# A Socio-Technical Approach to Crowd Management: Understanding Decision-Making and Movement Patterns with Virtual Reality and Machine Learning

Xiangmin Yang

July 2025

Supervised by

Professor Washington Ochieng

Professor Arnab Majumdar

Associate Professor Emilia Grass

A Thesis submitted as fulfilment of the requirements for the Degree of Doctor of Philosophy (PhD) and the Diploma of Membership of Imperial College London (DIC)

Centre for Transport Engineering and Modelling, Department of Civil and Environmental Engineering, Imperial College London, United Kingdom

# Statement of originality

I hereby certify that all the work presented in this thesis is original, and is my own. If information has been used from other sources, this has been fully referenced, or appropriately acknowledged.

Xiangmin Yang

# Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

# Abstract

Crowd disasters, which are events in which deaths and injuries occur due to the high density of people, happen all too often. With greater urbanisation and mobility, the frequency of such disasters increases every year. Numerous attempts to improve crowd safety by better monitoring and prediction of crowds, have been hindered by the lack of systematic and integration of human behaviour and crowd social dynamics, relying instead upon heuristics.

This thesis tackles this gap by adopting a socio-technical perspective that emphasises the role of human actors and the surrounding social environment to contribute to a more comprehensive understanding of crowd behaviour.

Given that crowd disaster databases suffer from poor definition for their characteristics and fail to adequately incorporate the effect of socio-external environments, this thesis begins by constructing a comprehensive database of 293 past crowd disasters using data from news sources. Cluster analysis identified nine distinct types of crowd disasters, with the jostle disaster having the greatest frequency (20%).

A jostle scenario is reproduced in a virtual reality (VR) experiment, given the logistical and ethical difficulties associated in undertaking a physical experiment. Over 70 individuals participated in three rounds, with over 22 simultaneous participants. This experiment not only assesses the effectiveness of VR as a tool for analysing movement patterns in crowd analysis but also identifies two distinct behaviour types: ‘patient’ and ‘impatient’.

A decision-making inference framework that integrates inverse reinforcement learning and model interpretability is proposed. Data from the VR experiments enables quantification of the impact of numerous factors, including interactions with surrounding crowds, on the individual’s decision making. This framework is also tested on data from a case study in

Nanjing. Impacts from built environment features on pedestrians are quantified, and comfort zones are identified.

This thesis concludes with discussion on the implications for improving crowd management practices.

# Acknowledgements

Firstly, I must express my gratitude to my supervisors for their support throughout my PhD study. Thank Professor Washington Ochieng for taking me as his student and all his continuous encouragement and guidance. Thank Professor Arnab Majumdar for all the discussions we had, no matter whenever I need help. Thank Associate Professor Emilia Grass for teaching me all the skills and tips for improving my research and writing.

Secondly, I must thank my parents, Dr. Wenhui Yang and Professor Jianying Guo. I could not have reached this stage without your love and support, and this gratitude extends to my entire family. I am proud to say that I finally fulfil the dream when you received your PhD degrees. A special thank you to Ms. Yanfei Yin for all your love, and I love you as well.

Thirdly, the journey towards PhD is never easy, but with my colleagues, it was at least not lonely. I would like to thank Mr. Tianyu Zhao, Dr. Jingshuo Qiu, Mr. Josh Kirk, Mr. Angelos Malandrakis, Dr. Fahmy Adan, Dr. Xuehao Zhai, Ms. Niki Boyce, and especially Dr. Felix Feng. I am also grateful to my co-workers, Miss Yuelin Liu, Ms. Zumeng Liu, Dr. Yi Yu and Dr. Liu Yang, for our successful collaborations.

Many aspects of my PhD research would not have been possible without the help from the administrative team at CEE and ISST. Thank the development team of VIRSE, Dr.Mark Sutton and Mr. Fred Tovey-Ansell, for their support of developing virtual environment. Gratitude towards all the participants of my experiment, whom I cannot name due to ethical restrictions. I also would like to thank Mr. Ian Fox for reaching out to me after a conference presentation and introducing me to stakeholders in the field, and to those stakeholders for generously taking my interviews.

Last but not least, I would like to thank my friends. To Mr. Shengyi Jiang, who insists on going by this title until he passes his PhD viva, thank you for your friendship and entering this journey together with me. For Ms. Ruyi Yang, thank you for always listening during my countless emotional moments. For Dr. Zhiyao Zhu and Mr. Zihan Yang, it is always enjoyable to play games with you. I also would like to thank Ms. Sipu Zheng, Mr. Bo Yang, Mr. Jinna Sun, Ms.

Xinyao Liang, Ms. Ziyi Han, Ms. Siyi Liu, Mr. Tianrui Hou, Mr. Yancong Huang, and the list could go on…

Xiangmin Yang

London, July 2025

# List of Acronyms

**ADE** – Average Displacement Error

**AIRL** – Adversarial Inverse Reinforcement Learning

**BC** – Behaviour Cloning

**CA** – Cellular Automata

**CCTV** – Closed-Circuit Television

**DIM-ICE** – Design, Information, Management – Ingress, Circulation, Egress

**DL** – Deep Learning

**DTW** – Dynamic Time Warping

**EIF** – Events Industry Forum

**FDE** – Final Displacement Error

**FIST** – Force, Information, Space, Time

**GAN** – Generative Adversarial Network

**GAIL –** Generative Adversarial Imitation Learning

**IRL** – Inverse Reinforcement Learning

**LSTM** – Long Short Term Memory

**MDP** – Markov Decision Process

**ML** – Machine Learning

**RAMP** – Routes, Areas, Movement, Profile

**RL** – Reinforcement Learning

**SFM** – Social Force Model

**SHAP** – SHapley Additive exPlanations

**UK** – United Kingdom

**UKCMA** – United Kingdom Crowd Management Association

**VIRSE** – Virtual Reality Student Experience

**VR** – Virtual Reality

# Table of Contents


**1. INTRODUCTION ..... 17**

**1.1 BACKGROUND ..... 17**
**1.2 AIM AND OBJECTIVES ..... 20**
**1.3 THESIS STRUCTURE ..... 21**
1.3.1 CHAPTER OUTLINES ..... 22

**2. CROWD MANAGEMENT AND CROWD DISASTER ..... 25**

**2.1 INTRODUCTION ..... 25**
**2.2 CROWD MANAGEMENT ..... 25**
2.2.1 PRE-EVENT ..... 26
2.2.2 IN-EVENT ..... 30
**2.3 SOCIO-TECHNICAL SYSTEM ..... 33**
2.3.1 COMPLEX SYSTEM ..... 33
2.3.2 SYSTEM THINKING ..... 36
**2.4 CROWD DISASTER ..... 38**
2.4.1 FORCE, INFORMATION, SPACE AND TIME (FIST) MODEL ..... 39
2.4.2 CASE STUDY: THE 2014 SHANGHAI CROWD DISASTER ..... 44
**2.5 SUMMARY ..... 47**

**3. CROWD ANALYSIS ..... 49**

**3.1 INTRODUCTION ..... 49**
**3.2 MACROSCOPIC APPROACH ..... 50**
3.2.1 DENSITY ESTIMATION ..... 51
3.2.2 ACTION RECOGNITION ..... 52
**3.3 MICROSCOPIC (AGENT-BASED) APPROACH ..... 55**
3.3.1 CROWD COUNTING ..... 56
3.3.2 CROWD MODELLING ..... 58
3.3.3 SUMMARY AND COMPARISON ..... 71
**3.4 MESOSCOPIC APPROACH ..... 72**
**3.5 SUMMARY ..... 73**

**4. METHODOLOGICAL OVERVIEW ..... 75**

**4.1 INTRODUCTION ..... 75**
**4.2 RESEARCH PHILOSOPHY ..... 75**
**4.3 RESEARCH APPROACH ..... 77**
**4.4 RESEARCH STRATEGY ..... 78**
4.4.1 CHAPTER 5 RESEARCH STRATEGY ..... 78
4.4.2 CHAPTER 6 RESEARCH STRATEGY ..... 79
4.4.3 CHAPTER 7 RESEARCH STRATEGY ..... 80

**5. CROWD DISASTER DATABASE ..... 82**

**5.1 INTRODUCTION ..... 82**
**5.2 LITERATURE REVIEW ..... 83**
**5.3 METHODOLOGY ..... 85**
5.3.1 DATABASE CONSTRUCTION ..... 85
5.3.2 DATABASE ANALYSIS ..... 93
**5.4 RESULTS AND DISCUSSION ..... 96**
5.4.1 DATA QUALITY ..... 96
5.4.2 DATA ANALYSIS ..... 97
**5.5 SUMMARY ..... 105**

**6. VIRTUAL REALITY EXPERIMENTS ..... 107**

**6.1 INTRODUCTION ..... 107**
**6.2 LITERATURE REVIEW ..... 108**
**6.3 EXPERIMENTAL METHODOLOGY ..... 110**
6.3.1 EXPERIMENT SETUP ..... 110
6.3.2 EXPERIMENT EXECUTION ..... 114
**6.4 DATA ANALYSIS METHODOLOGY ..... 115**
6.4.1 DATA PREPROCESSING ..... 115
6.4.2 CHANGE POINT DETECTION ..... 117
6.4.3 CLUSTER ANALYSIS ..... 118
**6.5 RESULTS AND DISCUSSION ..... 119**
6.5.1 EXPERIMENT ..... 119
6.5.2 DATA ANALYSIS ..... 122
6.5.3 IMPLICATION ..... 127
**6.6 SUMMARY ..... 128**

**7. CROWD DECISION MAKING INFERENCE FRAMEWORK ..... 130**

**7.1 INTRODUCTION ..... 130**
**7.2 LITERATURE REVIEW ..... 131**
7.2.1 PEDESTRIAN DECISION MAKING ..... 131
7.2.2 INVERSE REINFORCEMENT LEARNING (IRL) ..... 132
**7.3 METHODOLOGY ..... 134**
7.3.1 MODEL FRAMEWORK ..... 134
7.3.2 MODEL INTERPRETABILITY ..... 143
**7.4 RESULTS AND DISCUSSION ..... 144**
7.4.1 AIRL TRAINING ..... 144
7.4.2 DECISION MAKING INFERENCE ..... 145
7.4.3 IMPLICATIONS ..... 151
**7.5 SUMMARY ..... 153**

**8. IMPLICATIONS ..... 154**

**8.1 INTRODUCTION ..... 154**
**8.2 IMPLICATION FOR CROWD MANAGEMENT ..... 154**
8.2.1 EXTENDED FIST MODEL ..... 156
8.2.2 PATIENT AND IMPATIENT BEHAVIOURS ..... 157
8.2.3 APPLICATION OF THE SHAP-CALIBRATED SFM ..... 159

**8.3 IMPLICATION FOR USAGE OF VIRTUAL REALITY IN CROWD ANALYSIS ..... 160**
8.3.1 ENVIRONMENT DEVELOPMENT ..... 160
8.3.2 EXPERIMENT DESIGN ..... 162
8.3.3 EXPERIMENT EXECUTION ..... 165
**8.4 IMPLICATION FOR UTILISING PROPOSED DECISION MAKING INFERENCE FRAMEWORK ..... 166**
8.4.1 NANJING XINJIEKOU CASE STUDY ..... 167

**9. CONCLUSION AND FUTURE WORK ..... 183**

**9.1 REVISITING THE RESEARCH AIM AND OBJECTIVES ..... 183**
9.1.1 DEFINE QUALITATIVE CHARACTERISTICS OF CROWD DISASTER ..... 183
9.1.2 CONSTRUCT A DATABASE OF PAST CROWD DISASTERS ..... 184
9.1.3 ASSESS THE FEASIBILITY AND RELIABILITY OF USING VIRTUAL REALITY (VR) FOR INVESTIGATING CROWD BEHAVIOUR IN SPECIFIC SCENARIO. ..... 185
9.1.4 IDENTIFY MOVEMENT PATTERNS UNDER SPECIFIC SCENARIO SETTINGS ..... 186
9.1.5 PROPOSE AN APPROACH TO INFER DECISION MAKING BEHIND CROWD BEHAVIOUR ACROSS VARIOUS SCENARIOS. ..... 186
9.1.6 PROPOSED IMPROVEMENTS IN PRACTICAL CROWD MANAGEMENT ..... 187
**9.2 FUTURE WORK ..... 187**
**9.3 FINAL THOUGHTS ..... 188**

**BIBLIOGRAPHY ..... 190**

**APPENDIX I ..... 214**

**APPENDIX II ..... 227**

**APPENDIX III ..... 236**

**APPENDIX IV ..... 238**

# List of Figures

FIGURE 1.1 LOVE PARADE IN DUISBURG, GERMANY (MÜSELER, 2010) ..... 18

FIGURE 1.2 THESIS CHAPTERS AND CORRESPONDING OBJECTIVES ..... 24

FIGURE 2.1 THREE STAGES OF CROWD MANAGEMENT (SHARMA, ET AL., 2018), REPRODUCED WITH PERMISSION FROM SPRINGER NATURE ..... 26

FIGURE 2.2 RETRACTABLE BELT BARRIER ..... 30

FIGURE 2.3 PRESSURE BARRIER (4028MDK09, 2010) ..... 31

FIGURE 2.4 LANE FORMATION (ZHANG, ET AL., 2012), REPRODUCED WITH PERMISSION ..... 34

FIGURE 2.5 (A) ARCH FORMATION FROM COMPUTATIONAL SIMULATION (HELBING, ET AL., 2000), REPRODUCED WITH PERMISSION FROM SPRINGER NATURE; (B) ARCH FORMATION FROM PHYSICAL EXPERIMENT (YE, ET AL., 2024), REPRODUCED WITH PERMISSION FROM ELSEVIER ... 35

FIGURE 2.6 KEY FEATURES OF SYSTEM THINKING (SOTNIK, ET AL., 2021), REPRODUCED WITH PERMISSION FROM SPRINGER NATURE ..... 36

FIGURE 2.7 SOCIO-TECHNICAL SYSTEM FRAMEWORK (CHALLENGER, ET AL., 2010)..... 37

FIGURE 2.8 SHOCK WAVE PROPAGATION ALONG THE QUEUE (WANG, ET AL., 2019), REPRODUCED WITH PERMISSION FROM ELSEVIER..... 42

FIGURE 2.9 SHANGHAI 2014 CROWD DISASTER FROM A SYSTEM PERSPECTIVE (YANG, ET AL., 2024).. 47

FIGURE 3.1 TAXONOMY OF CROWD ANALYSIS ..... 49

FIGURE 3.2 FUNDAMENTAL DIAGRAM OF PEDESTRIAN FLOW (NARANG, ET AL., 2015)..... 51

FIGURE 3.3 SCHEMATIC DIAGRAM OF CELLULAR AUTOMATA (LI, ET AL., 2018), REPRODUCED WITH PERMISSION FROM SPRINGER ..... 62

FIGURE 3.4 NEURAL NETWORK (KUGUNAVAR & PRABHAKAR, 2021)..... 64

FIGURE 3.5 LSTM ARCHITECTURE (JING, ET AL., 2022) ..... 65

FIGURE 3.6 INTERACTIONS OF RL AGENTS WITH ENVIRONMENT (HASSANI, ET AL., 2022), REPRODUCED WITH PERMISSION FROM ELSEVIER ..... 69

FIGURE 4.1 RESEARCH STRATEGY FLOWCHART ..... 78

FIGURE 5.1 INPUT IN THE FACTIVA SEARCH BOX (YANG, ET AL., 2024) ..... 87

FIGURE 5.2 DATABASE DATA COLLECTION FLOWCHART (YANG, ET AL., 2024) ..... 89

FIGURE 5.3 SOCIAL SCENARIO DISTRIBUTION (YANG, ET AL., 2024) ..... 98

FIGURE 5.4 PHYSICAL ENVIRONMENT DISTRIBUTION (YANG, ET AL., 2024) ..... 99

FIGURE 5.5 INFORMATION DISTRIBUTION (YANG, ET AL., 2024) ..... 100

FIGURE 5.6 PHYSICAL SPACE DISTRIBUTION (YANG, ET AL., 2024) ..... 101

FIGURE 5.7 TIME DISTRIBUTION (YANG, ET AL., 2024) ..... 102

FIGURE 5.8 CASUALTY DISTRIBUTION (YANG, ET AL., 2024) ..... 103

FIGURE 6.1 LAYOUT OF VIRTUAL ENVIRONMENT ..... 111

FIGURE 6.2 EXPERIMENT SETUP, (A) THE TARGET GATE OF THE PARTICIPANTS; (B) SCORE BOARD; (C) AVATAR CONTROL; (D) INDICATION OF PHYSICAL CONTACT ..... 112

FIGURE 6.3 POSITION DATA FROM THE EXPERIMENT ..... 120

FIGURE 6.4 RESPONSES FROM QUESTIONNAIRE ..... 121

FIGURE 6.5 CHANGE POINT DETECTION ..... 123

FIGURE 6.6 DENDROGRAM OF HIERARCHICAL CLUSTERING WITH DTW ..... 125

FIGURE 7.1 STRUCTURE OF GENERATOR NETWORK ..... 139

FIGURE 7.2 STRUCTURE OF REWARD NETWORK AND SHAPING TERM ..... 142

FIGURE 7.3 IMPATIENT BEHAVIOUR DECISION MAKING ..... 147

FIGURE 7.4 PATIENT BEHAVIOUR DECISION MAKING ..... 149

FIGURE 8.1 POTENTIAL CONTRIBUTIONS OF RESEARCH FINDINGS ..... 155

FIGURE 8.2 XINJIEKOU COMMERCIAL STREET ..... 168

FIGURE 8.3 PEDESTRIAN DECISION MAKING DURING MORNING TIME ..... 172

FIGURE 8.4 PEDESTRIAN DECISION MAKING DURING NOON TIME ..... 173

FIGURE 8.5 PEDESTRIAN DECISION MAKING DURING EVENING TIME ..... 174

FIGURE 8.6 BUILT ENVIRONMENT FEATURES ANALYSIS DURING MORNING TIME ..... 176

FIGURE 8.7 BUILT ENVIRONMENT FEATURES ANALYSIS DURING NOON TIME ..... 178

FIGURE 8.8 BUILT ENVIRONMENT FEATURES ANALYSIS DURING EVENING TIME ..... 180

# List of Tables

TABLE 1.1 CROWD DISASTERS BY DECADE (YANG, ET AL., 2024) ..... 17

TABLE 1.2 EXAMPLES OF PAST CROWD DISASTERS ..... 17

TABLE 2.1 COMMON ELEMENTS OF CROWD MANAGEMENT PLAN (CHEBIB, ET AL., 2025) ..... 27

TABLE 2.2 DIM-ICE MATRIX (STILL, ET AL., 2020) ..... 29

TABLE 2.3 INDICATORS FOR FIST MODEL (YANG, ET AL., 2024) ..... 44

TABLE 3.1 CROWD MODELLING MODELS ..... 72

TABLE 5.1 ACCURACY OF FOUR CHARACTERISTICS (YANG, ET AL., 2024) ..... 97

TABLE 5.2 NUMBER OF INCIDENTS WITHIN EACH CLUSTER (YANG, ET AL., 2024) ..... 97

TABLE 5.3 TYPES OF CROWD DISASTER (YANG, ET AL., 2024) ..... 104

TABLE 6.1 EXPERIMENT CONSISTENCY ..... 121

TABLE 6.2 BEHAVIOURAL MEASURES OF DIFFERENT CLUSTERS (IN ABSOLUTE VALUE) ..... 126

TABLE 6.3 BEHAVIOURAL CONSISTENCY FOR BOTH TYPE ..... 127

TABLE 7.1 POLICY EVALUATION AND BENCHMARKS COMPARISON ..... 145

TABLE 7.2 SIMULATION RESULTS OF SFM ..... 152

TABLE 8.1 POLICY EVALUATION AND BENCHMARKS COMPARISON ..... 171

TABLE 8.2 SIMULATION RESULTS OF SFM ..... 182

# 1. Introduction

## 1.1 Background

Rapid urbanisation and globalisation, coupled with major developments in transportation that make long-distance travel faster, cheaper, and more accessible, have led to a major increase in the number of large gatherings around the world. Unfortunately, as demonstrated in Table 1.1, previous research (Yang, et al., 2024) has identified a considerable rise in the number of crowd disasters, i.e., incidents occurring during situations involving crowds that can result in fatalities and/or serious injuries.

| Time period | Number of crowd disaster |
|---|---|
| 1990-1999 | 41 |
| 2000-2009 | 114 |
| 2010-2019 | 124 |

*Table 1.1 Crowd disasters by decade (Yang, et al., 2024)*

| **Incident** | **Cause** | **Fatality** |
|---|---|---|
| 1989 Hillsborough disaster (BBC, 2022) | Influx of football fans into standing terrace | 97 |
| 2010 Love parade disaster (Helbing & Mukerji, 2012) | Crowd turbulence | 21 |
| 2014 Shanghai crowd disaster (Xinjiang Television, 2015) | Counterflow at staircase | 36 |
| 2015 Hajj crowd disaster (Almukhtar & Watkins, 2016) | Convergence of two large flows onto a street | 2600+ |
| 2022 Seoul Halloween crowd disaster (Reuters, 2023) | Inadequate preparation | 159 |
| 2022 Brixton crowd disaster (Warren, 2024) | Fans without tickets tried to force into the venue | 2 |

*Table 1.2 Examples of past crowd disasters*




*Figure 1.1 Love Parade in Duisburg, Germany (Müseler, 2010)*

Among the past crowd disasters, the reported cause, and number of recorded fatalities of several prominent crowd disasters are illustrated in Table 1.2. Some of these disasters occurred during major annual events e.g., the Hajj religious gathering, Union of European Football Associations Champions League Final. It is also notable that some events have been cancelled permanently due to crowd disasters e.g., the Shanghai New Year Light Show and Love Parade in Duisburg, Germany, with a scene picture demonstrated in Figure 1.1.

The severity of these disasters has resulted in the critical need for effective crowd management measures. Crowd management itself refers to proactive planning and operational strategies designed to ensure the safety, comfort, and orderly movement of attendees. It involves anticipating the potential risks through scenario planning and preparing accordingly. In particular, the preparation considers expected crowd behaviour in given circumstances and defines measures intended to encourage behaviours that enhance safety outcomes (Martella, et al., 2017). These preparations encompass all stages of event participation, including guiding people into the venue, facilitating their experience during the event, and ensuring their safe dispersal afterward, also known as pre-event, in-event and post-event, respectively (Sharma, et al., 2018).

While currently various analytical techniques, e.g. computer simulation modelling of crowds, are used to assist in crowd management, in recent years advances in machine learning have expanded the capability of crowd modelling, particularly by enabling faster simulation (Liu, et al., 2017). When suitable data are presented, machine learning-based methods could achieve a better accuracy in predicting crowd movement than conventional methods for particular tasks. For example, Social-LSTM (Alahi, et al., 2016) outperforms social force model (Helbing & Molnár, 1995) by roughly 31% in predicting future steps over selected datasets.

If such analytical tools are appropriately applied, they hold the potential to support pre-event planning by enabling planners to simulate alternative crowd scenarios, test different infrastructure layouts, and identify potential congestion points in advance. Such predictive simulations can contribute to a more robust logistical planning of events compared to the past, including route optimisation and emergency evacuation strategies (Sharma, et al., 2018).

During the in-event phase, crowd monitoring technologies, mostly vision-based systems, have significantly advanced with the development of machine learning, providing enhanced situational awareness in real time and overcoming the limited adaptability of conventional methods to complex and changeable scenarios (Zhao, et al., 2024). These systems support the early detection of anomalies such as bottlenecks or sudden shifts in density, enabling timely interventions to maintain flow and ensure public safety (Ameen & Stone, 2023).

However, despite significant advancements in the crowd management technologies, the frequency of crowd disasters remains stubbornly high, indeed with a rising trend. This has been attributed to an inadequate consideration of the human factors (Alasmari, et al., 2024). The current imbalance, with its dominant focus on technology and technical solutions, has been described as a culture of technocracy by Challenger, et al. (2010). Although this concept was proposed over a decade ago, barring a few notable efforts, relatively little research has been conducted to address the neglect of the social aspects in crowd analysis (Drury & Reicher, 2020). Instead, most efforts have continued to focus on refining machine learning based crowd modelling and crowd monitoring techniques. This thesis aims to address this deficiency in a systematic manner following a clearly defined strategy which combines numerous elements, e.g. development of a database of crowd disasters, virtual reality experimentation and the use of inverse reinforcement learning techniques.

