59e-12***)				**F-statistic (p-value)**	379.1778 (0.000***)
				Omnibus test (p-value)	8440.5058 (0.000***)				**Omnibus test (p-value)**	7235.9309 (0.000***)

Table E.7.: PR Stress and VR Standardised Acceleration models

	PR Stress: Standardised Velocity (Absolute, ms^{-1})					**VR: Standardised Velocity (Absolute, ms^{-1})**				
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	0.7331	0.1251	5.8617	4.76e-09***	Constant	1.4907	0.0565	26.4019	8.14e-152***
Positional	Distance	-0.0032	0.0174	-0.1854	0.8529	Distance	-0.0272	0.0095	-2.8538	0.0043**
	Distance^2	-0.0010	0.0004	-2.2611	0.0238*	Distance ^{2}	0.0006	0.0002	2.5550	0.0106*
	X	0.0061	0.0034	1.8087	0.0705†	X	-0.0084	0.0016	-5.1007	3.41e-07***
	X^2	0.0020	0.0006	3.6174	0.0003***	X^2	0.0009	0.0003	3.0998	0.0019**
	X^3	-2.86e-05	2.76e-05	-1.0361	0.3002	X^3	-0.0001	1.12e-05	-5.9079	3.50e-09***
	Abs(X)	-0.0374	0.0135	-2.7729	0.0056**	Abs(X)	-0.0827	0.0058	-14.3380	1.83e-46***
	Y	-0.0090	0.0037	-2.4540	0.0141*	Y	0.0464	0.0017	27.0814	1.58e-159***
	Y^2	-0.0030	0.0006	-4.7917	1.68e-06***	Y^2	-0.0003	0.0003	-0.9344	0.3501
	Y^3	1.61e-05	3.34e-05	0.4823	0.6296	Y^3	-0.0001	1.13e-05	-6.8173	9.46e-12***
	Abs(Y)	0.0189	0.0130	1.4549	0.1457	Abs(Y)	-0.0077	0.0070	-1.1112	0.2665
Demog.	Age	-0.0298	0.0038	-7.7729	8.61e-15***	Age	-0.0230	0.0018	-12.5215	7.05E-36***
	Exercise	0.0454	0.0117	3.8897	0.0001***	Exercise	0.0376	0.0049	7.7132	1.27e-14 ***
	F	-0.0308	0.0291	-1.0580	0.2901	F	-0.0937	0.0125	-7.4848	7.37e-14***
Adjusted r^2				0.038		**Adjusted r^2**			0.223	
F-statistic (p-value)				27.7657 (1.14e-62***)		**F-statistic (p-value)**			694.6589 (0.000***)	
Omnibus test (p-value)				6026.7071 (0.000***)		**Omnibus test (p-value)**			5605.92144 (0.000***)	

Table E.8.: PR Stress and VR Standardised Velocity models

PR Stress: Standardised Tangential Acceleration (Absolute, ms^{-2})						VR: Standardised Tangential Acceleration (Absolute, ms^{-2})				
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	0.2389	0.1270	1.8816	0.0599†	Constant	1.2538	0.0612	20.4933	1.13e-92***
Positional	Distance	0.0073	0.0177	0.4123	0.6801	Distance	-0.0957	0.0104	-9.2383	2.67e-20***
	Distance^2	0.0001	0.0005	0.1887	0.8503	Distance ^{2}	0.0031	0.0003	12.0896	1.45E-33***
	X	0.0043	0.0034	1.2660	0.2055	X	-0.0166	0.0017	-9.5908	9.42e-22***
	X^2	0.0010	0.0006	1.7295	0.0838†	X^2	0.0016	0.0003	5.4726	4.47e-08***
	X^3	-3.45e-06	2.80e-05	-0.1232	0.9019	X^3	1.02e-05	1.19e-05	0.8593	0.3902
	Abs(X)	-0.0309	0.0137	-2.2567	0.0241*	Abs(X)	-0.0593	0.0061	-9.7168	2.76e-22***
	Y	-0.0007	0.0037	-0.1803	0.8569	Y	0.0303	0.0018	16.6921	2.93e-62***
	Y^2	-0.0009	0.0006	-1.3990	0.1619	Y^2	0.0015	0.0003	5.0596	4.23e-07***
	Y^3	-4.71e-05	3.39e-05	-1.3881	0.1652	Y^3	-2.41e-05	1.20e-05	-2.0096	0.0445*
	Abs(Y)	-0.0106	0.0132	-0.8033	0.4218	Abs(Y)	-0.0146	0.0074	-1.9807	0.0476*
Demog.	Age	-0.0086	0.0039	-2.1997	0.0279*	Age	-0.0065	0.0019	-3.3526	0.0008 ***
	Exercise	0.0163	0.0118	1.3785	0.1681	Exercise	0.0232	0.0052	4.4966	6.93e-06***
	F	0.1060	0.0295	3.5899	0.0003***	F	-0.0410	0.0133	-3.0948	0.0020**
			Adjusted r^2	0.009				**Adjusted r^2**	0.129	
			F-statistic (p-value)	7.0296 (7.18e-13***)				**F-statistic (p-value)**	359.6860 (0.0000***)	
			Omnibus test (p-value)	9191.0761 (0.000***)				**Omnibus test (p-value)**	10918.3564 (0.000***)	