## 1.2 Aim and objectives

Given current challenges in crowd management, this research aims to contribute to a more comprehensive understanding of crowd disasters by incorporating social aspects, and to develop methods for analysing crowd movement and behaviour. By achieving this aim, crowds at events can be much better managed than at present, thereby avoiding situations where densities build up to dangerous levels, leading to individuals in such a crowd being either killed or seriously injured. This aim is to be attained through the following research objectives:

1. Define the relevant qualitative characteristics of crowd disasters.

This objective is to be achieved through reviewing relevant literature and conducting a case study on a past crowd disaster. The summarised characteristics aim to reflect the causal mechanism of a crowd disaster, as well as incorporating the effect of the external environment, with the focus on the social aspects based on a socio-technical system perspective.

2. Construct a database of past crowd disasters.

This objective is to be achieved through data collection on the past crowd disasters based on the defined qualitative characteristics of crowd disasters. Due to lack of sufficient and appropriate official reports, the primary data source is based on a review of past news reports, complemented by supplementary sources, e.g. Internet websites. A further pattern analysis is conducted based on the developed database, contributing to a new comprehensive understanding of crowd disasters and thereby laying the foundations for subsequent analyses in the thesis.

3. Assess the feasibility and reliability of using virtual reality (VR) for investigating crowd behaviour in specific scenario.

This objective is to be achieved through the reproduction of a selected scenario in the virtual environment, based on pattern identification of the crowd disaster analysis. The virtual environment implements specific features to reflect unique environment settings. Multiple

rounds of experiments are conducted to evaluate the consistency and reliability of behavioural data obtained in this virtual environment, thereby examining the suitability of VR as a research tool for this specific scenario.

4. Identify movement patterns under specific scenario settings.

This objective is to be achieved through analysing the movement data collected in the VR experiments. Due to the heterogeneity of crowds, different movement strategies are expected to be discovered, providing a deeper insight into understanding behaviours within the selected scenario.

5. Propose an approach to infer decision making behind crowd behaviour across various scenarios.

This objective is to be achieved through a review of the state-of-the-art methods of analysing crowd movement and subsequently to propose a novel approach to quantitatively understand the decision making underlying crowd behaviour. The proposed approach is applied to data collected from the VR experiment and further examined in an additional scenario setting.

6. Propose improvements in the practical crowd management based on the research results.

This objective is to be achieved through discussion of the implications of the research findings, both at the end of each original research chapter, and in the concluding chapter of this thesis. The proposed improvement build upon the current crowd management practices, contributing both to the real-time operational strategies and pre-event planning.

## 1.3 Thesis structure

This thesis is presented in nine chapters, with details of each chapter given in the next subsection, as illustrated in Figure 1.2.

In general, chapter 2 reviews various approaches used to understand and manage crowds, especially in terms of the causes and prevention of crowd disaster. Chapter 3 reviews the

conventional and state-of-art analytical models to analyse crowds. Building on these foundations, chapter 4 summarizes the research philosophy of this study, including the adopted research approach and strategy. Chapter 5 contributes to a new comprehensive understanding of a crowd disaster from a socio-technical system perspective. Under the new comprehensive understanding, chapters 6 and 7 present methods of exploring crowd movement patterns, as well as understanding their decision making. Chapter 8 discusses implications drawn throughout the study and chapter 9 concludes this thesis.

### 1.3.1 Chapter outlines

**Chapter 1** introduces the research topic of this thesis, as well as its background. The aim and objectives are further introduced in detail, together with the structure of this thesis.

**Chapter 2** reviews common approaches in crowd management. The concept of system thinking and socio technical systems is further introduced as a framework for understanding crowds. Based on the socio-technical system perspective, the characteristics of a crowd disaster, the undesirable consequences of failed crowd management, are reviewed, and gaps are identified through a case study relating to the 2014 Shanghai crowd disaster.

**Chapter 3** reviews conventional and state-of-the-art analytical methods used to study crowds. The two most common approaches, based on macroscopic and microscopic models, are discussed, including density estimation, action recognition, crowd counting and crowd modelling. A mesoscopic approach is introduced as a hybrid method, capturing the advantages of both macroscopic and microscopic approaches.

**Chapter 4** presents the methodological foundations of this research, driven by the nature of the research problem. A pragmatic approach is adopted to accommodate both objective system properties and subjective human behaviours. Based on the mesoscopic perspective, the research approach and strategies for subsequent investigations are outlined.

**Chapter 5** presents the construction and analysis of a crowd disaster database. A codebook is designed to encode causal factors and external environment impacts. Cluster analysis identifies nine types of crowd disasters, contributing to a new comprehensive understanding of crowd disaster and providing the empirical foundation for subsequent chapters.

**Chapter 6** investigates the jostle scenario, selected from nine types of crowd disasters identified in Chapter 5, through a virtual reality (VR) experiment. A virtual environment is developed to reproduce the scenario settings, and three rounds of experiments are conducted. Movement patterns are subsequently analysed, and two types of behaviour, namely impatient and patient, are identified through cluster analysis.

**Chapter 7** proposes a decision making inference framework combining inverse reinforcement learning and model interpretability. By retrieving and analysing a reward function, the impact of different factors on behavioural preferences are quantified. Differences between the two behavioural types identified in chapter 6 are quantified and visualised, with further implications discussed.

**Chapter 8** discusses the implications drawn from this study across three domains. Recommendations are made for crowd management strategies, particularly for the jostle scenario. Lessons learned from VR experiment design are summarized, and the scalability of the proposed decision inference framework is further explored using a case study of urban planning in Nanjing, China.

**Chapter 9** concludes this thesis by summarizing the research findings and revisiting research objectives to discuss how they have been achieved. Future works related to the topic of this thesis are also suggested.

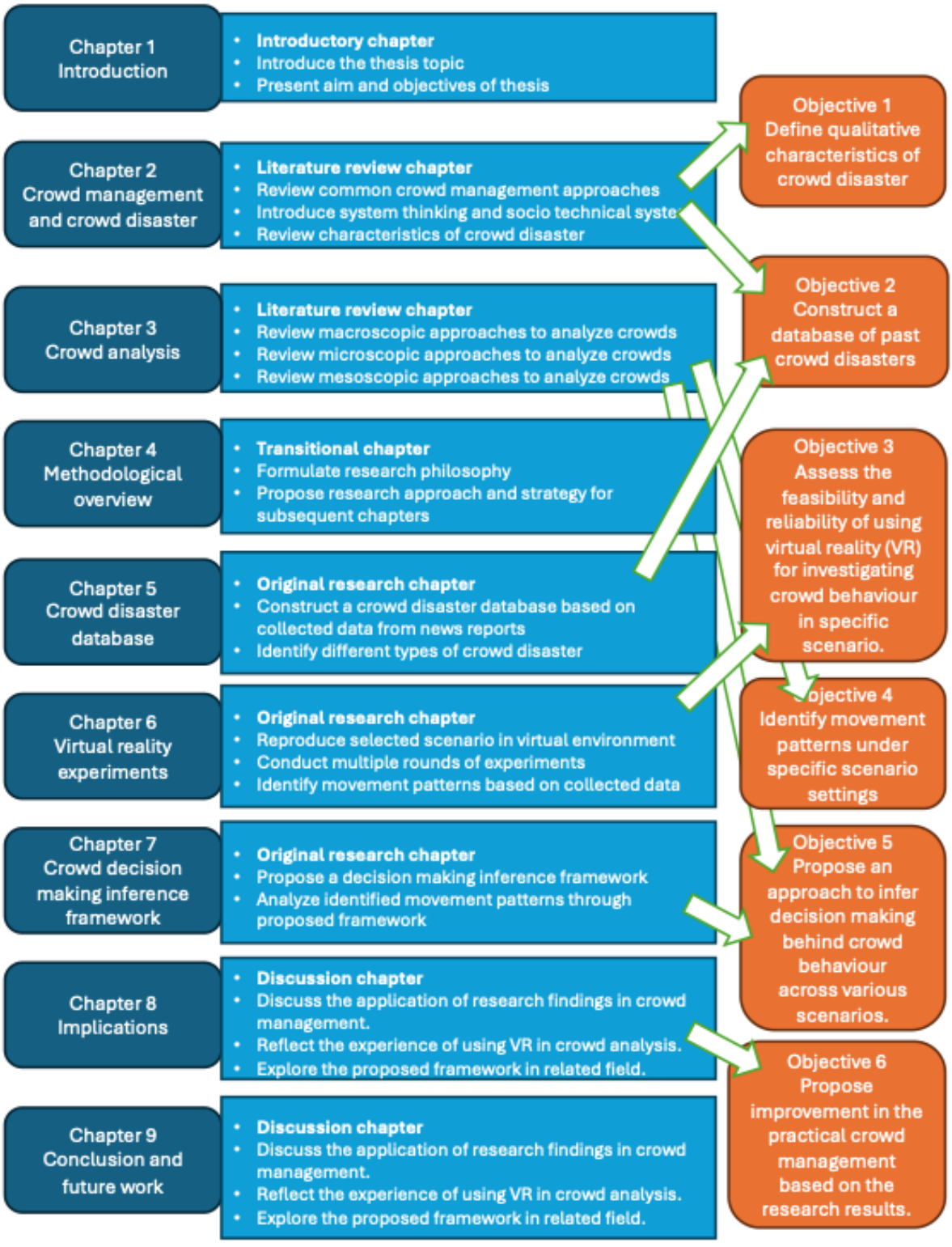



*Figure 1.2 Thesis chapters and corresponding objectives*

# 2. Crowd management and crowd disaster

## 2.1 Introduction

This chapter identifies gaps within current crowd management practices. By reviewing common crowd management approaches during both pre-event and in-event stages, it highlights the importance of understanding crowd behaviour for an efficient implementation of crowd management strategy, especially underrepresentation of social factors in current frameworks. To address this gap, the system thinking approach is introduced as a way of comprehending crowd, extended by the socio-technical system perspective to emphasise the impact of human actors and from social external environment. Characteristics of crowd disaster, as an extreme consequence of failed crowd management, are then reviewed, with a case study of 2014 Shanghai crowd disaster to further validate the importance of adopting social technical system perspective.

Section 2.2 reviews common approaches for crowd management. Section 2.3 introduces socio-technical system perspective as the foundations of understanding crowds in this research. Section 2.4 reviews characteristics of crowd disaster and presents the case study of 2014 Shanghai crowd disaster. Section 2.5 summarizes the chapter.

## 2.2 Crowd management

Crowd management refers to measures taken by authorities to ensure the safety, comfort, and orderly movement of attendees. This typically involves anticipating potential risks through scenario planning, identifying likely crowd behaviours, and preparing accordingly (Martella, et al., 2017).

Sharma, et al. (2018) identified three stages of crowd management including: pre-event, in-event and post-event. Each stage plays a critical role in addressing challenges proactively and effectively. While post-event practices focus on evaluation and learning, most preventive and

operational measures occur during the pre-event and in-event phases. This section therefore reviews the essential practices and methods typically employed in these two phases.

### 2.2.1 Pre-event

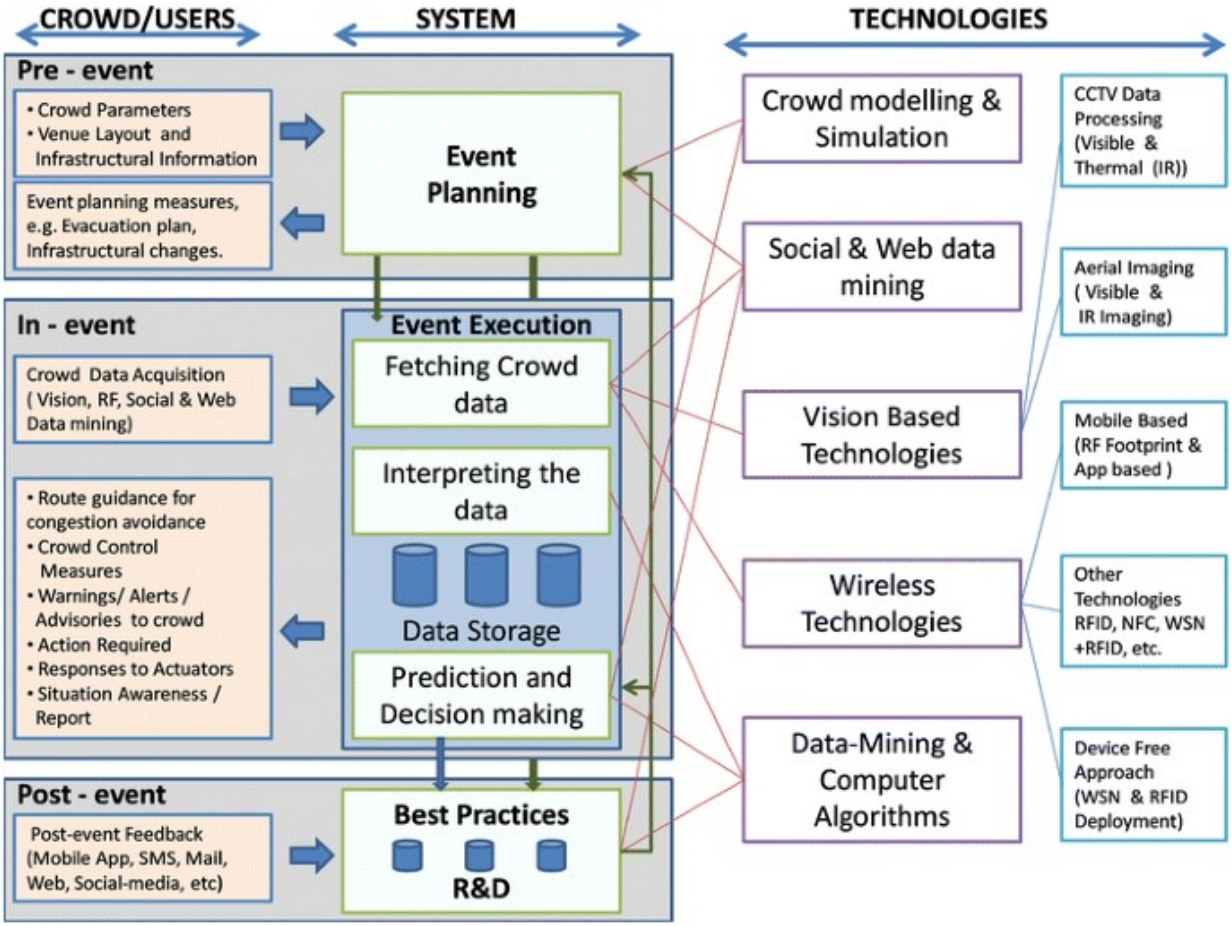



*Figure 2.1 Three stages of crowd management (Sharma, et al., 2018), reproduced with permission from Springer Nature*

As Figure 2.1 highlights, the pre-event phase of crowd management primarily involves planning and preparation stages of an event. A guide as to what this involves is highlighted by the "Safer Crowd, Safer Venues'' (Chebib, et al., 2025) document used by United Kingdom Crowd Management Association (UKCMA), and Table 2.1 summarises the common elements included in a crowd management plan.

| Element | Description |
|---|---|
| Statement of intent | Describes the purpose of the crowd management plan and how it fits in the overall event management. |
| Policies and procedures | Includes the entry policy and the show stop policy. Also included are policies that have direct impact on participant behaviour, e.g. alcohol policy. |
| Roles and responsibilities | Details the duties and expectations for each individual involved in the crowd management |
| Capacity | States the calculated capacity of the event, including the information on tickets and queuing system |
| Site plan | Shows the physical layout of the event venue and surrounding areas. |
| Crowd safety risk analysis | Identifies the potentials risks through a risk analysis model |
| Ingress and egress plans | describes the procedures of people arriving at and departing from the event venue. |
| Security and stewarding plan | outlines a dot plan for the places of the staff at each phase of the event |
| Protective measures | Detail the measures put in place to prevent attacks by a person or a group. |
| Contingency and emergency plans | Include the evacuation plans and show stop or pause plans. Threats to be considered include fire, extreme weather, deviant behaviour, attack and external communication such as shouting. |
| Briefing and debriefs | Consist of briefing documents and debrief forms for staff. |

*Table 2.1 Common elements of crowd management plan (Chebib, et al., 2025)*

Amongst these elements, risk analysis plays a central role in the planning stage, as it inherently involves the evaluation of many other elements, including the venue capacity, site design, ingress and egress planning, as well as the nature of the activities at the event. Hence, the

following section reviews commonly adopted models for conducting risk analysis in crowd management.

#### *2.2.1.1 Risk analysis*

A classical risk analysis model is the DIM-ICE model (Still, et al., 2020), and is widely used in the crowd management planning practice in UK (Chebib, et al., 2025). The acronym DIM represents Design, Information and Management, which are the three primary influences on the crowd behaviour.

**Design** refers to both immutable aspects, such as the physical layout of the venue, as well as mutable aspects including measures implemented by the organisers such as the barriers and the queuing system. Both aspects can significantly affect crowd behaviour throughout an event.

**Information** includes communication with participants and staff. Communication with participants typically aims to set expectations around behaviour, detail prohibited items, and recommend arrival and departure options. It may also include directional signage that guides movement, and dissuasive signage that discourages entry into restricted areas. Both types of signage can be either passive, providing static information, or active, adapting in response to changing conditions or operational requirements (Galea, et al., 2017).

For staff, communication focuses on shift schedules, briefing documents, and operational guidance. Additionally, relevant information from external agencies, such as any anticipated crowd surges or off-site disruptions, is also incorporated to support informed planning and response.

**Management** focuses on the allocation of resources, such as the distribution of staff and volunteers to reinforce crowd management plans and prevent violent behaviour.

| Normal | Ingress | Circulation | Egress |
|---|---|---|---|
| Design | | | |
| Information | | | |
| Management | | | |
| Emergency | Ingress | Circulation | Egress |
| Design | | | |
| Information | | | |
| Management | | | |

*Table 2.2 DIM-ICE matrix (Still, et al., 2020)*

In contrast, ICE stands for three phases of the in-event stage: Ingress, Circulation and Egress. In addition to these phases, the DIM-ICE model incorporates two behavioural modes: Normal and Emergency, to reflect different crowd dynamics. The outcomes of the risk analysis can be presented in a matrix format, as shown in Table 2.2. By completing each cell, DIM-ICE model helps to identify potential risks arising from design, information, or management across the different phases of the event and under two behavioural modes.

An alternative common risk analysis model also adopted by the UKCMA is the RAMP model (Still, 2021), which focuses on the following four key dimensions:

i) **R**outes likely to be used during the event,
ii) **A**reas where crowds are expected to gather,
iii) **M**ovement patterns within and between these areas and
iv) **P**rofile of the crowd, especially its demographic characteristics.

On one hand, the DIM-ICE model identifies potential risks under different behavioural modes, which requires a detailed understanding of interactions between external factors and crowd behaviour. On the other hand, the RAMP model takes a further step by dividing crowd behaviour into route choice, preferred areas and movement patterns, while also incorporating impact from participants' demographics. In summary, common risk analysis models place crowd behaviour at the centre of their frameworks, thus making their efficient implementation heavily dependent on the accuracy of behavioural understanding.

### 2.2.2 In-event

The primary objective of the in-event phase, as illustrated in Figure 2.1, is to monitor crowd states and inform decisions regarding appropriate management and intervention strategies in real time. The three phases of the in-event stage outlined in the DIM-ICE model, each involve specific methods to achieve this objective and detailed as follows.

i) Ingress

The 'Purple Guide', developed by the Events Industry Forum (EIF) (2022) to establish safe standards for outdoor events, notes that ticketing gates and ticket scanning are commonly used during the ingress phase to monitor crowd flow and manage venue capacity. Additionally, trained and competent crowd spotters can be employed to observe crowd movement, with the assistance of hand-held mechanical counters to gauge the numbers of individuals moving in, out or through a particular space. However, the implementation of ticketing has a direct impact on the crowd's behaviour, which requires further consideration, including the types of tickets (Couto, et al., 2015), layout of the ticketing gates (Fu, et al., 2024), etc.

Furthermore, the use of barriers requires further consideration as well. Such barriers can be broadly classified into the following two categories (Events Industry Forum, 2022):



*Figure 2.2 Retractable belt barrier*

a) differentiating spaces – retractable belt (demonstrated in Figure 2.2), ropes or tape are commonly utilized in low-risk area to define queuing areas and waiting lines.



*Figure 2.3 Pressure barrier (4028mdk09, 2010)*

b) pressure barriers (demonstrated in Figure 2.3) - these are used to address the potential risk with crowd surges and crowd action, moving from a designated safe zone into an area where a risk may be present.

The choice of barrier depends on the anticipated crowd behaviour, and beyond safety concerns, qualitative research (Li, et al., 2013) suggests that the configuration of the barrier affects how crowds perceive comfort, safety, and confinement, which in turn is reflected on the crowd behaviour.

ii) Circulation

During the circulation phase, continuous crowd monitoring becomes increasingly critical. Closed circuit television (CCTV) cameras are widely used to provide coverage of specific areas within and around the venue, especially due to the advances of crowd counting techniques (Wong & Law, 2023; Tomar, et al., 2022), which are discussed in the next chapter. Other techniques, such as Wireless Fidelity (Wi-Fi) (Weppner, et al., 2016), Radio Frequency ID (RFID) (Kaiser, et al., 2018) and Unmanned Aerial Vehicles (UAV) (Husman, et al., 2021), are also incorporated to complement CCTV by providing additional information on crowd density, movement patterns, and circulation dynamics.

In addition, suitably trained stewards may be stationed at key internal and external locations in order to monitor crowd flow, activity, and behaviour. All the collected information should be

communicated through a pre-agreed protocol with an event control room or the event organiser directly, for a rapid and informed decision-making in response to observed incidents (Events Industry Forum, 2022).

However, the effectiveness of both methods relies heavily on accurately identifying any early indicators of potential incidents. The earlier such signs are detected, the more time is available for the decision making of interventions, and the greater the likelihood of preventing the incident from escalating.

iii) Egress

The Egress phase often entails the largest scale of movement for any entire event, as all the participants leave the venue simultaneously. To accommodate this, more gates are normally used than during ingress, which in turn makes crowd monitoring more complex, and highlights the importance of ensuring all planned exits are opened, while the routes are clear of obstruction (Chebib, et al., 2025).

More importantly, the mood of participants shifts from that of excitement to impatience, as they seek to get back home as quickly as possible (Chebib, et al., 2025). This behavioural change could often lead to unsteady crowd flow, which may be further exacerbated by velocity variability arising from the diverse composition of the crowd (Larsson, et al., 2021).

In summary, both risk analysis in the pre-event phase and the selection and implementation of crowd management methods during the in-event phase require a deep understanding of crowd behaviour, as highlighted by Challenger, et al. (2010). Despite recent advances in this field, understanding the decision making of crowd behaviour, as well as the impact from social factors are still to be addressed (Haghani, et al., 2023).

## 2.3 Socio-technical system

The previous section has identified the importance of understanding crowd behaviour within crowd management practices. This section looks at the theoretical framework of understanding crowd, especially from a complex system perspective. There is considerable historic literature relating to sociotechnical systems as applied in various fields (Viterale, 2025). It is not the aim of this section to cover the historical context, assumptions, etc. of complex socio-technical systems which have been covered in considerable detail by other authors. Rather the following sections focus on those aspects of socio technical systems relevant to this thesis, beginning with complex system in the following section.

### 2.3.1 Complex System

A complex system, as defined by Newman (2011), is "a system composed of many interacting parts, such that the collective behaviour of those parts together is more than the sum of their individual behaviours."

Crowd, as every accumulation of its parts (each pedestrian) stand in relation with each other and also form a whole in certain respect, could be viewed as a perfect example of a complex system.

A key feature of a complex system is emergence, which is only observable at macroscopic level, arising from collective behaviour from microscopic level (Artime & Domenico, 2022). Furthermore, Boccara (2010) highlights that emergence is (i) difficult to anticipate from the knowledge of the agents' behaviour, and (ii) does not result from the existence of a central controller.

Two emergent phenomena that are widely observed in pedestrian crowds, lane formation and arch formation (Helbing, et al., 2005), are introduced and discussed in the following sections. It is worth noting that other emergent phenomena are also reported in crowd dynamics, such as freezing by heating (Helbing, et al., 2000) and zipper effect (Hoogendoorn & Daamen, 2005).

The former two phenomena are selected here due to their frequent occurrence in everyday crowd scenarios, as well as their relevance to and connection with later research findings.