Table E.9.: PR Stress and VR Standardised Tangential Acceleration models

PR Stress: Standardised Tangential Velocity (Absolute, ms^{-1})						VR: Standardised Tangential Velocity(Absolute, ms^{-1})				
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	0.5693	0.1249	4.5590	5.21e-06***	Constant	1.6084	0.0584	27.5218	1.27e-164***
Positional	Distance	-0.0179	0.0174	-1.0309	0.3026	Distance	-0.0844	0.0099	-8.5341	1.48e-17***
	Distance^2	-0.0004	0.0004	-0.9384	0.3481	Distance ^{2}	0.0023	0.0002	9.1873	4.29e-20***
	X	0.0129	0.0034	3.8482	0.0001***	X	-0.0085	0.0017	-5.1285	2.94e-07***
	X^2	0.0039	0.0006	7.0067	2.64e-12***	X^2	0.0027	0.0003	9.3621	8.36e-21***
	X^3	-0.0001	2.76e-05	-1.8700	0.0615†	X^3	-4.02e-05	1.14e-05	-3.5413	0.0004 ***
	Abs(X)	-0.0602	0.0135	-4.4664	8.06e-06***	Abs(X)	-0.0950	0.0058	-16.2810	2.47e-59***
	Y	-0.0037	0.0037	-1.0205	0.3075	Y	0.0385	0.0017	22.1921	3.28e-108***
	Y^2	-0.0043	0.0006	-6.8639	7.20e-12***	Y^2	-0.0004	0.0003	-1.3196	0.1870
	Y^3	-2.88e-06	3.34e-05	-0.0864	0.9311	Y^3	-0.0001	1.14e-05	-4.4712	7.81e-06***
	Abs(Y)	0.0481	0.0130	3.7112	0.0002***	Abs(Y)	0.0125	0.0070	1.7742	0.0760†
Demog.	Age	-0.0257	0.0038	-6.7160	1.99e-11***	Age	-0.0211	0.0019	-11.3544	8.15E-30***
	Exercise	0.0622	0.0116	5.3368	9.72e-08***	Exercise	0.0482	0.0049	9.7839	1.43e-22***
	F	0.0333	0.0290	1.1469	0.2515	F	-0.0728	0.0127	-5.7495	9.04e-09***
				Adjusted r^2	0.041				**Adjusted r^2**	0.206
				F-statistic (p-value)	29.9146 (6.92e-68***)				**F-statistic (p-value)**	625.8416 (0.000***)
				Omnibus test (p-value)	5972.7670 (0.000***)				**Omnibus test (p-value)**	8013.4157 (0.000***)

Table E.10.: PR Stress and VR Standardised Tangential Velocity models

PR Stress: Standardised Radial Acceleration (Absolute, ms^{-2})						VR: Standardised Radial Acceleration (Absolute, ms^{-2})				
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	0.3636	0.1271	2.8611	0.0042**	Constant	0.7974	0.0631	12.6323	1.75E-36***
Positional	Distance	0.0166	0.0177	0.9388	0.3479	Distance	-0.0299	0.0107	-2.7996	0.0051**
	Distance^2	-0.0003	0.0005	-0.7538	0.4510	Distance ^{2}	0.0005	0.0003	1.7608	0.0783†
	X	-0.0030	0.0034	-0.8662	0.3864	X	-0.0153	0.0018	-8.5359	1.46e-17***
	X^2	-0.0015	0.0006	-2.5534	0.0107*	X^2	-0.0005	0.0003	-1.4914	0.1359
	X^3	-2.53e-06	2.81e-05	-0.0901	0.9282	X^3	-1.38e-05	1.23e-05	-1.1231	0.2614
	Abs(X)	0.0036	0.0137	0.2609	0.7941	Abs(X)	-0.0164	0.0063	-2.6077	0.0091**
	Y	-0.0029	0.0037	-0.7754	0.4381	Y	0.0277	0.0019	14.7764	3.14e-49***
	Y^2	0.0011	0.0006	1.7316	0.0834†	Y^2	0.0009	0.0003	3.0084	0.0026**
	Y^3	-4.29e-05	3.40e-05	-1.2632	0.2066	Y^3	-0.0001	1.24e-05	-5.0539	4.36e-07***
	Abs(Y)	-0.0401	0.0132	-3.0427	0.0024**	Abs(Y)	-0.0163	0.0076	-2.1380	0.0325*
Demog.	Age	-0.0118	0.0039	-3.0306	0.0024**	Age	-0.0063	0.0020	-3.1445	0.0017 ***
	Exercise	-0.0007	0.0119	-0.0556	0.9556	Exercise	0.0052	0.0053	0.9818	0.3262
	F	0.0057	0.0296	0.1944	0.8459	F	-0.0578	0.0137	-4.2264	2.38e-05***
				Adjusted r^2	0.007				**Adjusted r^2**	0.073
				F-statistic (p-value)	5.6096 (1.12e-09***)				**F-statistic (p-value)**	191.9830 (0.000***)
				Omnibus test (p-value)	9522.9555 (0.000***)				**Omnibus test (p-value)**	12630.4076 (0.000***)

Table E.11.: PR Stress and VR Standardised Radial Acceleration models

PR Stress: Standardised Radial Velocity (Absolute, ms^{-1})						VR: Standardised Radial Velocity (Absolute, ms^{-1})				
Sub.	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**	**Variable**	**Coeff.**	**Std. error**	**t-value**	**p-value**
C	Constant	0.7365	0.1261	5.8391	5.45e-09***	Constant	1.3299	0.0604	22.0137	1.60e-106***
Positional	Distance	0.0039	0.0176	0.2240	0.8228	Distance	-0.0239	0.0102	-2.3341	0.0196*
	Distance^2	-0.0008	0.0004	-1.8379	0.0661†	Distance ^{2}	0.0002	0.0003	0.9664	0.3339
	X	-0.0010	0.0034	-0.3030	0.7619	X	-0.0060	0.0017	-3.4993	0.0005***
	X^2	-0.0005	0.0006	-0.8759	0.3811	X^2	-0.0004	0.0003	-1.3475	0.1778
	X^3	-1.21e-05	2.78e-05	-0.4364	0.6626	X^3	-7.50e-05	1.17e-05	-6.3826	1.77e-10***
	Abs(X)	-0.0016	0.0136	-0.1210	0.9037	Abs(X)	-0.0428	0.0060	-7.0999	1.28e-12***
	Y	-0.0123	0.0037	-3.3346	0.0009***	Y	0.0416	0.0018	23.1922	6.50e-118***
	Y^2	-0.0003	0.0006	-0.5154	0.6063	Y^2	0.0006	0.0003	2.1761	0.0296*
	Y^3	3.84e-05	3.37e-05	1.1387	0.2548	Y^3	-7.76e-05	1.18e-05	-6.5653	5.28e-11***
	Abs(Y)	-0.0204	0.0131	-1.5584	0.1192	Abs(Y)	-0.0208	0.0073	-2.8559	0.0043**
Demog.	Age	-0.0258	0.0039	-6.6526	3.07e-11***	Age	-0.0216	0.0019	-11.2126	4.06e-29***
	Exercise	0.0101	0.0118	0.8612	0.3891	Exercise	0.0205	0.0051	4.0129	6.01e-05***
	F	-0.0682	0.0293	-2.3236	0.0202*	F	-0.0907	0.0131	-6.9286	4.34e-12***
				Adjusted r^2	0.022				**Adjusted r^2**	0.151
				F-statistic (p-value)	16.0486 (3.04e-34***)				**F-statistic (p-value)**	430.9162 (0.000***)
				Omnibus test (p-value)	7461.8120 (0.000***)				**Omnibus test (p-value)**	12241.8710 (0.000***)