- Lane formation

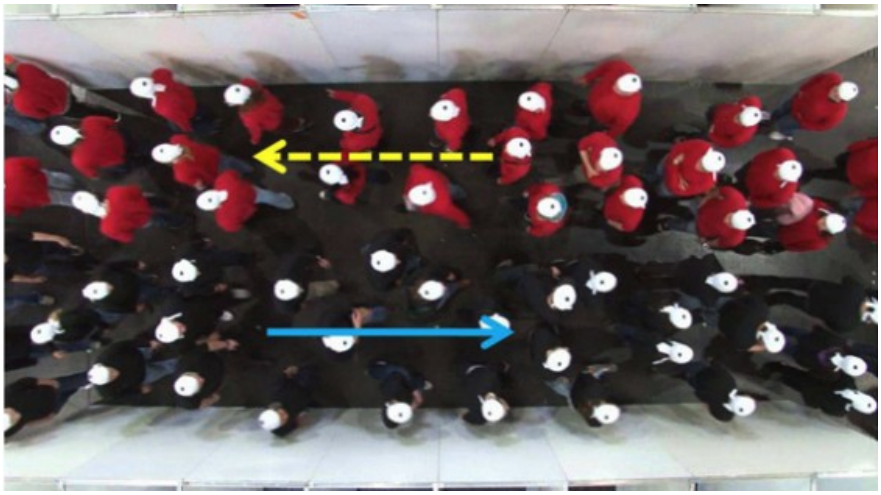


*Figure 2.4 Lane formation (Zhang, et al., 2012), reproduced with permission*

Lane forming phenomena occurs mainly during bidirectional flow and it is described as clustering of pedestrians moving in one direction (Sieben, et al., 2017). A visual illustration of lane formation is presented in Figure 2.4, which is captured from an experiment set up by (Zhang, et al., 2012). The segregation between opposing flows was formed autonomously after the experiment started.

Such phenomenon is assumed to originate from the collision avoidance behaviour of pedestrian, which pedestrians change speed and directions when encountering movement from the opposite direction. As the density increases, less space become available for such manoeuvre. Separating into lanes leads to fewer potential conflicts and thus to fewer forced changes of speed and direction of motion (Sieben, et al., 2017).

- Arch formation

Similarly, arch forming phenomena occurs when crowds trying to pass through a narrow bottleneck under high desired velocity (Helbing, et al., 2000). The term ‘desired velocity’ refers to the velocity that each pedestrian aims to achieve, and is discussed further in the next chapter.

This phenomenon is described as a spontaneous formation of an arch-shaped congestion in front of the bottleneck.

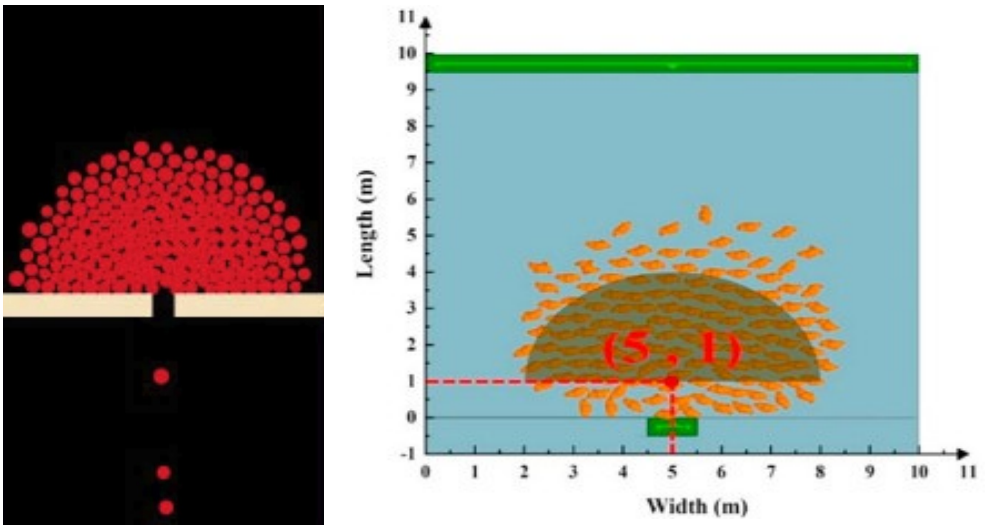



*Figure 2.5 (a) arch formation from computational simulation (Helbing, et al., 2000), reproduced with permission from Springer Nature; (b) arch formation from physical experiment (Ye, et al., 2024), reproduced with permission from Elsevier*

This arch formation is proposed by Helbing, et al. (2000) and further validated through computational simulation. Subsequent research also relied on computational simulation (Li, et al., 2019), due to the potential harm it might cause to participant, but physical experiment was successfully conducted to verify its existence (Ye, et al., 2024). A demonstration of arch formation both based on computational simulation and physical experiment is presented in Figure 2.5.

The underlying cause of the arch formation phenomena is attributed to a 'faster is slower' effect summarised by (Helbing, et al., 2000). This effect describes how pedestrians with high desired velocity tend to behave more aggressively and lead to more collisions and frictional interactions with others, which results in lower actual velocity and finally lead to clogging.

In summary, both lane formation and arch formation originate from simple behaviour rules followed by most crowds, and emerge without specific instructions or a central controller. This highlights the importance of understanding crowd from a complex system perspective. It suggests that not only the rules behind the decision making of each agent needs to be comprehended, but also properties of the crowds as an entire system.

### 2.3.2 System thinking

System thinking, introduced by Haberfellner, et al., (2019), is a structured approach to understand and manage complex system. It identifies two key features of a complex system: elements, the building blocks of the system (referring as parts in previous definition), and relations (also known as interactions), rules that links elements, to the two key features to understand a complex system.

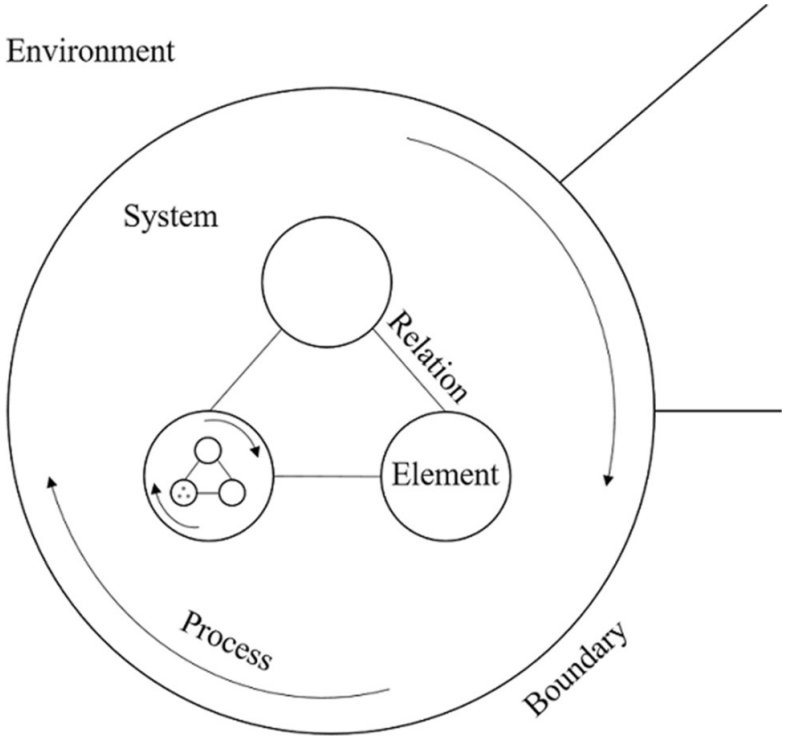



*Figure 2.6 Key features of system thinking (Sotnik, et al., 2021), reproduced with permission from Springer Nature*

Additionally, system thinking recognises the influence of the external environment, which is composed of systems or elements outside the system, that nonetheless affect, or are affected by, the system in significant ways. All these features and their interactions are illustrated in Figure 2.6.

On the other hand, the concept of the socio-technical system, first introduced by Trist and Bamforth (1951) while formalised by Trist (1981), emphasises the critical role of human behaviour and social dynamics. These elements, collectively referred to as the social subsystem, shape system performance alongside the traditional technical subsystem. This perspective

directly aligns with the knowledge gaps identified earlier concerning the limited integration of social factors in crowd management frameworks.

Socio-technical system theory highlights that effective system design and functioning cannot rely solely on technical optimization. It must also account for the intentions, interactions, and experiences of human actors embedded in the system. In doing so, it shifts the perspective from purely mathematical models to those that reflect the complexity of real-world systems composed of both social and technical subsystems.

The incorporation of the external environment, as emphasised in system thinking, has also been recognised as crucial in the design and management of socio-technical systems (Pasmore, et al., 2019). Such incorporation has also been adopted in a socio-technical framework proposed by Challenger, et al. (2010) for analysing crowd event, illustrated in Figure 2.7.

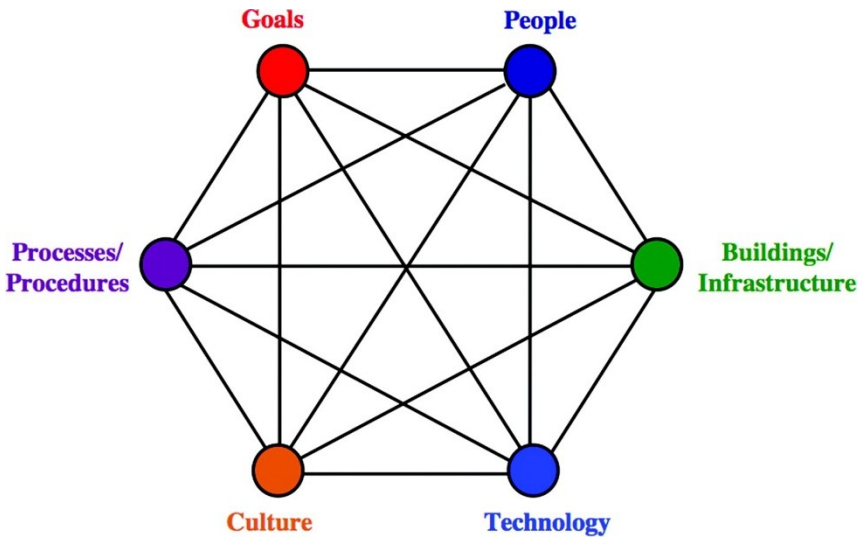



*Figure 2.7 Socio-technical system framework (Challenger, et al., 2010)*

This framework consists of six elements:

- People: Individuals involved, including event staff, emergency services, and attendees, each with their own behaviours and decision making processes.

- Goals: The objectives and priorities of people within the system.

- Processes/Procedures: Established protocols and routines that guide operations and responses during events.

- Technology: Tools and systems used for communication, monitoring, and managing crowds.

- Buildings/Infrastructure: The physical environment, including venue design, gates and barriers, which can influence crowd movement and safety.

- Culture: The shared values, norms, and attitudes within system and among individuals that affect perceptions related to safety and risk.

Among these six elements, Goals and Infrastructure are predominantly determined by the external environment. Goals reflect the social subsystem, encompassing collective intentions, stakeholder priorities, and human motivations. In contrast, Infrastructure represents the technical subsystem, comprising the physical and engineered features that shape how the crowds behave.

As highlighted by Challenger, et al. (2010), a culture of technocracy, which puts the focus heavily on technology and technical solutions, while ignoring the wider, more social aspects, has dominated in the field of crowd management. Although recent studies have emphasised the importance of considering the influences from the social factors, such as those utilising the social identity model (Reicher, 1996), which illustrates how perceived social group identification can facilitate understanding of crowd behaviour (Stott & Radburn, 2020) and emergency management (Drury, et al., 2019), no explicit link has been established between external social environment and system's elements (crowds), nor their relations.

## 2.4 Crowd disaster

Crowd disasters, which this thesis defines as incidents occur during a high density of crowd that can result in fatalities or serious injuries, are often considered as an undesirable consequence for failed crowd management (Fruin, 1993; Helbing & Mukerji, 2012). While

deaths and injuries in crowd disasters can arise from various mechanisms, such as falls followed by trampling (Chang, et al., 2024), high crowd density and the associated compressive effects are generally regarded as the dominant contributors (Fruin, 1993; DeAngeles, et al., 1998; Gill & Landi, 2004). This study therefore defines the primary scope of the analysis accordingly.

Unfortunately, recent years have seen a considerable rise in frequency of such crowd disasters. This underscores the importance of understanding the underlying characteristics and causes of crowd disasters. This section introduces a commonly used model of assessing crowd disasters below.

### 2.4.1 Force, Information, Space and Time (FIST) model

The Force, Information, Space and Time (FIST) model summarises the four primary causal factors of crowd disaster (Fruin, 1993):

i) "F" stands for crowd forces,
ii) "I" stands for the information that crowds act upon,
iii) "S" stands for the physical space where the disasters take place and
iv) "T" denotes the time duration.

This model has been widely used, ranging from understanding the causal factors of one specific incident (Pretorius, et al., 2013) to improving crowd management in confined spaces (Alkhadim, et al., 2018) and public venues (Kui, et al., 2009) by enhancing planning and coordination. Furthermore, it has been listed as a representative risk analysis model as well by the practical guides of UKCMA (Chebib, et al., 2025).

#### *2.4.1.1 Crowd Force (F)*

The main source of fatalities from any crowd disaster is the crowd force (Fruin, 1993; DeAngeles, et al., 1998; Gill & Landi, 2004). According to Helbing, et al. (2007), when the density of people reaches approximately seven persons per square meter, a transition from laminar to longitudinally unstable flows occur "inside" crowds. This phenomenon, known as

crowd turbulence, is defined as an unanticipated and unintended irregular motion of individuals in different directions due to strong and rapidly changing forces in crowds of extreme density (Helbing & Mukerji, 2012).

At extreme density, a crowd acts more like a fluid mass than a solid mass, so that individuals lose control of themselves, and everyone becomes a part of the fluid mass involuntarily. Forces can therefore be propagated through the crowd in order to lift people off of their feet and propel them distances of three or more meters (Helbing & Johansson, 2010). The lungs of individuals in the crowd may be compressed by the forces accumulating due to crowd turbulence, such that individuals cannot breathe sufficiently well in order to acquire the required amount of oxygen (Helbing & Mukerji, 2012). Consequently, compressive asphyxia has been identified as the main cause of death for crowd disasters (Fruin, 1993).

Considerable research (Fruin, 1971; Johansson, et al., 2008; Mori & Tsukaguchi, 1987; Polus, et al., 1983) has suggested that the local average speed of crowds drop sharply when the local density reaches seven persons per square meter, which highlights that people are forced into a restricted "shuffling" gait at this point.

Another form of crowd force is pushing. People unintentionally push each other during a queuing effect and, as the distance between them is short, there is inadvertent body contact (Helbing & Mukerji, 2012). The interactions among people then become physical and leading to injuries, which effectively become non-moving obstacles for others (Helbing, et al., 2002). Such pushing, followed by a "domino effect", can exert a force as high as 4500 N and which is capable of bending steel (Fruin, 1993).

As summarized by the Alkhadim, et al. (2018), the key indications of crowd force contain the following: breathing difficulties, crowd pushing, movement difficulty, crowd pressure (from high density), uncontrollable pushing, and suffocation.

#### *2.4.1.2 Physical space (S)*

The degree of crowding is sometimes restricted by the physical space ("S" element) of assembly facilities. For instance, the safe static crowd density has been identified to be below

five persons per square meter (Still, 2000), while the safe density for walking downstairs and upstairs is defined to be 3.5 and 3.2 persons per square meter, respectively (Ando, et al., 1988). Some physical facilities may have a lower safe density.

Meanwhile, the safe density is affected by the desired velocity of pedestrian. When people try to pass through a bottleneck with a high desired velocity (i.e., in a rushing situation), an irregular succession of arch-like blockings is formed at the entrance of the bottleneck. Friction between people is then amplified due to impatience resulting in pushing, and avalanche-like bunches of pedestrians leaving is found when the arch-like blocking breaks (Helbing, et al., 2000).

Collectively, these mechanisms indicate that crowd safety risks are closely associated with specific physical features where density accumulates or movement is constrained. In conclusion, the key indicators of physical space are the: existence of staircase, escalators and lifts; density at entrances and exits; densities at other physical facilities (Alkhadim, et al., 2018).

#### *2.4.1.3 Information (I)*

Fruin (1993) identified that information ("I" element) crowds act upon to be a major cause of crowd disasters. Perceived information has numerous forms, ranging from any visual information received by an individual's retina and their brain (Moussaïd, et al., 2011) to real time communication (e.g., announcement) from the event organiser. Previous research (Alkhadim, et al., 2018) focused on any information given by the event organiser, such as all types of instruction signs and warning signs. However, such information is hard to achieve without knowledge of the venue configuration or access to the official incident report. On the other hand, information can be further exchanged within crowds, forming social influences based on different social identities (Lovreglio, et al., 2016). As this research does not explicitly model individual social identities, information exchange within crowds is not considered within the scope of this study.

This research is interested in the rushing situation stated above. The term "panic" is commonly used by the media report to describe such situation. The panic situation, as defined by (Kelley, et al., 1965), is when people "are suddenly confronted with an imminent danger and provided

with a means of escape that is limited and for which they do not possess any prior organized mode of use." People tend to show herding behaviour (i.e., follow the actions of others) under a panic situation (Quarantelli, 1957) and thus alternative exits are often inefficiently used (Elliott & Smith, 1993; Helbing & Johansson, 2010; Sharbini & Bade, 2009). Such inefficient usage of emergency exits leads to both overcrowding and intentional pushing as people try to move faster than normal during panic situation (Helbing, et al., 2002).

However, the term panic only accounts for a small proportion of the incidents (Shipman & Majumdar, 2018), and lacks a clear and agreed-upon definition within the literature (Haghani, et al., 2019). Under many conditions, its use is inaccurate and not supported by empirical evidence. Additional indicators are therefore required to present other types of information that might lead to a high desired velocity. Reviewing numerous incidents, results in identifying information leading to a rushing situation including an unexpected incident (e.g., sudden rain) and limited desirables (e.g., charity distribution, discounted product). Combined with imminent danger, these are the key indicators of the information element.

#### *2.4.1.4 Time (T)*

Time factor, identified in the FIST model, relates mainly to the control of the pedestrian demand rate so that any traffic flow does not exceed the capacity of any element of the physical environment (especially the queuing area).

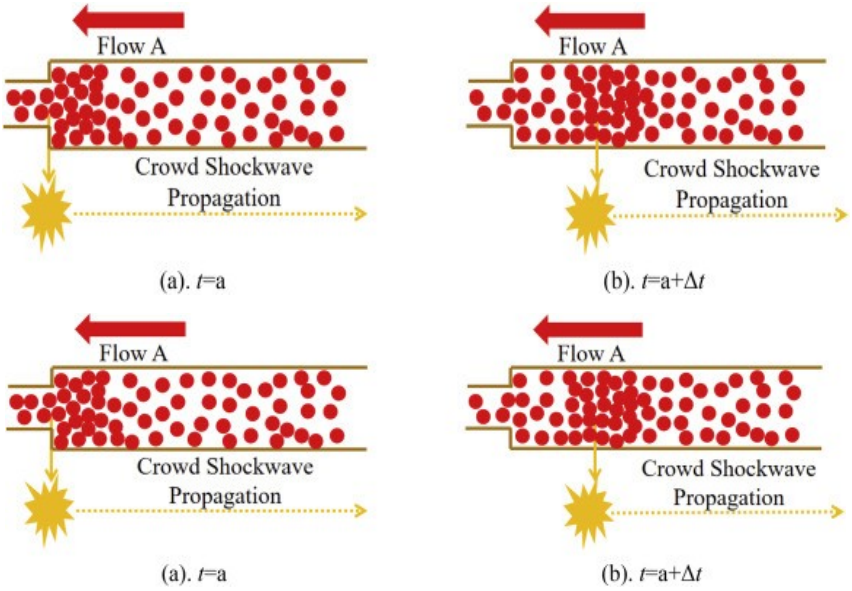



*Figure 2.8 Shock wave propagation along the queue (Wang, et al., 2019), reproduced with permission from Elsevier*

Crowd behaviour within a traffic queue is commonly characterised by a shock wave in traffic flow theory (Lighthill & Whitham, 1955). Shock waves originate from a concept in physics, which describes a sharp change in pressure, temperature, and density moving through a medium (López-Marín, et al., 2018). In the context of crowd dynamics, this concept has been adapted to describe the propagation of density changes, which typically moves backward through a queue (Wang, et al., 2019). An illustration of this density propagation is presented in Figure 2.8.

These shock waves arise due to fluctuations in movement, whereby individuals react to the movement of people ahead of them by adjusting their speed. Consequently, a movement disturbance, usually in the form of sudden acceleration or deceleration, propagates backward through the queue. The longer in duration that these shock waves are sustained, the greater the risk of a crowd disaster (Wang, et al., 2019).

A long waiting time though, can alter crowd behaviour, attributed to the so-called queuing effect identified by (Helbing, 1992). When a queue has been static or moving slower than expected, some of the waiting individuals begin to move forward a little thereby creating an illusion of progression, causing their successor to perform the same behaviour. Such behaviour unconsciously increases people density and has been demonstrated as one of the contributing factors for crowd disasters, such as the 2010 Love parade (Helbing & Mukerji, 2012).

While queue formation may also arise from other mechanisms, such as merging flows (Zhang, et al., 2013) and bottleneck formation (Kretz, et al., 2006), these processes primarily reflect spatial capacity constraints rather than prolonged waiting conditions. As such, although they contribute to density build-up, they are conceptually distinct from the time-related effects emphasised in the FIST model and are therefore not included as indicators of the time factor in this study.

#### *2.4.1.5 Key indicators*

As summarised by Alkhadim, et al. (2018), the key indicators of time are both crowd flows control and waiting time. These indicators, along with those identified for force, information, and space, provide a practical and observable approach for identifying the causal factors in crowd disasters. Collectively, they offer a structured and accessible framework for evaluating

risk factors during a crowd event analysis. A summary of these indicators is shown in Table 2.3 and serves as the basis for the subsequent stages of this research.

| Cause | Indicator |
|---|---|
| F (Crowd Force) | Breathing difficulties |
| | Crowd pushing |
| | Movement difficulty |
| | Uncontrollable pushing |
| | Suffocation |
| | Crowd pressure |
| I (Information) | Imminent danger |
| | Unexpected incident |
| | limited desirables |
| S (Physical Space) | Staircase, escalators and lifts |
| | Density at entrances and exits |
| | Densities at other physical facilities |
| T (Time) | Crowd flows control and Waiting time |

*Table 2.3 Indicators for FIST model (Yang, et al., 2024)*

### 2.4.2 Case study: The 2014 Shanghai crowd disaster

On 31 December 2014, a crowd disaster occurred near Chenyi Square on the Waitan area in Shanghai city. Nearly 300,000 people (mainly tourists) had gathered at the location for New Year celebrations to view a fireworks display. A disaster ensued at a staircase for access to the viewing platform, when spectators going downstairs collided with those going up to get a better view from the platform. In total, 36 people were reported to have died with 49 injured.

After the incident, an official report (Xinjiang Television, 2015) was published, outlining the results of an investigation by Shanghai City Government that identified several causes, and the translated version is attached in Appendix I. The report contains valuable official information about the crowd disaster, which is typically unavailable for most such incidents or accessible only through media reports (Bell, et al., 2018). Consequently, the 2014 Shanghai crowd disaster has been widely used as a case study in order to provide valuable insights into a crowd disaster

for either the pre-event stage (Dong, et al., 2017) or the in-event crowd aggregation detection (Duan, et al., 2020).

The socio-technical system framework for the crowd disaster introduced in section 2.3.2 is utilised here for a case study, and each cause identified in the report (Xinjiang Television, 2015) of the 2014 Shanghai crowd disaster has been categorized into one of the components of the FIST model. The result of case study is summarised below.

- Culture: "Reluctance to make a decision" has long been a feature of China's local bureaucratic culture, since the responsibility of making an incorrect judgement often falls on individuals who can subsequently be punished. Huangpu district government made an application on 13 November 2014 to Shanghai City Government to change the venue of New Year countdown event. The latter then clarified that the district government was responsible for its own decision, thus it took until 9 December 2014 for the district government to decide on the new venue of the event which only received official approval on 26 December 2014. Consequently, the new venue was officially announced to the public on the morning of 30 December 2014, just a day prior to the event.

- People: Most attendees were tourists from other cities, unfamiliar with the local area and with limited access to local information, which influenced their decision making (Labhiri, et al., 2024). In addition, the name of the new venue, Waitanyuan, is very similar to the name of the original, Waitan.

- Goal: The main goal of attendees was to watch the lightshow scheduled to commence at 23:30. Hence increasing numbers of people were trying to reach the viewing platform in order to get a good viewing spot, rather than leave, around that time.

- Process: Failure to predict the location of gatherings. The late announcement of the venue location, coupled with the tourists' lack of familiarity with the local area, meant that most attendees gathered in Waitan, in contrast to the main police force which had gathered at Waitanyuan. Furthermore, because the goal of attendees was not taken into account, the gathering at the viewing platform was not foreseen by the authorities.

Indeed, only seven police personnel had been allocated to the viewing platform prior to the event.