Table E.12.: PR Stress and VR Standardised Radial Velocity models

F. Recall survey Questions

The following tables show the survey questions distributed to participants three months after their experiments. The simplified question text is shown, including both positive and negative framings, alongside any associated coding scheme and the allocated domains. If the question also asked for a confidence rating, this is shown by an asterisk (*) after the question text. If the question was scored using a coding scheme, the appropriate item is indicated, and detailed in Table F.3.

F.1. PR experiment

Simplified question text	**Coding scheme**	**Domains**
What happened in this experiment? *	Coding scheme 1	1, 2
How many participants took part in the the experiment (including yourself)? *	± Deviation from ground truth	1, 6
How many participants were female (including yourself, if appropriate)? *	± Deviation from ground truth	1, 6
What was used to mark out the physical boundary between the two parts of the gym (in blue)? *	Coding schme 2	1, 2
Where were you standing when you noticed the aggressor? Please be as specific as possible, describing your position in relation to the numbers on the image. *	Positive deviation from ground truth	1,2,5,7
Which gender was the aggressor?	Binary deviation from ground truth	1, 2, 5
What was the colour of the aggressor's hair? *	Binary deviation from ground truth	1, 2, 5
What were the instructions given by the aggressor?*	Coding scheme 3	1, 2, 5

Simplified question text	Coding scheme	Domains
What was your very first (e.g. first 5 seconds) behaviour when you noticed the aggressor?*	Coding scheme 4	1, 4, 5,7
I did/did not verbally confront the aggressor*	Binary deviation from ground truth	1,4,5
No one/At least one person verbally confronted the aggressor*	Binary deviation from ground truth	1,5,6
I/No one confronted the aggressor physically*	Binary deviation from ground truth	1,4,5
At least one/No one of the other participants confronted the aggressor physically*	Binary deviation from ground truth	1,5,6
I did not/did take another participant by the hand*	Binary deviation from ground truth	1,4,6
No one/At least one other participant was holding hands *	Binary deviation from ground truth	1,6
Our initial collective reaction was to get together to ward off the aggressor/everyone for themselves*	N/A	None-Prosociality
I did not help/helped a participant that the aggressor had approached*	Binary deviation from ground truth	1,4,5,6
No one/At least one of the other participants helped a participant that the aggressor had approached*	Binary deviation from ground truth	1,5,6
I tried/did not try dodging the aggressor when they started pursuing *	Binary deviation from ground truth	1,4,5
At least one of the other participants/ No one tried dodging the aggressor when they started pursuing him/her*	Binary deviation from ground truth	1,5,6
PANAS (P and N scores)	± Deviation from ground truth	1,3
STAI	± Deviation from ground truth	1,3
Social responses- Solidarity	± Deviation from ground truth	1,3

Simplified question text	Coding scheme	Domains
Social responses- Satisfaction	± Deviation from ground truth	1,3
Social responses- Centrality	± Deviation from ground truth	1,3
Social responses- Individual self-stereotyping	± Deviation from ground truth	1,3
Social responses- In-group homogeneity	± Deviation from ground truth	1,3
Social responses- Shared identity	± Deviation from ground truth	1,3
Science says that people are very prosocial/individualistic when they are confronted by a threat. From your experience during the experiment, how much do you agree with this statement?	N/A	None-Prosciality