- Technology: Due to the reasons outlined above, inadequate technology was used. The viewing platform is supposed to be unidirectional, with one staircase going up and the other down. However, only warning tape had been used to enforce this unidirectional system. Due to the unequal inflow and outflow, people started to rush to the exit staircase. The warning tape was easily torn down by this strong pedestrian flow and the unidirectional system started to fail. At 23:23, the inflow and outflow clashed and formed a stalemate. Soon afterwards, one person in the stalemate fell and caused a massive domino effect in the crowd (Crowd pressure identified).

- Infrastructure: The Waitan area is located near the Huangpu River, which naturally makes this a restricted area i.e., only three directions are accessible for pedestrians and there are fewer escape options during emergencies (Insufficient physical space identified).

  Based on (Ando, et al., 1988), pedestrian flow on a staircase is more unstable than that on a flat surface. There are several staircases in Chenyi square, and the local authority failed to consider this when performing their safety checks. At 21:14, they reported 50% of the maximum capacity, which based on their experience, was approximately 160,000 people. However, subsequent video footage analysis estimates that the maximum capacity should have been approximately 240,000.

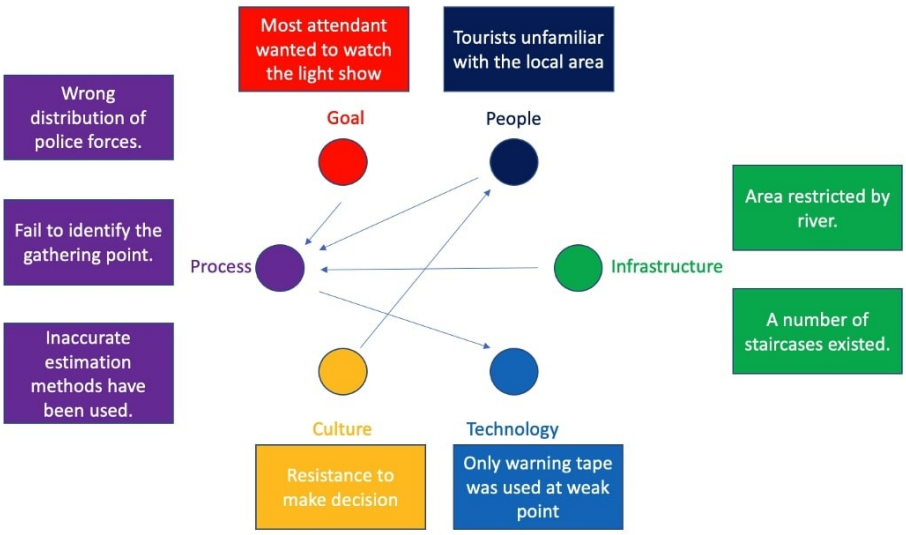



*Figure 2.9 Shanghai 2014 Crowd disaster from a system perspective (Yang, et al., 2024)*

The interactions of these components are illustrated in Figure 2.9, where three components contributed to a failure in the process and also led to an insufficient usage of crowd management measures. Within the three contributing components, only Infrastructure can be partially captured by FIST model (staircase), while Goal and People need to be further understood by other characteristics.

As discussed earlier, Goal primarily reflects the social aspect of the external environment. Its inadequate representation in any representative risk analysis tool further validates the existence of a technocracy in current practices and underscores the importance of understanding People's behaviour under different Goal settings.

## 2.5 Summary

This chapter has reviewed the key approaches and theoretical foundations relevant to crowd management and crowd disasters. Crowd management practices are reviewed across the pre-event and in-event phases, highlighting how effective implementation relies heavily on a deep understanding of crowd behaviour. Despite the advancements in this field, an identified limitation remains in the understanding of the decision making processes behind individual and the influence of social factors.

On the other hand, the chapter has explored the understanding of crowd from a complex system perspective, with two key emergent phenomena, lane formation and arch formation, discussed. Building on this, the socio-technical systems perspective was introduced to emphasise the interplay between social and technical subsystems in managing crowds.

Finally, this chapter has reviewed the characteristics of crowd disaster through the FIST model. A case study of the 2014 Shanghai crowd disaster demonstrates the importance of incorporating socio-technical system perspective, especially in considering the impact from social external environment.

In summary, this chapter has identified a key gap in the crowd managing field: the insufficient integration of social factors. The socio-technical systems approach has been proposed as a promising solution to address this gap and serves as the conceptual foundation for the subsequent chapters.

# 3. Crowd analysis

## 3.1 Introduction

While the previous chapter examines qualitative approaches to understanding crowd behaviour, this chapter reviews existing analytical approaches used to study crowd and build a predictive and interpretable understanding. In the field of crowd analysis, the two most common approaches employed are (Thida, et al., 2013):

- "bottom-up" approach that focuses each individual within the crowd from a microscopic perspective.
- "top-down" approach that considers crowd as an entire entity from a macroscopic perspective.

Such a dichotomy has further facilitated the taxonomy of two major branches in crowd analysis : crowd statistics, which focuses on quantifying aspects of crowds such as size and density, and crowded scene analysis, which aims to understand and interpret crowd behaviours and movement patterns within a scene (Bendali-Braham, et al., 2021; Aldayri & Albattah, 2022).

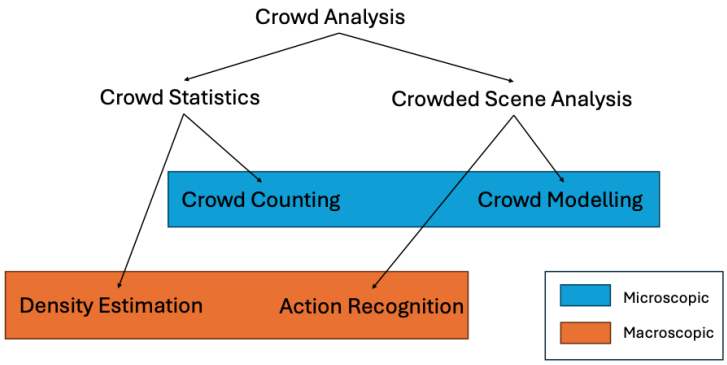



*Figure 3.1 Taxonomy of crowd analysis*

As shown in Figure 3.1, both branches can be further split into two sub-branches following the dichotomy introduced above. Crowd statistics entails crowd counting that considers each

individual (Chan, et al., 2008), and density estimation that views a crowd as an entity (Sindagi & Patel, 2017).

Similarly, the two sub-branches of crowded scene analysis are action recognition and crowd modelling. In order to further understand the movement of crowd in a scene, crowd modelling primarily examines the quantitative quality (e.g., velocity, position) of each individual (Lamba & Nain, 2020), while action recognition aims to classify crowds according to different types of behaviour (Yan, et al., 2019).

Accordingly, this chapter reviews representative models commonly employed across four sub-branches of crowd analysis following the dichotomy introduced above. In section 3.2, macroscopic approaches including density estimation and action recognition are viewed. In section 3.3, microscopic approaches, also known as agent-based approaches, are reviewed, including crowd counting and a specific emphasis on crowd modelling. In section 3.4, the idea of mesoscopic approach is introduced, together with a representative model. Section 3.5 summarizes this chapter.

Note that notation is defined locally in this chapter, with each sub-section using its own notation. Due to the diverse range of methodologies reviewed, some symbols are used with different meanings across models. While introducing entirely new notation for each methodology would be impractical and potentially counterproductive, established notations that are widely recognised in the literature have been retained for clarity and familiarity.

## 3.2 Macroscopic approach

Macroscopic models focus on the local average quantities, such as density and velocity and do not describe individual pedestrian's behaviour (Elaiw, et al., 2019). While multiple macroscopic variables can be derived from crowd statistics, density estimation is discussed here as a prominent instance, as density has been identified as the main source of injury and death during crowd disasters (Fruin, 1993; DeAngeles, et al., 1998; Gill & Landi, 2004). Consequently, estimating density is essential for crowd management to detect overcrowding (Bendali-Braham, et al., 2021).

### 3.2.1 Density estimation

The density $\rho$ of a pedestrian flow can be characterised by speed $V$ and flow rate $Q$, i.e., number of pedestrians $P$ passing through a certain place. Their relationship is captured through the equation:

$$\rho = \frac{Q}{V} \tag{3.1}$$

Which is the Fundamental Diagram of pedestrian flow.

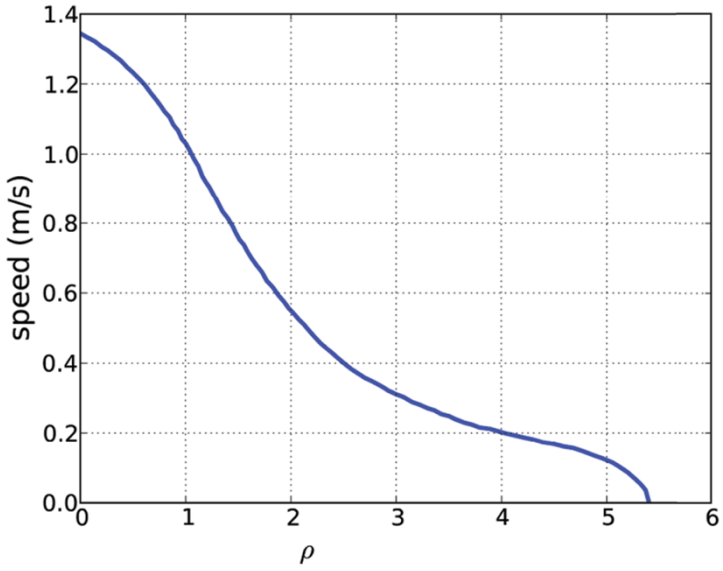



*Figure 3.2 Fundamental Diagram of pedestrian flow (Narang, et al., 2015)*

Empirical studies were conducted to verify the existence of such relationship, with results presented in Figure 3.2. It can be deduced from the graph that as density $\rho$ rises, mean speed $V$ decreases, almost linearly. This trend explains the existence of extreme density discussed in section 2.4.1, where pedestrians are forced into a restricted "shuffling" gait at this point.

Equation 3.1 is often used as basis for the dimensioning and design of pedestrian facilities, to facilitate safety and evacuation studies (Helbing, et al., 2007). However, there are challenges in its application to crowd management. Empirical studies have shown that different infrastructures yield varying fundamental diagrams (Vanumu, et al., 2017), and cultural differences also influence pedestrian flow characteristics (CY & WHK, 1998), as well as by

population characteristics, such as elderly (Ren, et al., 2019) and children (Fang, et al., 2019). In conclusion, the fundamental diagram fails to account for at least two key components of socio-technical systems: infrastructure and culture.

In recent research, density estimation is often performed together with crowd counting (Bendali-Braham, et al., 2021), which is discussed in next section.

### 3.2.2 Action recognition

Action recognition in crowd analysis can be broadly categorised into two approaches: local, interest-point-based methods and global, frame-based methods (Hassner, et al., 2012). Local approaches focus on behaviour around space-time key-points, distinct positions in a video where there is significant change in motion or appearance. Descriptive features are extracted at these key-points to capture localised motion patterns. In contrast, global approaches examine broader frame-level patterns, frequently incorporating crowd counting to inform action recognition in dense environments.

In this research, global approach is adopted in order to capture the gradual evolution of crowd behaviour across the entire timeframe. According to T & HR (2024), advanced methods in action recognition can be broadly categorised into supervised and unsupervised approaches.

#### *3.2.2.1 Supervised approach*

In supervised approach, each data instance is annotated with an action label, and a model is trained to learn the mapping between the input data and its corresponding label (T & HR, 2024). Most input data used in action recognition consist of video footage (Bendali-Braham, et al., 2021), and the specific set of action labels varies across studies depending on their objectives and context.

Action labels are commonly based on mode of locomotion, including activities such as walking, standing, turning, running and sitting (Roggen, et al., 2011). In many cases, an additional layer of labels is added based on the objective of the research. For example, Lin, et al. (2023) divided

each mode of locomotion into alone and together, reflecting the presence or absence of group behaviour.

Furthermore, contextual labels of actions are also included to capture interactions within the scene. As demonstrated in Ehsanpour, et al. (2021), human-object interactions are explicitly annotated, including “talking on the phone”, “reading” and “eating”.

Beyond predefined action categories, another widely used technique in crowd management is anomaly detection (Alasmari, et al., 2024). The term anomaly originally refers to irregular patterns and in the context of crowd analysis, can either be defined qualitatively as behaviour that does not correspond to expectations in crowd analysis (Aldayri & Albattah, 2022), or quantitatively as a rare event that occurs infrequently within large-scale trajectory data (Corbetta, et al., 2017). Under supervised approach, anomaly behaviour is annotated against “normal” behaviour (Direkoglu, 2020).

Supervised approach, however, is heavily dependent on annotation, and the process of labelling is often labour-intensive and time-consuming (Zhao, et al., 2024). Consequently, weak-supervised and semi-supervised approaches are proposed, in which only a portion of data is annotated (T & HR, 2024). Action labels are often influenced by the annotator’s judgement, which may introduce human errors (Demrozi, et al., 2023). The inherent complexity and variability of human behaviour further complicate the process of fitting actions into discrete categories (Sedaghati, et al., 2025).

#### *3.2.2.2 Unsupervised approach*

In contrast to supervised approach, unsupervised approaches can mitigate some of above challenges by learning patterns directly from the data without relying on predefined labels (Vahdani & Tian, 2023). In crowd analysis, unsupervised approach can be employed to discover hidden behavioural patterns, as well as extract abstract representations of collective intent (Titus, et al., 2021).

*Cluster analysis*

Cluster analysis is a classical unsupervised learning technique that aims to "divide a supposedly heterogeneous set of data into homogeneous clusters" (Everitt, 1972, p. 143). It operates on a dataset of $n$ observations $\chi = \overrightarrow{x_1}, \overrightarrow{x_2}, \dots, \overrightarrow{x_n}$, where each observation $\vec{x}$ is a vector composed of features by selection. Clustering methods are commonly categorized into two main types: typically includes partitional clustering and hierarchical clustering (Pitafi, et al., 2023).

- Partitional clustering

Partitional clustering aims to produce a partition $C = C_1, C_2, \dots, C_K$, which divides $\chi$ into $k$ non-overlapping clusters, where $k$ is specified in advance. The objective is to minimize total intra-cluster dissimilarity, typically formulated as:

$$\sum_{k=1}^{k} \sum_{\overrightarrow{x_n} \in C_K} \phi(\overrightarrow{x_n}, C_K) \tag{3.2}$$

where $\phi(\overrightarrow{x_n}, C_K)$ is a cost function to measure dissimilarity between data point $\overrightarrow{x_n}$ and cluster $C_K$ (Jain, et al., 1999).

One classical example of the cost function is the squared error:

$$\|\overrightarrow{x_n} - Z_K\|^2 \tag{3.3}$$

Where $Z_K$ is the centroid of cluster $C_K$. It is used in the widely known K-means algorithm (MacQueen, 1967).

- Hierarchical clustering

Similar to crowd analysis, hierarchical clustering has two common approaches: bottom-up, named as agglomerative clustering, and top-down, named as divisive clustering (Pitafi, et al., 2023).

Agglomerative clustering begins by treating each data point $\overrightarrow{x_n}$ as an individual cluster. At each iteration, the two clusters with the smallest dissimilarity, measured by a chosen cost function, are merged. This process continues until all points are combined into a single cluster representing the entire dataset (Patel, et al., 2015).

The divisive approach, in contrast, employs the top-down method, beginning with the entire dataset grouped as a single cluster. The algorithm then recursively splits this cluster into smaller subsets. To determine the optimal division, the method typically evaluates all possible bipartitions, ensuring that each resulting subset contains at least one data point. The process continues until the predetermined number of clusters $k$ is reached (Pitafi, et al., 2023).

In the context of crowd behaviour analysis, both partitional and hierarchical clustering techniques have proven effective, particularly in identifying distinct modes of locomotion (Roggen, et al., 2011; Yang, et al., 2018). Additionally, cluster analysis can reveal complex, hidden patterns embedded within the data that may not be apparent through direct observation (Zhou, et al., 2020).

Beyond action recognition, clustering methods have also proven effective in other key applications, including crowd counting (Rong & Li, 2020), anomaly detection (Yuan, et al., 2015), as well as identifying behavioural patterns within crowd disasters (Lu, et al., 2021).

In summary, cluster analysis serves as an efficient tool for discovering patterns in crowd behaviour, enabling both action recognition and broader situational understanding in complex environments.

## 3.3 Microscopic (agent-based) approach

Microscopic approach, in contrast, focuses on modelling the state and behaviour of each individual pedestrian, capturing interactions and decision making at the agent level (Elaiw, et al., 2019). Consequently, it is often referred as agent-based approach (Bendali-Braham, et al., 2021), where the term agent refers to a discrete entity within a simulated environment that acts according to predefined behavioural rules and interacts with its surroundings (Macal & North, 2010).

### 3.3.1 Crowd counting

Crowd counting determines the total number of individuals present within a given scene (Wang, et al., 2025). The estimated number of people, also known as the crowd size, is an indispensable metric for crowd management, especially identifying the existence of extreme density discussed above (Fiandeiro, et al., 2023).

Crowd counting can be performed using a range of sensors, including ultrasonic sensor and infrared sensor (Drăgulinescu, et al., 2018), as well as radio devices (Nguyen, et al., 2024), with cameras remaining the most commonly employed and dominant type (Irfan, et al., 2016).

Based on the visual inputs from camera, crowd counting methods are generally categorized into three types: detection-based approaches, regression-based approaches, and density estimation-based approaches (Wang, et al., 2025).

*Detection-based approaches*

Early crowd counting methods primarily focused on detecting individual pedestrians in images or video frames. These approaches rely on specific visual features to support classification, and are typically categorised as either monolithic or parts-based detection methods (Fiandeiro, et al., 2023).

Monolithic detection methods are based on holistic pedestrian features (Wang, et al., 2025), such as:

- Haar wavelets, which transform data into a representation that highlights differences at different scales (Viola & Jones, 2004).

- Histogram of Oriented Gradients (HOG), which describes appearance and shape of each individual within an image by the distribution of intensity gradients or edge directions (Dalal & Triggs, 2005).

- edgelets, which are small edge fragments (short line segments) used to describe object boundaries (Bhuvaneswari & Rauf, 2009).

A classifiers algorithm is then used based on selected features to identify individuals within the scene, and then the crowd size is achieved. Such approach, while effective in low-density settings, tend to perform poorly in high-density environments due to low-resolution images, intense crowding, foreground obstruction, and difficulties judging perception (Fiandeiro, et al., 2023).

To address these limitations encountered in high-density situation, parts-based detection methods were introduced, focusing on detecting particular body parts, most notably heads and shoulders, to estimate the number of individuals in a scene. Compared to global features, local features often yield better performance, as detecting local areas is easier than whole pedestrians. However, their effectiveness also diminishes in dense crowds where body parts, especially contours of heads, are blurred or indistinguishable (Wang, et al., 2025).

*Regression-based approaches*

To tackle with challenges encountered during high-density situation, regression-based approaches were proposed. Instead of identifying each individual within the scene, this method adopts a more direct approach by constructing a regression model that capture the mapping between selected visual features and the number of individuals (Wang, et al., 2025).

However, this method poses limited spatial generalization, which means it can only process images captured from a fixed perspective, as spatial information is not taken into account (Fiandeiro, et al., 2023).

*Density estimation-based approaches*

The concept of density map was introduced by Lempitsky & Zisserman (2010) to address limited spatial generalization. A density map is defined as a function of pixels in an image that estimates the local density at each location. The total number of individuals in the scene can be estimated by integrating the local density across the entire image.

Building on this concept, density estimation-based approaches reformulate crowd counting as a regression problem over spatial density distributions. Similar to regression-based approaches, they bypass the challenges of detecting each individual by learning the mapping between local visual features and corresponding density values. Furthermore, spatial correlations are also captured and thus significantly simplify the counting task. This shift from discrete detection to continuous density estimation provides a more scalable and robust framework, particularly for analysing densely populated scenes (Wang, et al., 2025).

Beyond providing better counting performance (Rodriguez, et al., 2011), density maps also offer valuable insights into density estimation, which are critical for crowd management as discussed above (Bendali-Braham, et al., 2021).

In summary, crowd counting has been progressed steadily with better accuracy achieved. This technique has been crucial especially for the monitoring period of crowded events (Lamba & Nain, 2017).

### 3.3.2 Crowd Modelling

Crowd modelling focuses on understanding the movement trajectories of individuals within a scene. It involves tracking motion patterns over time, analysing spatial-temporal behaviours, and making predictions about future positions or crowd dynamics. This information is crucial for identifying unusual movement, forecasting congestion, and supporting proactive decision making in crowd management (Zhang, et al., 2024).

#### *3.3.2.1 Social Force Model*

Social forces, or social fields is a concept proposed by Lewin (1951) to describe the dynamic interaction between individuals, the surrounding crowd, and the external environment. This concept was later adopted by Helbing & Molnár (1995) in the development of the well-known Social Force Model (SFM).

The term social force is derived from the concept of physical force. Although it does not directly exert on human body, social force represents the effect of external environment, as well as surrounding crowds, on the behaviour of agent $\alpha$.

In SFM, each agent $\alpha$ is assumed to have a desired destination $\overrightarrow{r_\alpha^k}$ and a desired velocity $\overrightarrow{v_\alpha^0}$. Both properties describe the state agent $\alpha$ aims to achieve. It is further assumed that agent $\alpha$ always attempts to reach desired destination $\overrightarrow{r_\alpha^k}$ in the shortest possible way, defined by the desired direction $\overrightarrow{e_\alpha}$.

The total force $\overrightarrow{F_\alpha}(t)$ is then defined as the sum of all influences acting on agent $\alpha$. $\overrightarrow{F_\alpha}(t)$ induces in a temporal change $\frac{d\overrightarrow{\omega_\alpha}}{dt}$ of the preferred velocity $\overrightarrow{\omega_\alpha}$, which is the velocity agent $\alpha$ intends to take at the current moment.

Formally, the Social Force Model is expressed as:

$$\frac{d\overrightarrow{\omega_\alpha}}{dt} = \overrightarrow{F_\alpha}(t) + fluctuations \tag{3.4}$$

Where $fluctuations$ are used to consider random variations of the behaviour, and $\overrightarrow{F_\alpha}(t)$ is defined as:

$$\overrightarrow{F_\alpha}(t) = \vec{F}_\alpha^0\left(\overrightarrow{v_\alpha}, \overrightarrow{v_\alpha^0}\right) + A_\beta \sum_\beta \overrightarrow{f_{\alpha\beta}}(\overrightarrow{r_{\alpha\beta}}) + A_B \sum_B \overrightarrow{F_{\alpha B}}(\overrightarrow{r_{\alpha B}}) + A_\iota \sum_\iota \overrightarrow{F_{\alpha\iota}}(\overrightarrow{e_\alpha}, \overrightarrow{r_{\alpha\iota}}, t) \tag{3.5}$$

Among all the forces described above, two primary types can be identified: attractive forces, which drive agent $\alpha$ to move closer to a target, and repulsive forces, which encourage agent $\alpha$ to stay away from a target. Each force component is scaled by a corresponding strength factor $A$, which reflects the influence of that specific type of interaction to agent $\alpha$.

$\overrightarrow{F_\alpha^0}$ is an attractive force that reflects the intent for agent $\alpha$ to reach the desired destination with desired velocity $\overrightarrow{v_\alpha^0}$:

$$\overrightarrow{F_\alpha^0} = \frac{1}{\tau_\alpha}\left(\overrightarrow{v_\alpha^0} - \overrightarrow{v_\alpha}\right) \tag{3.6}$$

Where $\overrightarrow{v_\alpha}$ is the actual velocity, and $\tau_\alpha$ is the relaxation time to adapt to desired velocity.

$\overrightarrow{f_{\alpha\beta}}$ is a repulsive force that reflects the intent for agent $\alpha$ to avoid contact with surrounding crowd $\beta$:

$$\overrightarrow{f_{\alpha\beta}}\left(\overrightarrow{r_{\alpha\beta}}\right) = -\nabla_{\overrightarrow{r_{\alpha\beta}}} V_{\alpha\beta}(b) \tag{3.7}$$

Where $\nabla_{\overrightarrow{r_{\alpha\beta}}}$ is a repulsive potential that decreases monotonically with $b$ and $\overrightarrow{r_{\alpha\beta}} = \overrightarrow{r_\alpha} - \overrightarrow{r_\beta}$. $\overrightarrow{r_\alpha}$ denotes the position of agent $\alpha$ and $\overrightarrow{r_\beta}$ denotes the position of surrounding crowd $\beta$.

$b$ represents the semi-minor axis of an elliptical equipotential contour centered on agent $\alpha$, which models the comfort zone. This ellipse is oriented in the walking direction, reflecting the greater amount of space pedestrians require in front of them to take steps or navigate efficiently.

$\overrightarrow{F_{\alpha B}}$ is a repulsive force that reflects the intent for agent $\alpha$ to avoid contact with objects $B$ in external environment:

$$\overrightarrow{F_{\alpha B}}\left(\overrightarrow{r_{\alpha B}}\right) = -\nabla_{\overrightarrow{r_{\alpha B}}} U_{\alpha B}(\| \overrightarrow{r_{\alpha B}} \|) \tag{3.8}$$

where $U_{\alpha B}(\| \overrightarrow{r_{\alpha B}} \|)$ is a monotonic decreasing potential as well and $\overrightarrow{r_{\alpha B}} = \overrightarrow{r_\alpha} - \overrightarrow{r_B^\alpha}$ with $\overrightarrow{r_B^\alpha}$ denotes the location of object $B$.

Additionally, $\overrightarrow{F_{\alpha\iota}}$ is an attractive force originated from points of interest $l$, such as friends or landmarks, typically aligned with the desired direction $\overrightarrow{e_\alpha}$. This force is time-dependent, as the agent's level of interest may decrease over time.

SFM has been proven efficient for capturing emergent crowd behaviour such as lane formation (Belotti & Martins, 2021). Notably, the simulation that validates the existence of arch formation was based on SFM (Helbing, et al., 2002). Consequently, SFM could be viewed as a perfect example of system thinking, as it takes into account of elements, relations and external environment, and is capable of capturing macroscopic level phenomena.

However, SFM has also faced criticism, particularly regarding the interpretation of behavioural terminology such as "panic" (Haghani & Ronchi, 2024), and the difficulty of calibrating model parameters such as the strength factor $A$ (Kretz, et al., 2018). Furthermore, SFM assumes each agent as a rigid body (Wang, et al., 2024), which cannot capture the physical compression that occurs in high density crowd situations.

Despite this limitation, SFM has laid important foundations for subsequent research in crowd modelling, such as being used as a benchmark (Jain, et al., 2025), or ground-truth physics model (Zhang, et al., 2022). Notably, the term desired velocity has been extremely valuable even beyond the scope of SFM. For instance, it has been frequently employed in the literature to characterise rushing behaviours in crowds (Peng, et al., 2016), which is also adopted in this thesis as discussed in section 2.4.1.3.

#### *3.3.2.2 Cellular Automata*

Cellular Automata (CA), originally proposed by Von Neumann (1966), is a grid-based dynamic model. In contrast to SFM, CA discretizes time, space, and cell states (Li, et al., 2019), and is composed of four key components (Li, et al., 2018):

- Cells: Basic units in the grid, each representing a small spatial unit of the environment.
- Cellular space: the entire grid of cells that defines the environment.
- Cell states: discrete values assigned to cells at each time step, indicating their current condition. For example, empty, occupied by a pedestrian, or an obstacle.

- Transition rules: a set of predefined rules that determine how a cell's state will change at the next time step, typically based on the current state of the cell and the states of its neighbouring cells.

A schematic representation of the CA model is shown in Figure 3.3.

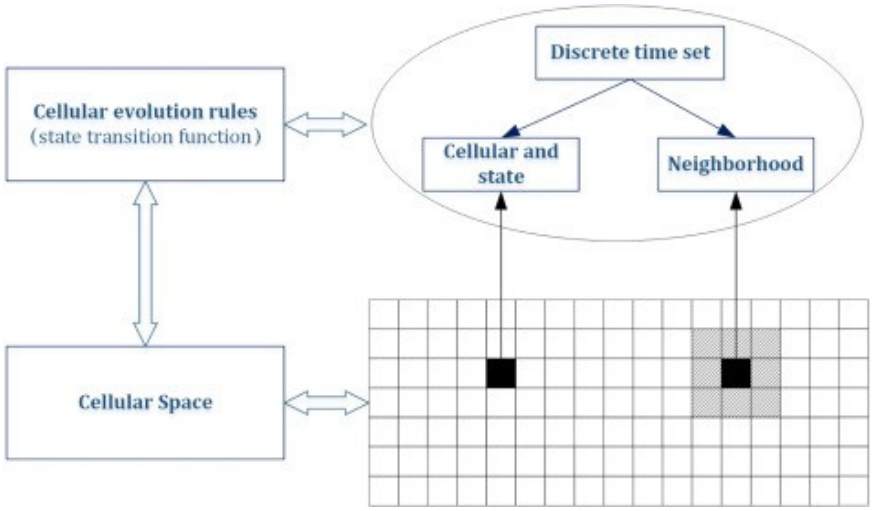



*Figure 3.3 Schematic diagram of cellular automata (Li, et al., 2018), reproduced with permission from Springer*

In crowd modelling, the size of each cell is often assumed to be 0.4m * 0.4m assuming the shape of square. However, the shape of hexagon has also been adopted to overcome the isotropy problem, where diagonal movements in a square grid are artificially longer than horizontal or vertical movements (Li, et al., 2019).

More importantly, CA introduces the idea of pedestrian's impact size, which is more aligned to real human being. In contrast, the original formulation of SFM treats pedestrians as point particles with no spatial extent.

A classical transition rule used by CA in crowd modelling is floor filed (Schadschneider, et al., 2009), composed of a static floor field $S$ and a dynamic floor field $D$ (Kirik, et al., 2007):

- Static floor field $S$ describes the shortest distance to the desired destination, which does not change over time.

- Dynamic floor field $D$, in contrast, is a time-dependent virtual trace left by moving pedestrians, which gradually fades and spreads, and influences others to follow similar paths, mimicking herding behaviour in crowds.

The probability $p_{ij}$ for an agent moving to cell $ij$ is then expressed as (Xing, et al., 2024):

$$p_{ij} = \mathrm{N} \cdot \exp\left(k_S S_{ij} + k_D D_{ij}\right) \cdot \epsilon_{ij} \tag{3.9}$$

Where $k_S$ and $k_D$ are sensitivity parameters, both reflecting the tendency of agent moving to the direction specified by the floor field. The term $\epsilon_{ij} = 0$ serves as an occupancy indicator: it equals 0 if cell $ij$ is occupied else 1. N is the normalization factor ensuring total transition probability across all available neighbouring cells adds up to 1.

The agent then selects a neighbouring cell to move into, based on the computed transition probabilities. This highlights a key distinction between CA and the SFM: while the future trajectory in SFM is typically deterministic, CA adopts a stochastic approach. In CA, the pedestrian's movement is governed by probabilistic rules, which better reflect the variability and unpredictability inherent in real human behaviour.

While CA offers a powerful framework for modelling pedestrian flow, the discretization adopted by CA can introduce unrealistic behaviour, particularly when using square grids, which may oversimplify the complexity of real pedestrian dynamics (Li, et al., 2019). Furthermore, the use of uniform cell sizes in CA implicitly assumes that all agents have the same physical size, an assumption that does not hold for heterogeneous populations such as adults and children (Ronchi, 2021). To address this limitation, subsequent studies have proposed multi-cell representations in which agents occupy multiple cells to better reflect differences in body size (Bazior, et al., 2025).

#### *3.3.2.3 Deep Learning*

Neural network-based deep learning (DL) is a subset of machine learning (ML) that model complex patterns in data using architectures inspired by the human brain (LeCun, et al., 2015).

Unlike SFM and CA introduced above, DL is a data-driven methods requires a substantial volume of input data. A neural network model defines a mapping:

$$y = f(x; \theta) \tag{3.10}$$

Where $x$ denotes the input value and $y$ the output value. $\theta$ stands for a set of learnable parameters and the objective is to learn the optimal parameters $\theta$ that result in the best function approximation (Goodfellow, et al., 2016).

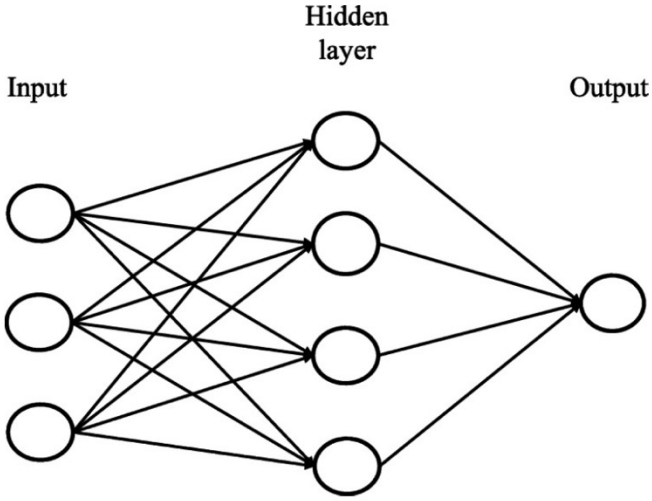



*Figure 3.4 Neural network (Kugunavar & Prabhakar, 2021)*

As illustrated in Figure 3.4, a typical neural network model consists of interconnected layers of nodes (neurons), where each layer transforms the input data into more abstract representations. The network typically includes an input layer, one or more hidden layers, and an output layer.

Each neuron $j$ processes $n$ input value $x_i$ using weighted connections $w_{ji}$ and a nonlinear activation function $g$, denoted as:

$$y_j = g\left(\sum_{i=1}^{n} w_{ji} x_i + b_j\right) \tag{3.11}$$

Where $y_j$ is the output of neuron $j$ and $b_j$ is the corresponding bias term (Sazlı, 2006). Both weights $w_{ji}$ and $b_j$ belong to the learnable sets of parameters $\theta$. This formulation enables the network to learn non-linear relationships between input and output.

In the field of crowd modelling, both models introduced above have effectively incorporated spatial correlations with surrounding crowds and the external environment. However, temporal dependencies, particularly the agent's past trajectories, remain less thoroughly addressed. Among the two, only the CA incorporates movement history through the dynamic floor field that reflects past activity of surrounding crowds in the environment.

To capture temporal dependencies in movement, Long Short-Term Memory (LSTM), a widely used deep learning architecture for time-series data, has been widely used due to its ability to incorporate and learn from past trajectories (Alahi, et al., 2016).

*Long Short-Term Memory*

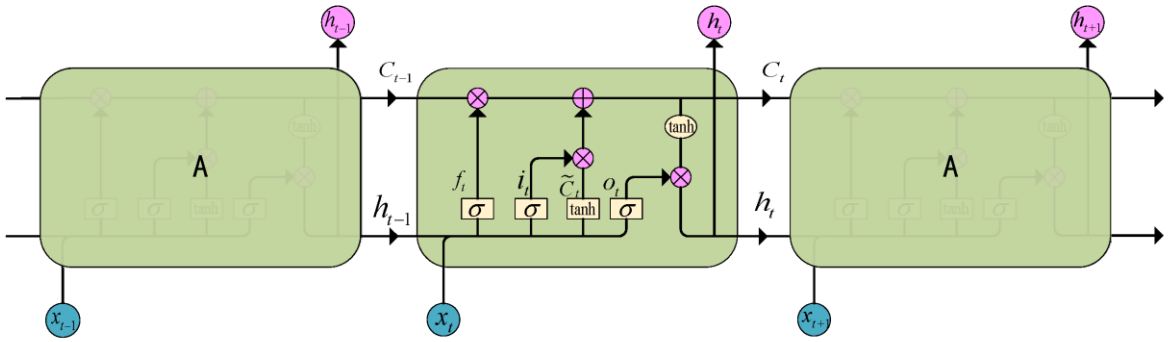



*Figure 3.5 LSTM architecture (Jing, et al., 2022)*

LSTM, originally proposed by Hochreiter & Schmidhuber (1997), is designed to capture long-term dependencies. As depicted in the centre of Figure 3.5, an LSTM unit is composed of a cell state $C_t$, which serves as the memory of the unit, and three core gates: the forget gate $f_t$, the input gate $i_t$, and the output gate $o_t$.

These gates regulate the flow of information at each time step. The forget gate $f_t$ determines which part of the previous cell state $C_{t-1}$ should be discarded. The input gate $i_t$ controls the incorporation of new information $\widetilde{C}_t$, generated from the current input $x_t$ and previous hidden state $h_{t-1}$.

The updated cell state $C_t$ is computed as:

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \tag{3.12}$$

Where $f_t \odot C_{t-1}$ denotes the retained past information and $i_t \odot \widetilde{C_t}$ represents the newly selected content.

Finally, the output gate $o_t$ filters what portion of the updated cell state is passed forward to generate the hidden state $h_t$, which also serves as the output at timestep $t$.

This gated structure allows LSTMs to retain relevant information across long sequences. As illustrated in Figure 3.5, at each timestep $t$, the input $x_t$ is processed into a single LSTM unit to produce a hidden state $h_t$. In the end, a hidden state that contains crucial information alongside the sequence is generated, making it a strong candidate for capturing temporal dependencies in movement.

A classical application of LSTM in crowd modelling is Social-LSTM, developed by Alahi, et al. (2016), and it is a DL model designed to predict human trajectories based on sequential positional data. The model takes as input a series of observed agent coordinates over a fixed time window, typically represented as $(x, y)$ at each timestep t.

These coordinates are used to model individual motion dynamics through an LSTM cell, which learns the temporal dependencies in an agent's past trajectory. By encoding the trajectory into hidden states, the model captures patterns in walking behaviour, enabling it to predict future positions.

Additionally, Social-LSTM captures spatial correlations with surrounding crowds as well, which is a distinctive feature of Social-LSTM, named as social pooling.

At each timestep t, social pooling is performed by constructing a spatial grid around the target agent and aggregates the hidden states of neighbouring LSTMs located within this grid. This

aggregated representation is passed to LSTM at the next timestep, allowing the model to implicitly account for social interactions such as collision avoidance or group movement.

Social-LSTM has demonstrated strong performance in pedestrian trajectory prediction and continues to serve as a benchmark for evaluating more advanced prediction models (Manh & Alaghband, 2019; Salzmann, et al., 2021; Gupta, et al., 2018). From a system thinking perspective, while Social-LSTM does not explicitly incorporate influences from external environment, subsequent studies, such as Scene-LSTM model (Manh & Alaghband, 2019), had demonstrated that LSTM-based model are capable of capturing such influences.

However, despite the good performance on predicting future trajectory, such models suffer from explainability issues, mainly due to the "black box" nature of neural network which limits insight into their internal decision making processes. Consequently, although interaction patterns may be captured by the model, it remains difficult to extract meaningful insights or interpret the underlying rules governing those interactions (Sheu, 2020).

*Generative Adversarial Network*

Generative Adversarial Network (GAN), introduced by Goodfellow, et al. (2014), consists of two neural networks, a generator and a discriminator, that are trained simultaneously in an adversarial framework. While the generator aims to produce data samples that are indistinguishable from real data, the discriminator seeks to correctly differentiate between real and generated samples. This dynamic encourages the generator to learn the underlying data distribution, leading to highly realistic outputs in domains such as image generation (Zhang, et al., 2018) and video synthesis (Chen, et al., 2019).

In the field of crowd analysis, SoPhie (Sadeghian, et al., 2018) is a classical model that utilizes GAN to predict future trajectory. It is capable of extracting features from surrounding crowds, known as the social features $C^{\mathrm{So}}$, as well as from a top-down image of external environment, referred as the physical features $\mathrm{C}^{\mathrm{Ph}}$.

The generator $G$ and discriminator $D$ in SoPhie both employ LSTM as well. For generator $G$, it utilizes a decoder LSTM to produce future trajectory points. For agent $i$ at each time step $\tau$, the decoder receives a concatenated context vector:

$$C_{\tau}^{G}(i) = \left[C_{\tau}^{\mathrm{So}(i)},\ C_{\tau}^{\mathrm{Ph}(i)},\ z\right] \tag{3.13}$$

Where $z \sim \mathcal{N}(0, I)$ is a latent noise vector sampled to induce diversity in the predicted trajectories. The decoder LSTM then generate $\tau^{\mathrm{th}}$ future trajectory based on context vector along with the previous hidden state $h_{\tau-1}^{\mathrm{dec}(i)}$.

$$\widehat{Y}_{\tau}^{i} = \mathrm{LSTM}_{\mathrm{dec}}\left(C_{\tau}^{G}(i),\ h_{\tau-1}^{\mathrm{dec}(i)}\right) \tag{3.14}$$

On the other hand, the discriminator $D$ takes input in the format of trajectory which is randomly sampled either from the ground truth or generated trajectory $\widehat{Y}_{\tau}^{i}$ to produce a predicted label.

It can be observed that unlike Social-LSTM, the trajectory generated by SoPhie stochastic rather than deterministic, due to the introduction of the latent noise vector $z$. This introduction has again captured unpredictability inherent in real human behaviour and make SoPhie a common benchmark in the field of crowd analysis (Huang, et al., 2021; Li, et al., 2025).

#### *3.3.2.4 Reinforcement Learning*

While DL aims to model the relationship between past trajectory and future trajectory, Reinforcement Learning (RL) takes an agent-centric perspective, seeking to replicate pedestrian behaviour by learning how agents make sequential decisions.

In RL, an environment is constructed to allow agents to explore and learn. This environment provides both the spatial context, such as physical infrastructures, and the dynamics, including motion rules and interactions with surrounding crowds. The entire trajectory of pedestrians could be viewed as a Markov Decision Process (MDP) denoted by the tuple $\mathrm{M} = (\mathrm{S}, \mathrm{A}, \mathrm{T}, \mathrm{R}, \gamma)$ (Sutton & Barto, 2018), where:

- S is the state space, which represents the information available to the agent at each timestep. It can include agent-specific variables, such as position and velocity, and environment-specific variables, such as the positions of static obstacles or other agents. The state space can be flexibly designed to reflect different objective of research.

- A is the action space, describing the possible actions the agent can take. For example, selecting a velocity vector, heading direction, or discrete movement decision.

- $T(s_t, s_{t+1}, a_t)$ is the transition probability from state $s_t$ to state $s_{t+1}$ under action $a_t$, determined by the environment dynamics.

- R is the reward function, which encodes the agent's objectives and behavioural preferences. It reflects the intentions of the agent, such as minimizing travel time, avoiding collisions, or staying close to a group, by assigning a reward or penalty to relevant behaviours.

- γ is the discount factor ranged on [0,1), which determines how much future rewards are valued relative to immediate ones. A lower γ encourages short-term behaviour, while a higher γ favors long-term planning.

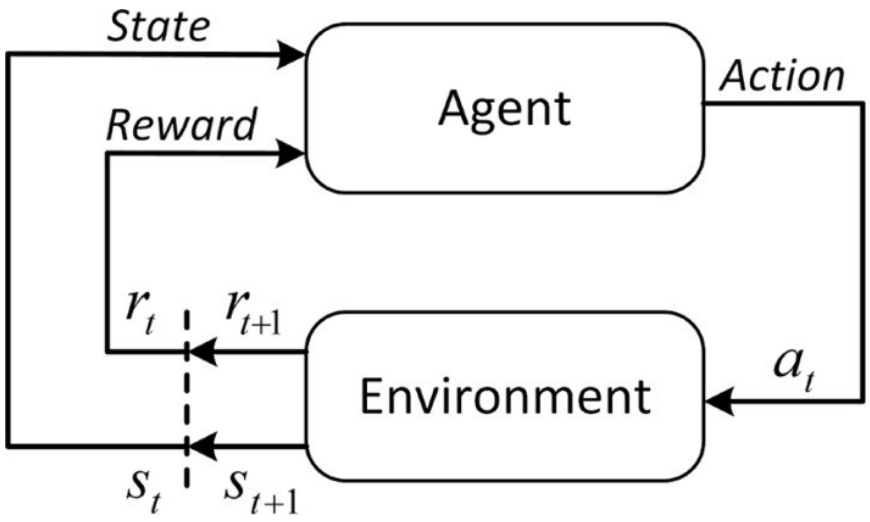



*Figure 3.6 Interactions of RL agents with environment (Hassani, et al., 2022), reproduced with permission from Elsevier*

The interactions of an agent with the environment are illustrated in Figure 3.6. At each timestep $t$, the agent selects an action $a_t$ based on observed state $s_t$, and receiving a reward $r_t$ based on the feedback from the external environment.

The agent's behaviour is defined as a policy $\pi(a \mid s)$, which is the probability of choosing action $a$ when in state $s$. The objective of RL is to learn the optimal policy $\pi^*$ that maximizes cumulative reward over time (Sutton & Barto, 2018):

$$\pi^*(s) = \arg\max_{a \in \mathcal{A}} E\left[r_t + \gamma V^*(s_{t+1}) \mid s_t = s, a_t = a\right] \tag{3.15}$$

Where $V^*(s)$ denotes the best possible cumulative reward the agent can achieve starting from state $s$:

$$V^*(s) = \max_{\pi} E_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_0 = s\right] \tag{3.16}$$

Based on the training objective of the optimal policy $\pi^*$, it is reasonable to conclude that, in contrast to Social-LSTM's emphasis on past trajectory patterns, RL-based models are primarily focused on the expected future trajectory.

In the field of crowd modelling, RL has been adopted mainly due to its ability to capture human-like decision making, which provide more insights for understanding crowd behaviour (Wang, et al., 2024). However, similar to SFM, a good practical performance of RL significantly depends on the design of reward function, which involves delicate calibration (Fu, et al., 2018).

#### *3.3.2.5 Imitation learning*

Another approach that adopts the decision making perspective is imitation learning, which is also commonly formulated within MDP. In contrast to RL, imitation learning aims to recover the policy $\pi(a \mid s)$ of the agents directly from collected data, typically in the form of state-action pair denoted as $(s, a)$, and is a data-driven method as well. This method bypasses the often difficult process of designing an appropriate reward function (Wang & Li, 2021).

*Behaviour Cloning (BC)*

BC is a classical imitation learning algorithm that is categorised as supervised learning. The objective of BC is to learn a $\pi(a \mid s)$ that is capable of generating trajectories that resemble those in a set of expert demonstrations. These expert demonstrations are often denoted as a collection of trajectories $\mathcal{D} = \{\tau_i\}_{i=1}^{N}$, where $\tau_i = \{(s_t, a_t)\}_{t=0}^{T_i}$, $N$ is the total number of trajectories, and $t$ indexes each individual timestep within each trajectory (Osa, et al., 2018).

This objective is achieved by finding the parameters $\theta$ of policy $\pi_\theta$ that minimize the discrepancy $\mathcal{L}$ between the predicted actions and the expert actions, denoted as:

$$\min_{\theta} \sum_{i=1}^{N} \mathcal{L}(\pi_\theta(s_i), a_i) \tag{3.17}$$

This approach leverages the strengths of supervised learning, allowing the use of various function approximators, including neural networks, to model complex policies.

However, a notable limitation of BC is its vulnerability to distributional shift (Foster, et al., 2024). Since the model is trained on the distribution of states encountered in the expert demonstrations, it may perform poorly when it encounters states outside this distribution during deployment, especially for social interactions and environmental influences. This issue can lead to compounding errors over time, as the policy's mistakes may lead it into unfamiliar states where it has not been trained to act appropriately.

### 3.3.3 Summary and Comparison

In this section, six representative crowd modelling models are reviewed, examining both their strengths and limitations. All the models align with the principle of system thinking, capturing the intrinsic properties of agents, social interactions, and the influence of the external environment, as summarized in Table 3.1.

| Model Name | Intrinsic Properties | Social interaction | External Environment |
|---|---|---|---|
| SFM | Intention | Attractive/Repulsive force | Attractive/Repulsive force |
| Floor field-based CA | Intention | Herding behaviour | Obstacle avoidance |
| S-LSTM | Past trajectory | Hidden states from surrounding crowds | No, but can be incorporated |
| SoPhie | Past trajectory | Hidden states from surrounding crowds | Features extracted from a top-down image |
| RL | Intention, Action, Expected future trajectory | Rewards from interacting with surrounding crowds | Rewards from interacting with environment |
| BC | Action | Encoded via expert state inputs | Encoded via expert state inputs |

*Table 3.1 Crowd modelling models*

## 3.4 Mesoscopic approach

Given both macroscopic and microscopic approaches have strengths and weaknesses, a hybrid approach, known as mesoscopic perspective (Elaiw, et al., 2019), has developed to gain the benefits of both approaches while reducing their disadvantages. This approach aims to bridge the gap between the macroscopic and microscopic perspectives by combining "top-down" and "bottom-up" modelling approach.

One example of mesoscopic approach is Trajectron++, developed by Salzmann, et al. (2021). At the core of Trajectron++ is a Conditional Variational Autoencoder architecture, which enables the model to capture a distribution over multiple plausible future trajectories, rather than producing a single deterministic path, similar to the generator of SoPhie.