Table F.1.: PR experiment recall survey

F.2. VR experiment

Simplified question text	Coding scheme	Domains
What happened in this experiment? *	Coding scheme 1	1, 2
How many avatars did you see in the experiment (including yourself)? *	± Deviation from ground truth	1, 6
How many avatars were female (including yourself, if appropriate)? *	± Deviation from ground truth	1, 6
What were the instructions given by the aggressor? *	Coding scheme 3	1, 2, 5
Which gender was the aggressor?	Binary deviation from ground truth	1, 2, 5
What was the colour of the aggressor's hair? *	Binary deviation from ground truth	1, 2, 5
What was the colour of the aggressor's shirt? *	Binary deviation from ground truth	1, 2, 5

Simplified question text	**Coding scheme**	**Domains**
Where were you standing during the introductions part of the experiment? Please be as specific as possible, describing your position in relation to the numbers on the image.*	Positive deviation from ground truth	1, 2, 7
Where were you standing when you noticed the aggressor? Please be as specific as possible, describing your position in relation to the numbers on the image.*	Positive deviation from ground truth	1, 2, 5, 7
I did/did not verbally confront the aggressor*	Binary deviation from ground truth	1,4,5
No one/At least one person verbally confronted the aggressor*	Binary deviation from ground truth	1,5,6
I/No one confronted the aggressor physically*	Binary deviation from ground truth	1,4,5
At least one/No one of the other participants confronted the aggressor physically*	Binary deviation from ground truth	1,5,6
I did not/did take another participant by the hand*	Binary deviation from ground truth	1,4,6
No one/At least one other participant was holding hands *	Binary deviation from ground truth	1,6
Our initial collective reaction was to get together to ward off the aggressor/everyone for themselves*	N/A	None-Prosociality
I did not help/helped a participant that the aggressor had approached*	Binary deviation from ground truth	1,4,5,6
No one/At least one of the other participants helped a participant that the aggressor had approached*	Binary deviation from ground truth	1,5,6
I tried/did not try dodging the aggressor when they started pursuing *	Binary deviation from ground truth	1,4,5
At least one of the other participants/ No one tried dodging the aggressor when they started pursuing him/her*	Binary deviation from ground truth	1,5,6

Simplified question text	Coding scheme	Domains
I did not/did go behind the dividing wall*	Binary deviation from ground truth	1,4,7
Nobody/ at least one participant went behind the dividing wall*	Binary deviation from ground truth	1,6,7
I went/ did not go around the central column	Binary deviation from ground truth	1,4,7
At least one other participant/No one went around the central column*	Binary deviation from ground truth	1,5,7
I did not/did try to leave the room*	Binary deviation from ground truth	1,4,7
No one/At least one other participant tried to leave the room*	Binary deviation from ground truth	1,5,7
PANAS (P and N scores)	± Deviation from ground truth	1,3
STAI	± Deviation from ground truth	1,3
Social responses- Solidarity	± Deviation from ground truth	1,3
Social responses- Satisfaction	± Deviation from ground truth	1,3
Social responses- Centrality	± Deviation from ground truth	1,3
Social responses- Individual self-stereotyping	± Deviation from ground truth	1,3
Social responses- In-group homogeneity	± Deviation from ground truth	1,3
Social responses- Shared identity	± Deviation from ground truth	1,3
Do you remember how much you thought the other participants were real? *	± Deviation from ground truth	1,3

Simplified question text	Coding scheme	Domains
Science says that people are very prosocial/individualistic when they are confronted by a threat. From your experience during the experiment, how much do you agree with this statement?	N/A	None-Prosciality

Table F.2.: VR experiment recall survey

F.3. Coding schemes

Below are the coding schemes used for the recall surveys.

Item	Scoring
1	Mention of the aggressor for relevant groups or cognitive and physical tasks otherwise
2	Mention of table tennis /ping pong boards
3	Mention of a chase for the relevant experiments
4	Answer matched behaviour observed in videos i.e.: Whether the statement is true within the 5 first sec after the aggressor started going towards the pps = 1; otherwise, if something is false within the statement score = 0

Table F.3.: Coding schemes for recall survey