Specifically, Trajectron++ expresses the conditional distribution $p(y \mid x)$ of future trajectories $y$ based on input $x$ using a categorical latent variable $z \in \mathcal{Z}$. A latent variable is an unobserved variable that captures the underlying intentions behind high-level behaviour modes that is not able to be explicitly labelled, as discussed in section 3.2.2.2. The model incorporates latent variable $z$ and decomposes the conditional prediction as:

$$p(y \mid x) = \sum_{z \in \mathcal{Z}} p(y \mid x, z) \cdot p(z \mid x) \tag{3.18}$$

This formulation allows the model to generate diverse trajectory samples for each agent, capturing multimodality while remaining conditioned on individual inputs. Moreover, the discrete nature of $z$ improves interpretability, as one can visualize which high-level behaviours belong to each $z$ by generating a representative trajectory.

However, unlike SoPhie, which introduces a latent variable purely as a source of stochasticity, the latent variable used by Trajectron++ could be seen as a subset of action recognition, as it looks at discrete behavioural distribution of all the crowds. Consequently, Trajectron ++ could be viewed as a perfect example of mesoscopic approach, due to its combination of action recognition and crowd modelling.

## 3.5 Summary

In this chapter, key methods of crowd analysis are reviewed, with a particular focus on those most relevant to crowd management. Building on the qualitative attributes of crowd behaviour discussed in Chapter 2, this chapter adds a layer of quantitative understanding, establishing links between the two, for instance, highlighting the reason behind the critical role of extreme crowd density.

Several of the methods reviewed here are applied in this research. In particular, the cluster analysis introduced in section 3.2.2.2 is employed due to its effectiveness in identifying patterns. Additionally, the crowd modelling methods reviewed in section 3.3 provide a foundational basis for understanding the intrinsic properties of crowd movement, as well as their interactions with the surrounding crowds and external environment. More importantly, these methods are also served as benchmarks for evaluating accomplished crowd modelling

model. In turn, the result of this research is able to contribute back to these established models, such as the SFM, by offering new insights that can inform their refinement and practical deployment.

Finally, the mesoscopic approach introduced in section 3.4, as a combination of macroscopic approach and microscopic approach, provides guidance for research approach throughout this research and is elaborated further in the next chapter.

# 4. Methodological Overview

## 4.1 Introduction

Following identification of gaps in current crowd management and a review of analytical methods in understanding crowds, this chapter outlines the philosophical and methodological foundations of this research.

Given the dual nature of crowd, which is both physically observable and socially constructed, this research adopts a pragmatic stance that integrates both realist and constructivist perspectives. Such an approach enables the use of quantitative modelling techniques alongside scenario-based analysis.

The chapter begins by discussing the research philosophy in section 4.2, followed by the adopted research approach in section 4.3, and concludes with the research strategy used in subsequent chapters in section 4.4.

## 4.2 Research Philosophy

A research philosophy aims to answer two questions behind a study (Saunders, et al., 2019):

- ontology, i.e., what is understood to constitute reality.

- epistemology, i.e., how knowledge about that reality can be acquired.

From an ontological standpoint, this research acknowledges the existence of an objective reality in the form of physical constraints, spatial configurations, and social mechanisms that shape crowd movement. This aligns with a realist ontology (Bhaskar, 2008), which assumes the existence of causal mechanisms that operate independently of human perception. Crowd behaviours, especially movement trajectory and emergent phenomena such as arch formation,

are therefore considered as observable, modellable, and explainable through simulation and quantitative methods.

Epistemologically, on the other hand, the research also recognises that human perception, interpretation, and decision making are socially and situationally constructed. This reflects a constructivist epistemology (Schwandt, 1994), wherein knowledge is viewed as being co-created through interactions with the social scenario, and surrounding crowds. Human reactions to stimuli and decisions to comply or diverge from expected behaviours are influenced by subjective factors, such as perceived scenario settings and individual goals. These elements are best understood through qualitative, scenario-based methods that account for contextual variation.

Within this epistemological perspective, it is also recognised that there are differing views on whether human behaviour in crowds can ever be fully captured through deterministic modelling. While some researchers argue that behavioural uncertainty is irreducible due to the complexity and social nature of human decision-making, others suggest that current limitations primarily reflect incomplete data and understanding. Recent work has highlighted that, although physical mechanisms governing crowd interactions are comparatively easier to characterise, behavioural aspects of crowd dynamics remain subject to significant uncertainty (Lovreglio, et al., 2014).

Combining stances from both sides, this research adopts a pragmatic worldview, drawing from both realist and constructivist perspectives to account for the complexity of human behaviour in crowd dynamics. This combination of philosophical positions enables inquiry from a socio-technical system perspective, consistent with the interdisciplinary and applied nature of the research. It allows for both the generalisation of structural crowd dynamics across scenarios and the contextual understanding of behaviour within specific settings.

Pragmatism (Creswell, 2014) supports the integration of multiple methods and perspectives based on their utility in answering the research question. In this research, pragmatism legitimises the combination of data-driven machine learning models, qualitative scenario analysis and Virtual Reality based experiment, allowing the researcher to engage with both empirical patterns and contextual meaning.

## 4.3 Research Approach

The mesoscopic approach adopted in crowd analysis, as previously discussed, integrates both ‘top-down’ and ‘bottom-up’ perspectives and has become as a valuable framework for understanding crowd behaviour.

For instance, group behaviour analysis can be viewed as a mesoscopic approach between crowd modelling and action recognition (Bendali-Braham, et al., 2021) as it groups crowd that share similar behaviour pattern into a single entity, while further exploring the behaviour of each entity. This research adopts a mesoscopic perspective to better understand crowd behaviour within different social settings, particularly due to the need to incorporate the influence of the social environmental context, as discussed in chapter 2.

While action recognition often considers actions that are commonly observed in daily scenarios (Ullah, et al., 2019; Lamba & Nain, 2017), the results of crowd modelling (especially motion prediction) are typically limited to originally analysed scenes (Qiao, et al., 2022; Qiao, et al., 2019). To bridge the gap, scenarios that share common patterns among different scenes are considered, but still maintaining distinct characteristics compared to alternative scenarios.

To achieve such objective, a top-down approach is employed to categorise crowd disasters into distinct types by examining them at the macro level. These categories then guide the selection of representative scenarios for experimental data collection and further analysis.

At each stage of this process, a bottom-up approach analysis is conducted to identify behavioural patterns that emerge from the aggregated data at the individual and group levels. This combination allows us to capture both the systemic property of crowd disasters and the nuanced behaviours that contribute to them.

## 4.4 Research Strategy

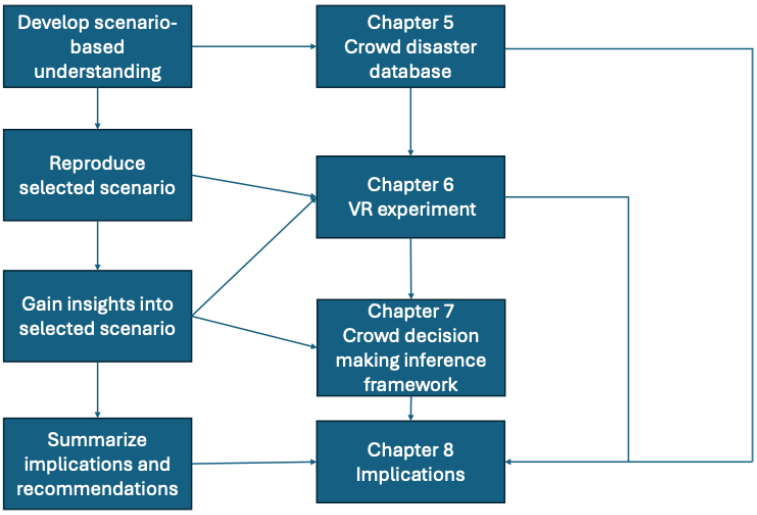



*Figure 4.1 Research Strategy flowchart*

This section outlines the research strategy employed in subsequent chapters, as illustrated in Figure 4.1. The flowchart presents the sequential and interlinked components of the research process, highlighting how each chapter contributes to the overall study. The detailed strategy used for chapters 5, 6 and 7 are discussed in the following subsections.

### 4.4.1 Chapter 5 Research Strategy

The research strategy adopted in Chapter 5 centres on the construction and analysis of a crowd disaster database, designed to contribute to a new comprehensive understanding into the characteristics of crowd disasters. The aim is to enable a qualitative exploration of incident characteristics, taking into account of the impact of external environment.

The data source derives from a comprehensive review of news archives and publicly available incident reports from 1989 to 2021. Each case was assessed and coded according to a predefined code book, developed based on causal factors extracted from the FIST model, as well as an external environment element, including physical environment and social scenario.

In addition, data quality is checked using commonly established metrics for assessing qualitative incident database. The resulting database enables clustering of similar incident types,

leading to the identification of scenario-level patterns, and the subsequent classification of nine distinct types of crowd disasters.

The findings from this chapter serve as the empirical foundation for this research, providing a comprehensive understanding of the crowd disasters from a socio-technical system perspective by incorporating external environment factors.

Furthermore, the identification of distinct types of crowd disaster has enabled the subsequent study for a more focused understanding of crowd behaviour under specific scenarios. This classification provides both a conceptual framework and empirical basis for the design and implementation of targeted virtual reality (VR) experiments in Chapter 6, which aims to reproduce the settings of a selected crowd disaster type.

### 4.4.2 Chapter 6 Research Strategy

The research strategy adopted in Chapter 6 is centred on the design and conduct of a controlled VR experiment to investigate crowd behaviours under a jostle scenario, selected from scenario types identified in chapter 5 due to its high frequency.

Using VR enables the safe and repeatable reproduction of high-density crowd conditions, which would be impractical or unsafe to replicate in physical environments. The objective is to extract meaningful behavioural insights from movement patterns, contributing to a deeper understanding of crowd dynamics in competitive, non-emergency settings.

A bespoke multi-user VR environment was then developed to simulate a supermarket sales scenario, in which participants compete to pass through a narrow entrance to retrieve limited items. This setting reproduces a common observed configuration in real-world jostle-related crowd incidents. The experiment involved three rounds, each with over 20 participants navigating the same scenario, and trajectory data were collected at high temporal resolution.

The effectiveness of the reproduction was validated through the emergence of relevant phenomena, as well as questionnaires collected from the participants. Following validation, clustering was applied to the trajectory data to identify distinct movement patterns. This

enabled the classification of behavioural types, most notably distinguishing between “patient” and “impatient” types.

This chapter constitutes the experimental part of this research, focused on data collection followed by an initial stage of data analysis. The findings validated VR as a powerful tool for the investigation of crowd analysis, and a diversity of the types of behaviour within the crowd has also highlighted the importance of deeper understanding of crowd behaviour.

Both the data collected in this chapter and the identification of the two types of behaviour were used in chapter 7 for a scenario-specific decision-making inference.

### 4.4.3 Chapter 7 Research Strategy

The research strategy adopted in chapter 7 focuses on the development of a decision-making inference framework. Analysing the patient and impatient behavioural types identified in chapter 6, this chapter seeks to explore how different factors, both internal and external, influence individual behavioural tendency. The objective is to link observed trajectory features with underlying decision dynamics, thereby offering a more interpretable understanding of crowd behaviour.

To achieve this, the strategy integrates Inverse Reinforcement Learning (IRL), a data-driven approach used to infer the reward functions that best explain observed behaviour. Within the IRL framework, pedestrian trajectories are interpreted as outcomes of sequential decision-making processes, where individuals repeatedly select actions in response to their local environment. IRL therefore allows the implicit objectives to be estimated directly from observed trajectories.

Following the training of a reward network via IRL, a model interpretability method is applied to examine how each feature contributes to the predicted reward values. This step provides transparent insight into which factors are most influential in shaping decision making. By analysing these feature attributions across behavioural clusters, the research identifies both shared and distinct decision drivers between “patient” and “impatient” participants.

The findings from this chapter offer actionable insights for developing crowd management strategies in jostle scenarios, particularly by emphasising the importance of understanding crowd behaviours. Moreover, the decision making inference framework proposed in this Chapter is not only applicable to crowd analysis but also for broader pedestrian-focused domains, such as urban planning, as further discussed in chapter 8.

# 5. Crowd disaster database

## 5.1 Introduction

This chapter develops a holistic qualitative understanding of crowd disasters, with particular focus on contextual fluctuations in crowd behaviour arising from complex interactions with the external environment.

To achieve this objective, four primary causal factors of crowd disaster identified in the FIST model (Fruin, 1993) are selected as qualitative characteristics of crowd disasters in this section. Incident scenario (e.g., sports event) and the physical environment (e.g., stadium) are also incorporated as two additional characteristics to reflect the impact from external environment, especially the social scenario, as emphasised in the case study of the 2014 Shanghai crowd disaster in section 2.4.

Furthermore, a code book (i.e. a form of data dictionary) of crowd disasters was developed based upon selected characteristics. While code books have been developed in other domains, to the best of the authors' knowledge, this is the first such book for crowd disaster. A total of 293 crowd disasters ranging from 1989 to 2021 have been identified from news articles and coded based on the code book.

A major outcome of this chapter is the identification of nine types of crowd disasters and their underlying factors, thereby enabling targeted, rather than generic, crowd management measures and further investigation. The data are verified based on its quality and analysed through a cluster analysis algorithm (K-Mode).

Section 5.2 reviews the relevant literature and identifies current research gaps. Section 5.3 introduce the methodologies for database construction and data analysis respectively. Section 5.4 discusses the database and section 5.5 concludes the chapter.

The main content of this chapter is published in a peer-reviewed journal article:

Yang, X., Liu, Y., Majumdar, A., Grass, E., & Ochieng, W. (2024). Characteristics of crowd disaster: Database construction and pattern identification. International Journal of Disaster Risk Reduction, 111, 104653, https://doi.org/10.1016/j.ijdrr.2024.104653.

## 5.2 Literature review

As highlighted by Hsieh, et al. (2009, p. 217) an "improved collective understanding of human stampedes[1] has crucial planning implications and is indispensable in mitigation of this unique form of crowd disaster." Two perspectives are often taken to seek patterns among different crowd disasters. One perspective considers crowd management.

For instance, Fourati, et al. (2017) noted common problems faced by Hajj in Mecca such as the failure to follow the schedule. Dickie (1995) identified common threads leading to major crowd disasters in the U.K. from 1863 to 1994, including inadequate planning and flaws within infrastructures. Likewise, Still, et al. (2020) analysed 68 crowd safety projects that the lead author had experienced in order to recognise potential place of crowd safety issues from the perspective of design, information and management.

For the other perspective, considerable research has studied the common characteristics of crowd disasters, including venues (Fruin, 1993), causal factors (mechanism) (Shao, et al., 2019; Lu, et al., 2021; Ngai, et al., 2009), fatalities (Lu, et al., 2021; Hsieh, et al., 2009), and other epidemiological characteristics (Hsieh, et al., 2009).

In this chapter, the latter perspective is adopted as it enables a broader understanding of crowd disaster beyond flaws within management procedure, as well as from a socio-technical system perspective. Consequently, it is proposed to incorporate the FIST model (Fruin, 1993), introduced in section 2.4, as causal characteristics of crowd disaster.

[1] Although the term "stampede" is frequently used in media reporting and earlier academic literature, it is avoided in this study because it implies panic-induced rushing behaviour, which as criticised in section 2.4.1.3, does not accurately reflect the underlying mechanisms of most crowd disasters. The term "crowd disaster" is therefore used consistently throughout.

The 2014 Shanghai disaster case study, also presented in section 2.4, shows that in addition to the causal factors derived from the FIST model, it is essential to consider the social scenario and the physical environment to capture the influence from external environment. Consideration of such characteristics does not imply any causality, in contrast to the factors derived from the FIST model.

Accordingly, the scenario of the event, due to its ability to capture the common features of participants and their goals, is incorporated as well. Another characteristic, environment, is also introduced to comprehend the nature of the spacious settings in the infrastructure (allowed degree of freedom in the movement direction), apart from the local physical space studied in FIST model. Additionally, casualty is added as the last characteristic to reflect the severity of the consequences.

Once key characteristics are identified, developing a database is invaluable for holistic pattern identification (Hsieh, et al., 2009). To the best of the author's knowledge, databases in the literature, mostly categorise incidents based on either the severity of their consequences in order to identify the types of crowd disaster with the highest fatality rates (Lu, et al., 2021; Ngai, et al., 2009) or the distribution of underlying mechanisms with respect to a particular time period (Hsieh, et al., 2009). Recent work by Feliciani, et al. (2023) adopts a broader analytical scope, examining crowd disasters between 1900 and 2019 and analysing trends in occurrence frequency and fatalities over time. Notably, the purpose of the gathering is also systematically recorded and made publicly available through an open-access database (Feliciani, 2023).

Such databases suffer from several drawbacks, notably the failure to introduce a specific inclusion criteria, i.e., the criteria (e.g., time, information source) for an incident to be included in the database, and the lack of precise definitions, both for a crowd disaster itself and each of its characteristics. This has resulted in discrepancies in the databases, i.e., not all collected incidents describe the same disaster, and this provides challenges for research reproduction. This chapter addresses this gap by adopting a codebook (attached in Appendix II) which states the inclusion criteria of the database and provides an exact definition for each characteristic.

## 5.3 Methodology

### 5.3.1 Database construction

#### *5.3.1.1 Source Identification*

The ideal primary source of any crowd disaster data is the published official report and the accompanying video footage. However, due to difficulties in accessing the relevant data from organisations (Bell, et al., 2018), news articles (in print or online) and solely online sources are often used as alternatives. Consequently, an online news database, Factiva (Dow Jones & Reuters, 2023), has been selected as the main source of data in this chapter due to its comprehensive coverage of global news and its availability through Imperial College London's institutional subscription.

Although near miss incident can substantially expand available evidence and provide valuable information, it is virtually not published by anyone due to fear of negative backlash (Haghani, et al., 2023). This scarcity is further reinforced by media reporting biases toward events involving injury or harm (Kitzinger, 2009), meaning that near miss incidents are unlikely to appear in the news database. Consequently, this study focuses on crowd disaster that resulted in injury or death, and since crowd force is the main cause of fatality and injury (Fruin, 1993) during crowd disasters, the term crowd disaster in this research is defined as:

*"An incident in which people are injured or die caused by a high density of surrounding people."*

The inclusion criteria of the database are then developed based on the following definition:

I. The Incident must occur between the 1st January of 1989 and 31st December of 2021.
II. The incident must have been reported in either Chinese or English by a well-known news publisher. The choice of report languages was determined by the language proficiency of the researchers. A well-known news publisher is defined as a

professionally operated media organisation indexed in recognised news databases or officially registered with national media regulatory authorities.

III. People must have been either injured or died in the incident due to crowd force originating from the high density of surrounding people.

Based on the above inclusion criteria, "(crowd and (stampede or crush) and (die or injured) and (tragedy or disaster)) or 踩踏事件" is entered in the search form box of Factiva, and a date range from 01/01/1989 to 31/12/2021 is selected. The term "踩踏事件" means stampede in Mandarin and is entered to explore Chinese news articles in order to increase the inclusiveness of the database. Correspondingly, English, Chinese – Simplified and Chinese – Traditional are selected in the Language section. A screen shot is provided in Figure 5.1 for reproduction purpose.

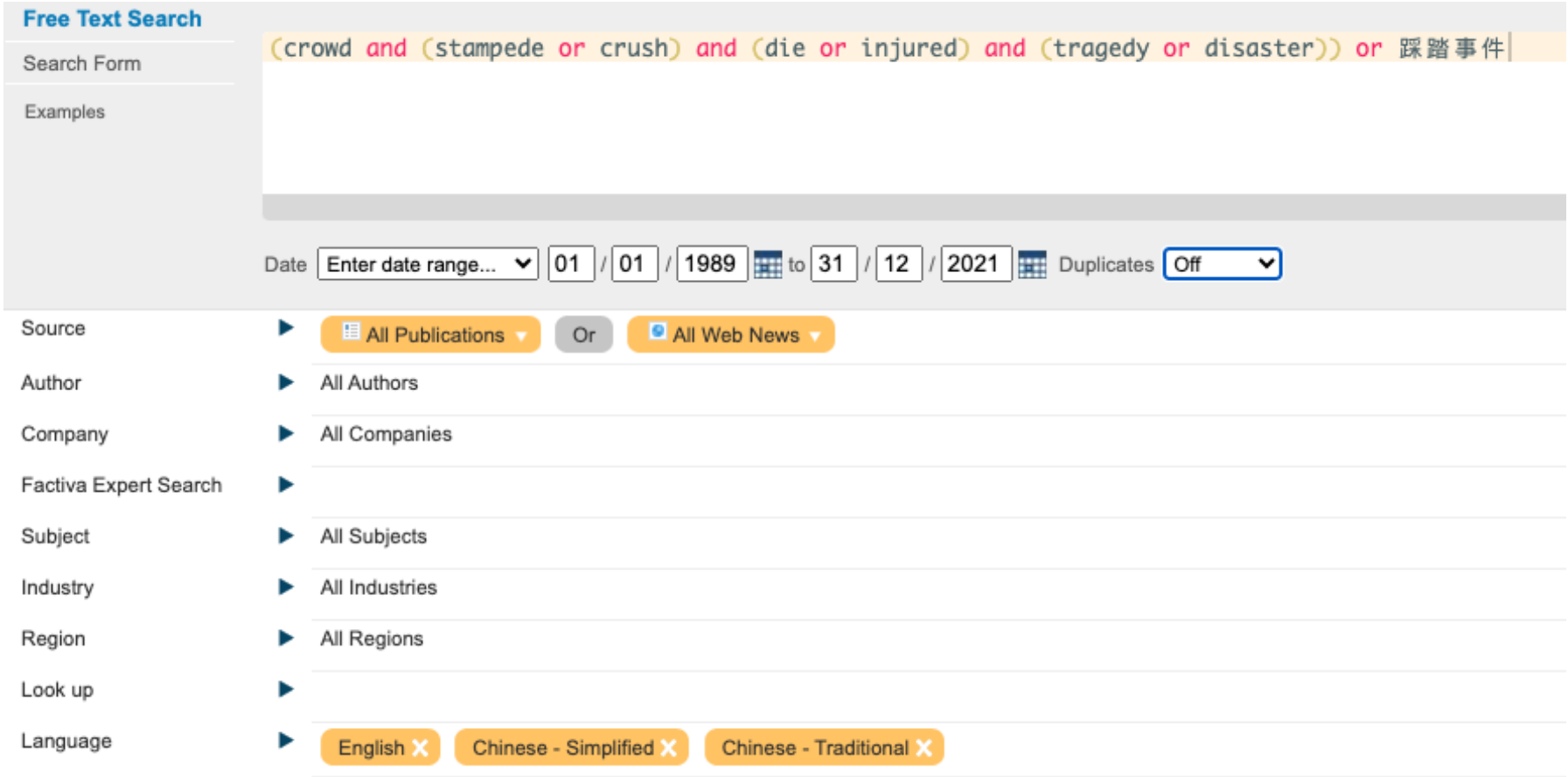



*Figure 5.1 Input in the Factiva search box (Yang, et al., 2024)*

Two supplemental Internet sources are also used in this research, Wikipedia and a website named ‘working with crowds’ (Working with crowd, 2023) that focuses on crowd safety management. These sources were not treated as primary data sources, but rather as initial identification tools to locate commonly recognised crowd disasters that may not be comprehensively indexed within news databases.

Although Wikipedia has been criticized for its credibility (Chesney, 2006) alongside with other Internet sources, they can serve as structured aggregators of publicly documented events. To ensure reliability, all incidents identified through these platforms were cross-verified against reports from well-known news publishers before inclusion in the database.

Incidents retrieved from above sources are further filtered following the inclusion criteria stated above and subsequently added to a list of crowd disasters. Depending on availability, a maximum of three news articles were downloaded for each crowd disaster to serve as the data source.

To maximise the reliability of news obtained from the Internet, the principle of maximum variation sampling is adopted. This approach involves maximising the number of news publishers and the variety of written languages based on availability. Optimally, an incident should be reported by at least three different news publishers in traditional Chinese, simplified Chinese, and English. When more than three news publishers are available, priority is given to the most recent publications and those from local news sources or with later publication date are prioritized.

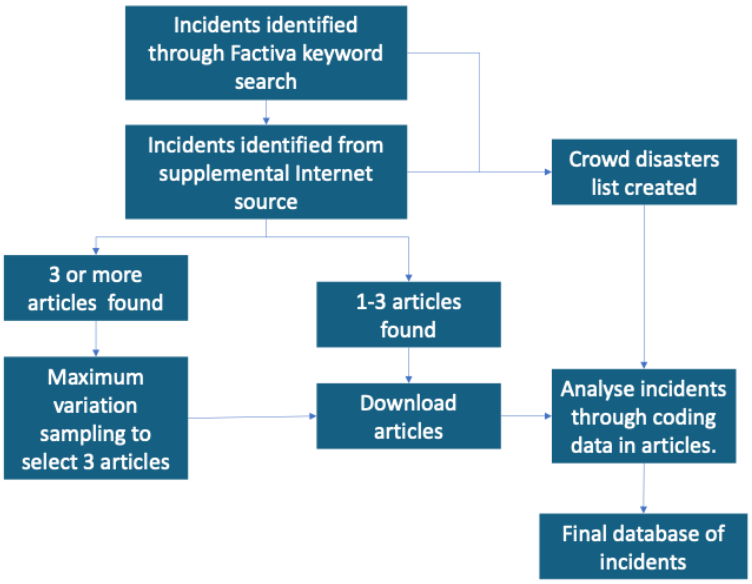



*Figure 5.2 Database data collection flowchart (Yang, et al., 2024)*

The process of adopting the principle removes the biases that might exist within a single news publisher, and ensures a more complete view as different publishers tend to report the incident from different angles. Finally, it facilitates the reliability of the data as the existence of each characteristic is reported by different data sources. The data collection procedure is illustrated in Figure 5.2.

Based on the above procedure (illustrated in Figure 5.2), a total of 293 crowd disasters was collected, in collaboration with another researcher. Out of the 246 incidents identified through Factiva keyword search, 47 are identified solely from Chinese news sources. Additional 47 incidents are identified through supplemental Internet sources.

### *5.3.1.2 Data Coding*

In order to systematically collect and compare characteristics of crowd disasters, data coding is utilised to extract data for the database. Data coding is a qualitative method of recording observations in a data format in which a defined taxonomy is used by researchers to denote specific characteristics as they are witnessed (Coates, et al., 2021). The process is explained below.

To construct the database, the content of the news articles for each incident is analysed by data coding, and its relevant characteristics data field are denoted by predefined subcategories. As stated above, the characteristics of crowd disaster entail social scenario, physical environment as well as the causal factors derived from the FIST model.

A code book (attached in Appendix II), composed of a comprehensive definition and taxonomy for the characteristics, has been constructed for the guidance of data coding, with terminology aligned with the Glossary for Research on Human Crowd Dynamics (Adrian, et al., 2025) where the definitions were consistent with the scope of this study.

The subcategory of each characteristic is developed based on indicators summarised in Table 2.3, literature review of relevant crowd management papers (Hsieh, et al., 2009; Lu, et al., 2021; Shao, et al., 2019) and the researchers' own observations during source identification. Due to the complexity of real-life situations and limitations related to news articles, in those cases where the certain characteristics of an incident fails to fit into any existing subcategory, a subcategory "other" is introduced. When certain characteristics of an incident cannot be identified, a "No' sub-category is added into the causal factor characteristics and an "Unknown" subcategory is added into scenario characteristics.

Numerous scenarios were identified during the source identification process. Scenarios with low frequency (i.e., fewer than five cases) were not classified as individual subcategories and were instead grouped under "Other." Seven types of social scenarios were selected due to their higher occurrence:

1) Sports event.
2) Performance.
3) Religion.
4) Jostle.
5) Organized event for celebrating a festival or memorable stuff other than religion.
6) Nightclub.
7) School.

Due to numerous physical environments present in crowd disaster, different taxonomies are proposed. For instance, International Organization for Standardization classifies building types

according to their functional usage. In contrast, the taxonomy adopted in this study is first structured according to whether the environment is indoor or outdoor, and outdoor environments are then further differentiated based on the degree of movement freedom available, reflecting the focus of this study on mobility. Four common types of physical environment have then been identified.

1) Stadium.
2) Building.
3) Open area.
4) Restricted area.

Six types of information have been identified:

1) Fire.
2) Imminent danger.
3) Unexpected incident.
4) Desirable 1, mainly describing a situation that things or items desirable for most participants are limited in amount or has to be done in a timely manner (e.g., discounted products during a sales event or limited free giveaways).
5) Desirable 2, mainly describing a situation that quality of things or items desirable for most participants is decreased with distance (e.g., obtaining a favourable viewing position near the stage at a music festival).
6) Official announcement.

As stated in section 2.4.1.3, all the subcategories of information must lead to a higher desired velocity, otherwise no information element is recorded. Previous research has also identified false alarm as a type of information leading to crowd disaster (Barr, et al., 2024). However, all six types of information identified above may also be transmitted in the form of a false alarm, provided that the crowd perceives the information as authentic. In this framework, a false alarm is therefore treated as a mode of information transmission rather than a distinct information category and is not coded separately.

Three types of physical space have been identified:

1) Staircase and escalators.
2) Entrance and exit.
3) Bridge and Tunnel.

Furthermore, due to the effect of desired velocity upon the critical density of physical space, such as arch formation phenomena introduced in section 2.3.1, each subcategory is divided into two situations: low desired velocity and high desired velocity.

As bidirectional pedestrian flow in a confined space has been a point of interest in this field (Flötteröd & Lämmel, 2015), especially the lane formation phenomena introduced in section 2.3.1, a further categorization is applied to represent both unidirectional and bidirectional flow. Consequently, a total of 17 subcategories are entailed under this characteristic (including "other" under 4 situations and "no").

Two types of time have been identified:

1) Ingress.
2) Egress.

These two subcategories are based on the researchers' observation during source identification and different space availability during ingress and egress and indicate long waiting times.

Four levels of casualty have been identified:

1) 0.
2) 1-10.
3) 11-50.
4) Over 50.

Detailed definitions for each subcategory are outlined in Appendix II.

### 5.3.2 Database analysis

#### *5.3.2.1 Data quality*

The three data quality dimensions utilised in Bateman & Majumdar (2020) for database assessment used in this chapter are outlined in turn below.

*Consistency*

This data quality dimension assesses the consistency and compatibility between the taxonomy schemes used for each characteristic and the raw data. Any characteristic data field that cannot be put into existing subcategories (i.e., that data field contains "other") is marked as 'inconsistent', and the overall consistency is calculated as:

$$consistency = (0.2N_{inconsistent} + N_{consistent})/N \tag{5.1}$$

Where $N$ is the number of relevant data fields. The result evaluates the ability of taxonomy schemes employed in the code book to capture characteristics, i.e., the percentage of data fields can be categorised under the taxonomy scheme.

*Completeness*

This data quality dimension assesses the proportion of missing data within the database. Each data field contains "Unknown" is treated as missing data as there always a scenario for each incident. The completeness is calculated as:

$$completeness = 1 - 1/N * \sum_{i=1}^{M} R_{missing}/R \tag{5.2}$$

Where $N$ is the number of data fields, $M$ is the number of characteristics and $R$ is the number of relevant data fields inside each characteristic.

*Accuracy*

This data quality dimension assesses the accuracy of data extraction process. As two researchers have been involved in the data coding process, an inter-rater reliability test (McHugh, 2012) is used to evaluate the agreement rate between the two coders. A high agreement rate ensures a sufficiently accurate standard has been applied in the data extraction process.

To perform the inter-rater reliability test, 15 incidents are randomly sampled by each coder out of the incidents coded by the other coder (representing approximately 10%) and coded again. The results are compared with the original data and the statistic Cohen's Kappa $\kappa$ (McHugh, 2012) is calculated as:

$$\kappa = (p - p_e)/(1 - p_e) \tag{5.3}$$

Where $p$ is the proportion of data fields both coders agree upon, and $p_e$ is the probability that agreement arose by chance, calculated as (Li, et al., 2023):

$$p_e = \sum_{j=1}^{n} p_{ij}\, p_{i'j} \tag{5.4}$$

Where $p_{ij}$ is the proportion of data fields that rater $i$ assigned to category $j$.

A value of $\kappa$ bigger than 0.8 can be viewed as almost perfect agreement and a value between 0.6 and 0.8 can be viewed as substantial agreement (Landis & Koch., 1977).

### *5.3.2.2 Cluster Analysis*

To address the objective of pattern identification, cluster analysis is employed here to identify different types of crowd disasters due to its wide application in the field of crowd analysis introduced in section 3.2.2.2. More specifically, the k-Modes algorithm is utilized in this chapter because all the data inside the database is categorical.

The k-Modes algorithm is a type of partitional clustering, which assigns each object to a cluster through minimizing the cost function between the object and cluster centroid, which serves as centres of each cluster (Rodriguez, et al., 2019). The cost function, as introduced above, is developed based on a matching dissimilarity measure, which calculates the 'distance' between categorical objects (Huang, 1998).

Let $X$ be a categorical object denoted by $[x_1, x_2, \dots, x_m]$, where $x_j \in DOM(A_j)$, for $1 < j < m$ and $A_j$ stands for jth attribute of object. For instance, a typical categorical object in this research is denoted as (sports events, stadium, no information, entrance and exit, ingress, 11-50).

The k-Modes algorithm looks for a way of dividing n categorical objects $X_n$ (denoted by a $k \times n$ $\{0,1\}$ matrix $W$) into k clusters (with centroids $Z_k$) that minimize the cost function $F_n$:

$$F_n(W, Z) = \sum_{l=1}^{k} \sum_{i=1}^{n} w_{il} d_n(Z_l, X_i) \tag{5.5}$$

Where $w_{il}$ is entries of $W$ and $d_n$ is the dissimilarity measure.

In this research, the dissimilarity measure between two objects used in Ng, et al. (2007) is utilised and it is defined as:

$$d_n(Z_l, X_i) = \sum_{j=1}^{m} \phi(z_{l,j}, x_{i,j}) \tag{5.6}$$

Taken the typical formulation in partitional clustering, where

$$\phi(z_{l,j}, x_{i,j}) = \begin{cases} 1, if\ z_{l,j} \neq x_{i,j} \\ 1 - |c_{l,j}|/|c_l|, otherwise \end{cases} \tag{5.7}$$

Where $|c_l|$ is the number of objects in $l$th cluster and $|c_{l,j}|$ is the number of objects in $l$th cluster contains same attribute $A_j$ as centroid $z_l$.

## 5.4 Results and discussion

### 5.4.1 Data quality

A total of 1758 datafields were tested based on the three data quality dimensions discussed above. For consistency and completeness, Dupuy (2012) suggested a minimum score of 60% to ensure sufficient data quality.

*Consistency*

Three out of six characteristics required the addition of an "Other" subcategory during the data coding process, including: social scenario, physical environment and physical space. A total of 112 datafields are labelled with "other" and hence the consistency of this crowd disaster database is calculated as 94.90%.

*Completeness*

Only five data fields within physical environment characteristics are labelled as "Unknown" since relevant information is missing in the identified data sources. Consequently, the completeness of this crowd disaster dataset is calculated as 99%.

*Accuracy*

There is perfect 100% agreement within "social scenario" and "casualty" characteristics between two researchers. The agreement rates and corresponding κ values are presented in Table 5.1 for the remaining four characteristics. All the characteristics besides "Physical space" have demonstrated a near perfect agreement between two researchers.

The slightly lower agreement rate in the 'Physical Space' category stems from the subjective nature of assessing desired velocity and the challenges in judging bidirectionality in pedestrian flow. Specifically, from the analysis, 2 out of 6 instances of disagreement within this subcategory were due to differing assessments of desired velocity, while 3 instances were related to disagreements on pedestrian flow directionality. However, this still indicates a

reasonably high level of accuracy of this crowd disaster database, and the effect of disagreement for directionality upon accuracy is further moderated as it is omitted in the data analysis process.

| | Physical Environment | Information | Physical Space | Time |
|---|---|---|---|---|
| Agreement rate | 96.67% | 90% | 80% | 96.67% |
| $\kappa$ value | 0.951 | 0.876 | 0.758 | 0.870 |
| Performance | Near perfect | Near perfect | Substantial | Near Perfect |

*Table 5.1 Accuracy of four characteristics (Yang, et al., 2024)*

### 5.4.2 Data Analysis

The optimal number of clusters k is determined to be nine through a silhouette analysis, which assesses the quality of cluster by comparing level of separation among clusters (Rousseeuw, 1987) based on the same dissimilarity measure $d_n$.

Table 5.2 demonstrates the outcome of the cluster analysis indicating the number of incidents in each cluster. Each cluster is identified numerically (e.g., Cluster 1, Cluster 2) rather than being named explicitly, in order to avoid premature interpretation and researcher bias. The distribution of subcategories within each characteristic and cluster are discussed below.

| Cluster | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Number of incidents | 25 | 24 | 24 | 24 | 50 | 43 | 25 | 19 | 59 |

*Table 5.2 Number of incidents within each cluster (Yang, et al., 2024)*

### *5.4.2.1 Social Scenario*

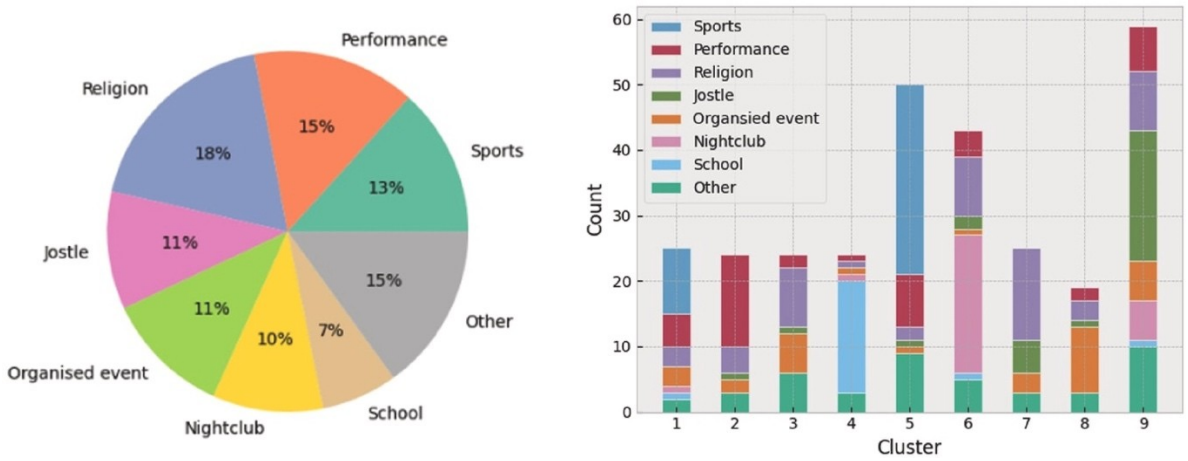



*Figure 5.3 Social Scenario Distribution (Yang, et al., 2024)*

Across the entire database, the most frequent social scenario identified for crowd disasters is that associated with religion, followed by performance, other and sports (See detailed distribution in Figure 5.3). The incidents classified as "other" under this characterization primarily refer to daily scenarios, including waiting for public transportation and walking on the street.

Within each cluster, religion is also the most frequent scenario within clusters 3 and 7, accounting for 37.5% and 56% respectively. The most frequent scenario within cluster 9 is "jostle", but it only accounts for 33.90%. It is the lowest occupancy rate for the most frequent scenario within a single cluster. However, it has overwhelmingly high occupancy rates for the rest of most frequent characteristics, indicating a weak correlation with social scenario for this type of crowd disaster.

It is also noticeable that although school scenario is the least frequent scenario across the entire database (only 7%), it has the highest occupancy rate within a single cluster, e.g. as high as 70.83% of cluster 4. This is mainly because most crowd disasters that occur in schools share unique and identical features which are discussed in the following section.

Overall, no dominant social scenario is observed within most clusters besides cluster 4. However, this does not indicate a mismatch between the proposed taxonomy and the clustered results, and modifying these categories after observing the clustering outcomes would

introduce outcome-driven adjustments and compromise the unsupervised nature of the analysis. (Hastie, et al., 2009). This principle is maintained across all subsequent categories.

### *5.4.2.2 Physical Environment*

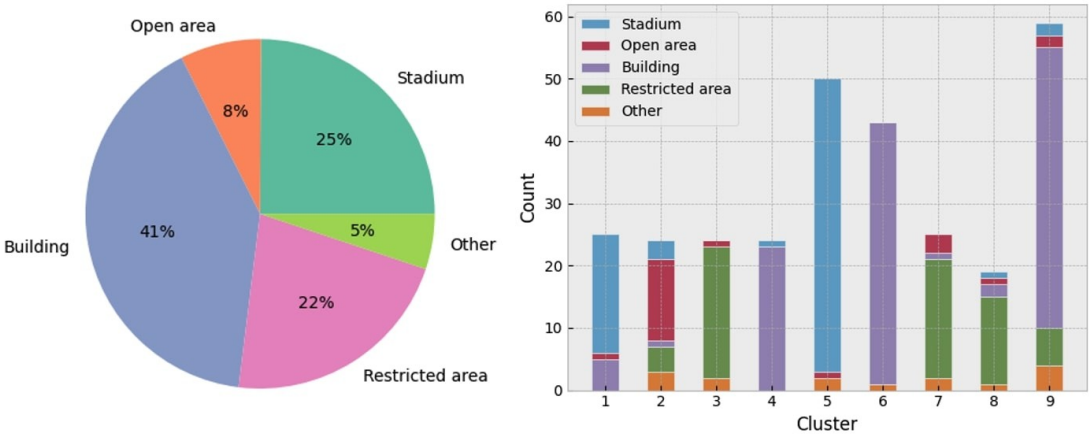



*Figure 5.4 Physical Environment Distribution (Yang, et al., 2024)*

Figure 5.4 demonstrates that approximately 66% of crowd disasters occur in indoor spaces (stadium and building), while only 30% occur in outdoor space (open space and restricted space). The incidents classified as "other" under this characteristic primarily refer to complex environments that cannot be easily categorised as a single category which only accounts for lower than 5%.

Building is the most frequent environment within clusters 4, 6 and 9, demonstrating overwhelming occupancy rates of 95.83%, 97.67%, 76.27% respectively. Unsurprisingly, there is a strong correlation between stadium and scenario sports, as both are the most frequent scenario and environment within clusters 1 and 5. Although there is a low proportion of open area across the database, it is still the most frequent environment within cluster 2, accounting for 54.17%. This is mainly because both the information and physical space that lead to crowd disasters in cluster 2 can barely be found in other clusters.

### *5.4.2.3 Information*

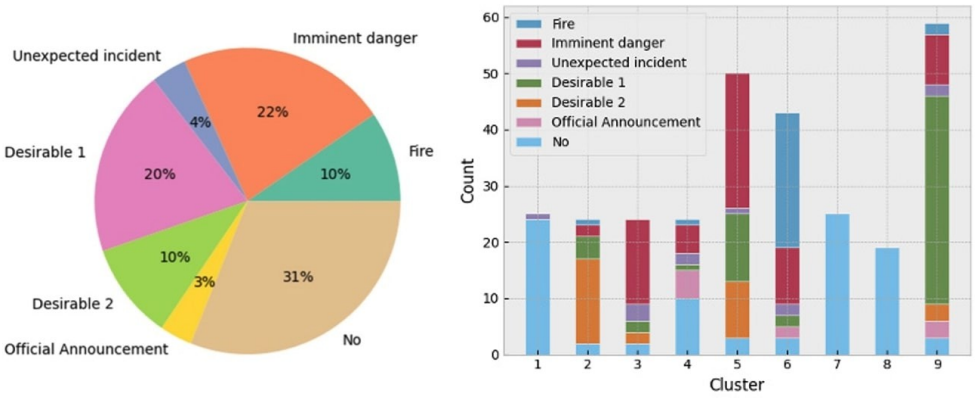



*Figure 5.5 Information Distribution (Yang, et al., 2024)*

A total of 69% of crowd disasters have been identified, entailing information causal factors. Very few of them are caused by an unexpected incident and official announcement (accounting for 4% and 3% respectively), while desirables and dangers are approximately equal in proportion.

On the other hand, three clusters have overwhelming occupancy rates of "no information" causal factor, with two 100% (cluster 7 and 8) and one 96% (cluster 1). The other cluster that has no information causal factor as its highest occupancy rate subcategory is cluster 4, which accounts for 41.67%. This is in contrast with the most frequent physical space for this type of crowd disaster (high desired velocity at staircase and escalator according to Figure 5.5). This indicates that a large number of people hold a high desired velocity without any external information given. The reason is primarily due to the most frequent scenario within the same cluster "school" refers mostly to primary school, while most victims are students.

Primary school children are viewed as being easily impatient due to their mental and physical developmental stage, and thus a high desired velocity. This poses a significant risk on staircases, especially since they often walk in groups during school time. It is also noticeable that "official announcement" has the second highest occupancy rate within this cluster (20.83%), and this typically refers to teachers' guidance in asking students to move quickly to reach destinations more promptly.

### *5.4.2.4 Physical Space*

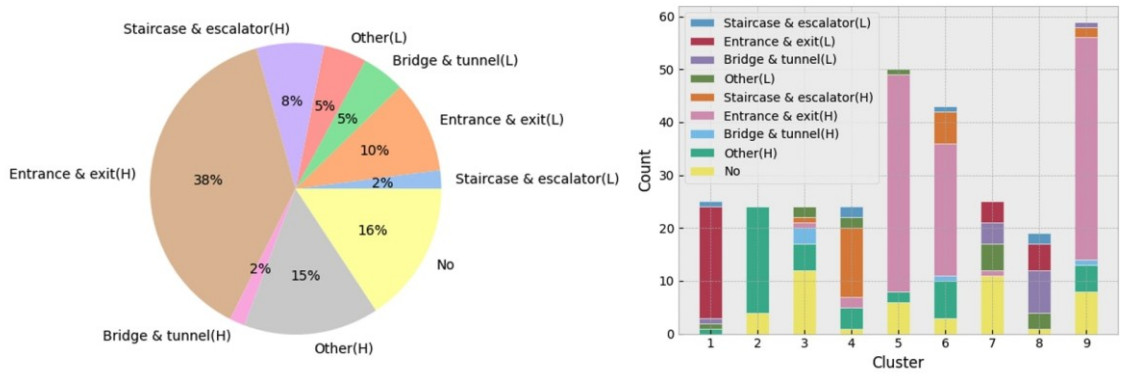



*Figure 5.6 Physical Space Distribution[2] (Yang, et al., 2024)*

In total, 16% of crowd disasters are identified as entailing no physical space causal factors. Among the remaining crowd disasters, only 13.41% contain bidirectional pedestrian flow. Due to its unbalanced proportion compared to unidirectional flow, as well as the significantly increased data dimensionality when accounting for it, the directionality of pedestrian flow is omitted in the cluster analysis and following discussion.

Figure 5.6 shows that 63% of crowd disasters are caused by physical space under high desired velocity, while low desired velocity only accounts for 22%. Entrance and exit have been identified as the most common locations for crowd disasters both within high and low desired velocity.

Other physical space under high desired velocity is the most common place for crowd disaster within cluster 2, accounting for 83.33%. This typically refers to fences around a stage or the stage itself, as most crowd disasters within this cluster occur during a performance, and the appearance of the performer(s) serves as information that incentivises participants to get as close as possible. This aligns with the definition of desirable 2, which holds the highest occupancy rate within this cluster (62.5%).

Further correlations between information and physical space are discovered. When presented with fire or other imminent danger within a stadium or other enclosed building (cluster 5 and cluster 6 respectively), people tend to flee as fast as they could, and such high desired velocity

[2] L stands for low desired velocity and H stands for high desired velocity.

often leads to severe congestion at the exit, which are the most frequent physical spaces, accounting for 82% and 58.14% respectively. Cluster 9 also has entrance and exit as its most common physical space, and is also driven by the external information that some desired items are offered with limited quantity (desirable 1) inside an enclosed building. People tend to get into the building as fast as possible to secure their possession of the item, thus leading to congestion at the entrance.

Cluster 7 is the only one that has no physical space causal factor as its highest occupancy rate subcategory, accounting for 44%. This indicates the already dense nature of the participants that require no deterioration from the physical space, and it is sourced from its most frequent scenario, religion, that normally no measures can be taken to limit the number of arrivals.

#### *5.4.2.5 Time*

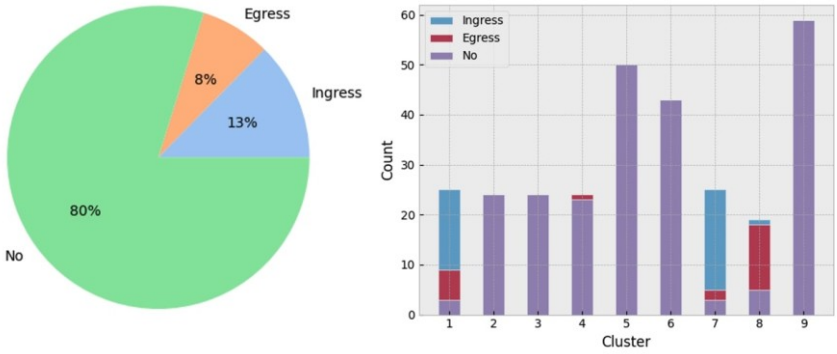



*Figure 5.7 Time Distribution (Yang, et al., 2024)*

Unlike other causal factors, time, i.e. whether there is a long waiting time at the time of the incident, has not been explicitly identified within many crowd disasters, accounting for only 21% of the entire database. Nevertheless, it still contributes considerably towards clusters 1, 7 and 8. Ingress has been identified in 70% crowd disasters within cluster 1 and 80% within cluster 7, while egress accounts for 73.68% of cluster 8. This indicates a distinct difference in the underlying causes among these three clusters in contrast to the others, as the incidents originated from mismanagement of flow rate, and could potentially take place at any physical space inside a building during organized event, as revealed by cluster 8.

The high occupancy rate of ingress within cluster 1 is mainly caused by low admission rates at stadium entrances, which leads to the impatience of crowds and queuing effects. A similar

situation also takes place within cluster 8, where the slower than expected flow rate is caused by physical space constraints (e.g., bridge and tunnel) during egress. As discussed later, such findings relating to ingress and egress enable targeted crowd management measures.

### *5.4.2.6 Casualty*

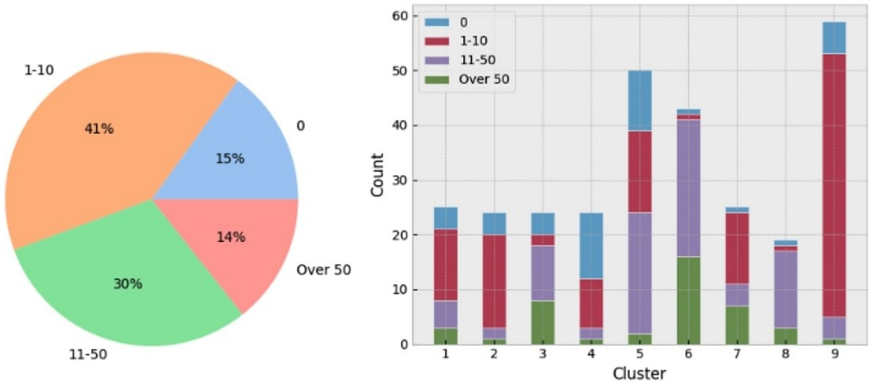



*Figure 5.8 Casualty Distribution (Yang, et al., 2024)*

Fortunately, casualties during most crowd disasters are below 10 persons. However, 14% of the crowd disasters still result in fatalities exceeding 50 persons, a significant concern for society.

The crowd disaster with highest fatality is in cluster 6, where 58.13% of the incidents lead to fatalities exceeding 50 persons. It is also noteworthy that many victims die from fire rather than crowd force within this cluster, and it is challenging to achieve a precise statistic to differentiate them.

Crowd disasters within cluster 4 share the highest non-fatal rate among all the nine types, reaching as high as 50%. This is assumed to originate from most victims, as they are still primary school students and have much less body mass than adults, thus less crowd forces are exerted. Future research should investigate further this type of crowd disaster in order to facilitate crowd management and building architecture design within schools.

In summary, the centroid of each cluster is presented in Table 5.3 (L and H denote low and high desired velocity as above). Crowd disasters within the same cluster are considered as the same type, given that they are more similar with each other with respect to their characteristics

based on the cluster analysis. For instance, type 2 crowd disaster tends to take place at a performance event, e.g., concert, music festival ('Social Scenario') hosted in an open area ('Physical Environment'), and it is mainly caused by crowds rushing towards other physical space (e.g., the stage) due to the presence of desirable 2 (e.g., the appearance of performer). The 'Time' element is not presented in this type of crowd disaster, and once it occurs, there is a high likelihood of resulting in 'Casualty'.

| | Social Scenario | Physical Environment | Information | Physical Space | Time | Casualty |
|---|---|---|---|---|---|---|
| Type 1 | Sports | Stadium | No | Entrance and exit(L) | Ingress | 1-10 |
| Type 2 | Performance | Open area | Desirable 2 | Other (H) | No | 1-10 |
| Type 3 | Religion | Restricted area | Imminent danger | No | No | 11-50 |
| Type 4 | School | Building | No | Staircase and escalator(H) | No | 0 |
| Type5 | Sports | Stadium | Imminent danger | Entrance and exit(H) | No | 11-50 |
| Type 6 | Nightclub | Building | Fire | Entrance and exit(H) | No | 11-50 |
| Type 7 | Religion | Restricted area | No | No | Ingress | 1-10 |
| Type 8 | Organized event | Restricted area | No | Bridge and Tunnel (L) | Egress | 11-50 |
| Type 9 | Jostle | Building | Desirable 1 | Entrance and exit(H) | No | 1-10 |

*Table 5.3 Types of crowd disaster (Yang, et al., 2024)*

Based on the results, event organisers can develop a more explicit crowd management strategy, as they are able to locate most common causal factors of crowd disaster based on the event scenario and the environment. For example, for a sports event that is hosted in a stadium, which shares same social scenario and physical environment with type 1 and type 5 crowd disaster, 'Guide to Safety at Sports Grounds' (SGSA, 2018) is commonly used to access the safe capacity and conduct safety planning for sports event in U.K. The safety capacity is calculated for every single part of the stadium, and extra work is encouraged for the entrance and exit capacities. This is due to type 1 crowd disasters usually being caused by a long waiting time during ingress at the entrance, and type 5 usually takes place because of rushing towards the exits during an emergency evacuation when imminent danger presents itself.

For other events that share similar scenario and environment settings with any type of crowd disaster, the organisers are encouraged to allocate more resources around such physical spaces, pay attention to the critical moment when certain information might emerge, and/or control the flow of ingress/egress.

## 5.5 Summary

This chapter has introduced the methodology behind the creation of a crowd disaster characteristics database, containing 293 crowd disasters worldwide between 1989 and 2021.

The main data source is from online news database Factiva (Dow Jones & Reuters, 2023), combined with supplemental Internet sources. The characteristics included in the database are retrieved from the FIST model (Fruin, 1993), combining with social scenario, physical environment and causality whose importance is shown by the case study of the 2014 Shanghai crowd disaster in section 2.4.2. To the best of the author's knowledge, this database is the most well-defined database for such incidents as it presents a structured taxonomy of crowd disasters.

The principle of maximum variation sampling is adopted during the data collection process. Data quality dimensions consistency, completeness and accuracy have also been utilised to validate the consistency of proposed characteristics taxonomy, the completeness of data and the accuracy of data coding. Such a data quality assurance procedure is not usually seen in the literature. A further data analysis has also been conducted using cluster analysis, and the K-Modes algorithm, which specifically focuses on categorical data, has identified nine clusters.

Based on the cluster analysis results, the distribution for each characteristic across the entire database and within each cluster has been discussed, and distinct features have been identified so that each cluster forms a unique type of crowd disasters. This classification forms the foundation for the subsequent chapters of this research, as it closely aligns with the principles of the mesoscopic approach. It enables the identification of scenarios that exhibit shared behavioural patterns while preserving distinct characteristics relative to alternative scenarios.

In the subsequent chapters, type 9 crowd disaster, which mainly occurred during the jostle scenario, is selected to be further investigated, because of its highest frequency. Moreover, emergent arch formation phenomena are commonly observed during this scenario, reflecting the complex system nature of crowds. As such, the jostle scenario serves as a representative scenario for studying crowds from a socio-technical system perspective.

# 6. Virtual reality experiments

## 6.1 Introduction

Following the results from the last chapter, this chapter investigates the jostle scenario, which describes a non-emergency daily situation where all individuals in the crowd rush into the building to retrieve desirable items.

An effective crowd management strategy requires an accurate understanding of crowd behaviour, which in turn relies on robust crowd analysis. Common crowd analysis methods, introduced in chapter 3, are reliant on the quality of the input data (Zhou, et al., 2010). However, available video datasets are suitable mainly for counting members but are considerably limited in behavioural analysis (Khan, et al., 2020), whereas physical experiments, on the other hand, face numerous challenges especially safety concerns when exploring certain scenarios (Moussaïd, et al., 2016).

Virtual Reality (VR), in contrast, provides experimentally controllable environment that overcomes key limitations of video footage-based crowd data (Kinateder, et al., 2018). It is a powerful tool in scientific research given its flexibility (Valmorisco, et al., 2024) and cost efficiency (Vizzari, 2020). Due to its ability to reproduce scenarios involving threats to life without causing any actual physical harm to participants, it holds considerable promise as an ideal alternative to physical experiments (Lawson, et al., 2019; Xu, et al., 2024). However, in order to realise this potential, it is necessary to carefully design scenarios that are ecologically valid.

In this chapter, the jostle scenario is investigated through VR experiments. To the best of the author's knowledge, this thesis conducts the largest scale VR-based experiment in crowd analysis by reproducing the jostle scenario in a virtual environment. Such an arrangement provides manually controllable experimental settings and enables both a systematic and reproducible crowd behaviour experiment. Three rounds of experiments were conducted in this environment, each involving over 22 participants in the same scene.

Two types of individual behaviour under the jostle scenario were then identified, namely patient and impatient, by analysing the patterns of crowd movement. This finding challenges the conventional classification adopted in crowd management practice that only considers normal and emergency behaviour and emphasises the need for a deeper understanding of crowd behaviour.

Section 6.2 reviews the relevant literature and identifies current research gaps, section 6.3 summarizes the setup and the conduction of the experiment, section 6.4 introduces the methodologies for data analysis, section 6.5 discusses the results and section 6.6 concludes the chapter.

## 6.2 Literature review

A typical crowd disaster under jostle scenario took place in 2008, when a Black Friday sales event at a Walmart in Valley Stream, New York, turned chaotic that over 2000 shoppers pushed through the glass door before opening time. Workers inside attempted to stop the crowd, and tragically, one worker was trampled to death (McFadden & Macropoulos, 2008). Furthermore, as chapter 5 indicates, this type of crowd disaster remains prevalent and was identified as the most frequent crowd disaster type. To the best of the author's knowledge, no paper in the public domain has investigated the jostle scenario in depth, providing an ideal scenario to explore.

Physical experiments conducted to reproduce similar scenarios are often faced with safety concern, as high desired velocity could lead to violent physical contact. Consequently, the behaviour of participants is normally restrained to prevent causing damage (Garcimartín, et al., 2016). Alternative approaches therefore need to be explored in order to allow more natural behaviour without causing safety concern, and VR has emerged as a perfect candidate (Lawson, et al., 2019; Xu, et al., 2024).

The usage of VR in scientific research has been seen to be valid across different domains including education, neuroscience and clinical study (Pietro, et al., 2018). In the field of crowd analysis, VR has become an essential tool for examining crowd behaviours during emergency evacuations, including wayfinding (Feng, et al., 2021; Dai, et al., 2024) and exit choices (Zhang, et al., 2025; Wang, et al., 2025; Chen & Li, 2025; Feng, et al., 2021). Beyond evacuation

contexts, VR has also been employed to investigate pedestrian crossing behaviour, both in terms of pre-crossing waiting time (Kalatian, et al., 2019) and interactions with vehicles during crossing (Wang, et al., 2021; Shipman, et al., 2022; Cavallo, et al., 2019).

In terms of validity of the VR experiments results, Olivier, et al. (2018) showed that the locomotion trajectory in VR closely mirrors those observed in real environments within certain VR scenes. Similarly, Moussaïd, et al. (2016) successfully demonstrated that crowd behaviour can be replicated in VR especially under certain mental settings and Beek, et al. (2024) found comparable crowd decision making between field study and matching VR experiment. Furthermore, Juřík, et al. (2023) illustrated that data obtained from VR experiment can be used to calibrate an agent-based model for a better accuracy.

On the other hand, flexibility is considered as a major advantage of VR, as it enables a completely bespoke environment setting (Valmorisco, et al., 2024) while ensuring complete control over the conditions (Vizzari, 2020; Kinateder, et al., 2018). Despite these advantages, recent studies using VR in crowd analysis are primarily concerned with individual participants at given points in time (Li, et al., 2025), investigating their interactions with built environment (Vizzari, 2020), algorithm controlled agents (Zhang, et al., 2023) and agents based on gathered data (Yin, et al., 2022). Very few studies have explored VR-based multi-participant experiments (Xia, et al., 2024; Chen & Li, 2025), leaving the potential of using VR to study interactions among crowd (Kinateder, et al., 2018) unexplored.

To address the above gaps, the objectives of this chapter include:

- Evaluate the feasibility and effectiveness of conducting multi-participant VR experiment in jostle scenario, particularly involving a larger number of participants than typically studied in existing crowd analysis research.

- Identify patterns of crowd movement through locomotion trajectory analysis within the jostle scenario.

## 6.3 Experimental Methodology

### 6.3.1 Experiment setup

During the database construction in chapter 5, it was noted that most instances of a jostle crowd disaster occur during a shop sales event, such as the 2008 Valley Stream crowd disaster introduced above. Typically, during such events, most members of a crowd aim to enter a building, in this case a shop, to purchase discounted products that are typically limited in quantity. Consequently, our experiment aims to reproduce such an event in the virtual environment, while eliciting participants' natural decisions and actions. This approach seeks to ensure behavioural realism, defined as the extent to which participants behave in VR as they would in the equivalent real-world situation (Arias, et al., 2021).

Using socio-technical system framework, the elements of the sales event are identified as:

- People: Participants of the sales event do not share common demographic characteristics.

- Goal: The goal of the participants is to enter the building faster than others in order to gain an advantage in retrieving the desired items.

- Procedure: Participants can start entering the building once the gate is open.

- Infrastructure: Physical infrastructure within the scene includes a narrow gate and wall of the building.

- Technology: An announcement system is needed to communicate with the participants, especially announcing the start and end of the event.

- Culture: Queuing is not mandatory, and overtaking is permitted.

Based on above elements, the virtual environment was developed in Unity (Unity Technologies, 2025) version 2022.3.24 using Virtual Reality Student Experience (VIRSE), a platform developed by Imperial College London (2025). A feature of VIRSE is that it enables the simplification of a multiplayer VR development by designating one client as the "host" to maintain the true state of the world, while other clients run predicted states. The host continuously updates the non-hosts, allowing participants to interact seamlessly in the same virtual environment, and thereby making VIRSE ideal for simulating crowd behaviours. Furthermore, VIRSE includes a real-time voice chat function, which can serve as the announcement system.

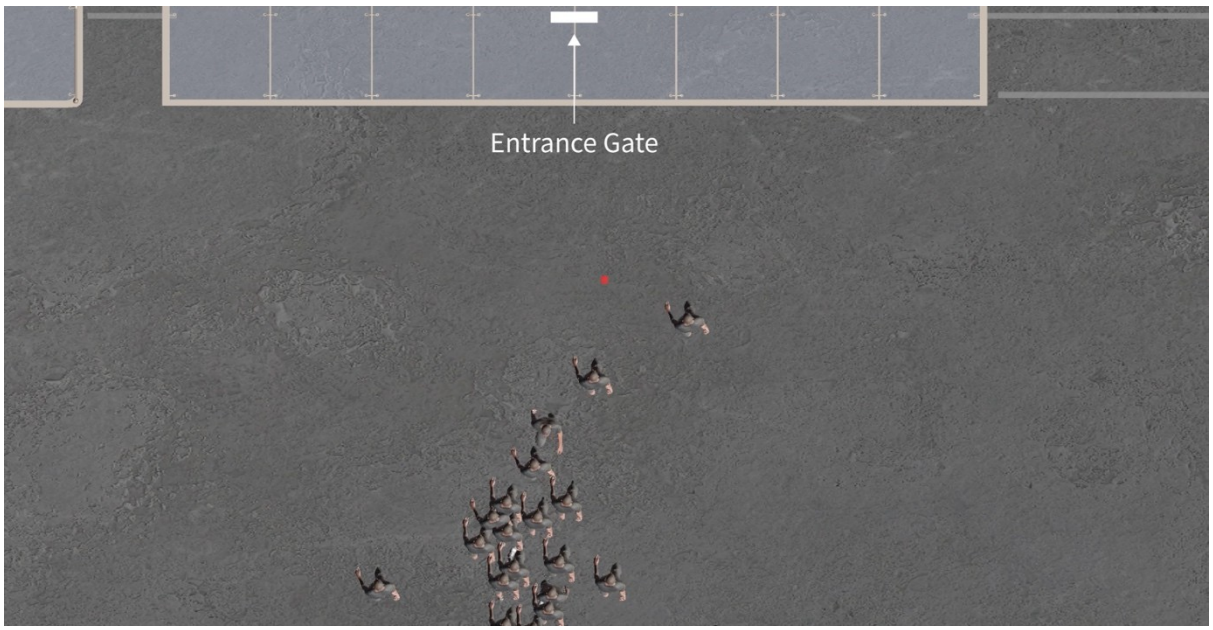



*Figure 6.1 Layout of virtual environment*

The virtual environment consists of a real-world-sized supermarket building with a controllable entrance door, surrounded by an open area. The layout of the virtual environment from a bird's-eye view is presented in Figure 6.1. Although the actual open area is substantially larger to greater movement freedom, only a 20 m × 10 m region is presented. The interior space of the supermarket was left empty as only the crowd behaviours prior to entry are of interest. To reproduce the goal of the sales event, all participants were instructed to pass through an entrance gate (as marked in Figure 6.1), restricted to one individual passing through at any one time, as quickly as possible (demonstrated in Figure 6.2 (a)).

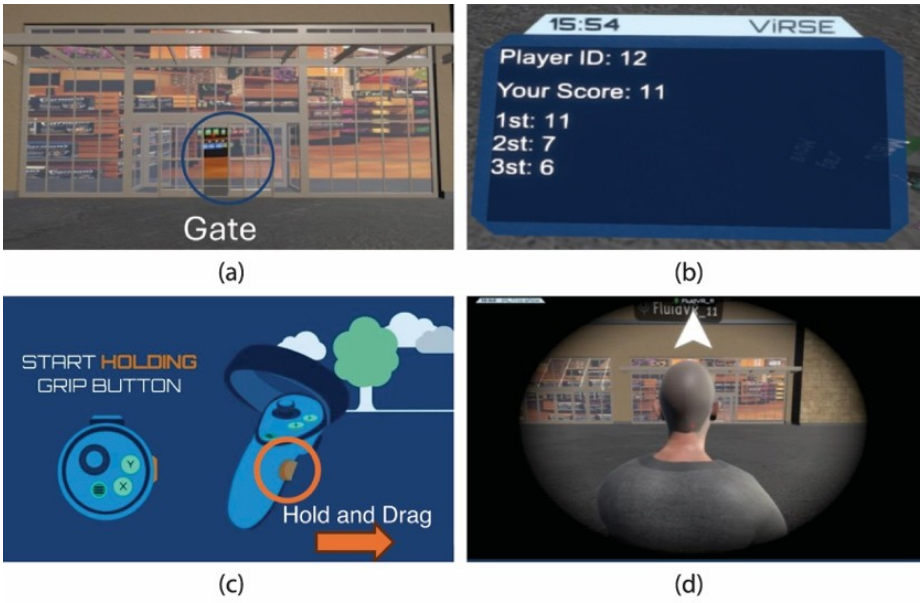



*Figure 6.2 Experiment setup, (a) the target gate of the participants; (b) score board; (c) avatar control; (d) indication of physical contact*

To fully exploit the virtual environment for data collection, once a participant passes through the gate, they are teleported back to a random location on an arc that is 10 meters away from the entrance, with a span of 60 degrees. The participant then rejoins the crowd and again attempts to enter the gate as quickly as possible. 10 meters was chosen as the appropriate distance to allow participants to rejoin the crowd quickly whilst providing sufficient distance to adjust their strategy, such as choosing the entry point into the crowds. The 60 degrees span was chosen to simulate a real-world condition in which most people typically approach the gate from approximately the same direction, e.g., from the parking lots of the supermarket.

Previous research has demonstrated that in VR scenarios, leaderboards are particularly effective in fostering competition (Banuri, et al., 2024; Marciani, et al., 2026) without negatively impacting the participants' intrinsic motivation (Mekler, et al., 2013). Therefore, a leaderboard-based game mechanic is incorporated to further stimulate the competitiveness of participants. Each time a participant successfully passes through the entrance gate, one point is assigned to their score. Participants can access their scoreboard anytime, which displays their current score along with the top three scores (demonstrated in Figure 6.2 (b)). The introduction of gamification should alter participant behaviour (Ourdas & Ponis, 2023). However, as

competitiveness is an inherent feature of the jostle scenario, the leaderboard was designed to reinforce this competitive context rather than introduce artificial behavioural distortion.

The character model for the avatar is downloaded from Mixamo (Adobe, 2025), an online platform that provides 3D (three-dimensional) character animation. Since this research focuses on analysing locomotion trajectory, all the avatars share the same appearance and thereby excluding the effect of demographics. The avatar collider was defined as a circle to approximate the human body, with a diameter of 0.4 m, adopting from the impact size in CA (Li, et al., 2019). Collisions between avatars were handled using Unity colliders without dynamic physics response, thereby preventing participants from physically pushing against one another. This design choice was made avoid exaggerated or abusive pushing behaviour that may arise in virtual environments, where participants are aware that no real physical harm can occur. Furthermore, since this study looks at behaviour at individual level, and as demonstrated in section 2.4.1.1, in dense crowds, physical pushing can induce loss of balance and involuntary displacement, meaning that observed motion may no longer reflect an individual's behavioural intention but instead the resultant of collective physical forces.

Each participant controls their avatar using an Oculus Rift-S VR headset. Only the left controller is used for movement control, and the direction of movement is determined by holding the grip button and dragging towards the desired direction (as demonstrated in Figure 6.2 (c)). The movement distance is set at three times the distance of the hand dragging motion. If the controller moves from wrist to elbow a distance of approximately 25 cm (Gordon, et al., 1989) twice in one second, corresponding to a typical walking cadence (Tudor-Locke, et al., 2018), the resulting virtual speed is equivalent to approximately 1.50 meters per second, which matches the average walking pace of an apparently healthy adult (Murtagh, et al., 2021). Participants retained the freedom to adjust their walking speed by varying the drag distance or movement frequency.

Furthermore, the sensation of physical contact is an important feature that needs to be reproduced in the virtual scenario, as it is unavoidable within the crowd. As the VR headset is unable to provide physical feedback, a substitution based on Nanay's (2018) demonstration have been identified that one sense can replace another. This substitution can generate mental imagery that replicates the sensation of the missing sensory input.

A visual cue is therefore implemented as a substitute for physical contact. In particular, a black border filter that narrows the participant's field of view, appears on the interface when a controlled avatar has physical contact with others, see Figure 6.2 (d). This effect mimics the sensation of crowd pressure by visually restricting the peripheral vision. In this experiment, physical contact is defined as the distance between the centre points of two avatars less than 0.4 meters, i.e., when their colliders are in contact. Additionally, a white arrow appears on the interface, indicating the direction, up, down, left, or right, of the physical contact. The implementation of lateral arrows can also serve as compensation for the reduced field of view in VR headsets (Xiao, et al., 2025).

This experiment was approved by Science, Engineering and Technology Research Ethics Committee of Imperial College London under application ID 6580476.

### 6.3.2 Experiment execution

Three experimental rounds were conducted, each involving over twenty participants, with the sample size determined by the capacity of the VIRSE lab. The participants were recruited through posts shared on a university mailing list and social media. Since participants in the jostle scenario did not exhibit any distinct demographic patterns, no inclusion criteria were defined, aside from the ability to read English and an age requirement of 18 to 35. The age requirement was set because competitiveness varies across different age groups (Mayr, et al., 2011). No demographic data of participants were recorded as discussed above.

The protocol for each round is as follows:

i) participants initially underwent a tutorial session to familiarize themselves with the settings, approximately 10 minutes in duration;

ii) followed by a break of 10 minutes.

iii) Data were only collected in the main session after the break, and participants were free to leave earlier or take breaks any time after being teleported. The main session lasted for approximately 10 minutes;

iv) Each participant was asked to complete a questionnaire following the main session.

All the data collected in this experiment were fully anonymous, i.e., impossible to trace back to the actual individual, with a randomly generated identification (ID) assigned to each avatar. Under each ID, position information (x,y coordinates), as well as the physical contact status, in the form of a 4-dimensional binary array, are recorded every 0.5 seconds for the entire duration of the main session.

The questionnaires were filled in using the same ID. Following similar approaches taken by Arias, et al. (2021), participants need to rate their level of immersion and realism based on their experience during the experiment. Although visual realism does not necessarily imply behavioural realism, subjective realism and immersion ratings are commonly used as practical proxies for assessing the realism of VR scenarios (Arias, et al., 2021). Additionally, participants were asked to rate the effectiveness of the visual cue as substitution for physical contact. All three questions are answered in a scale of 1 – 10, with 1 indicating "not at all" and 10 indicating "extremely". The original questions are attached in the Appendix III.

Data in this experiment was registered through the Data Asset Registration Tool of Imperial College London in compliance with General Data Protection Regulation (GDPR).

## 6.4 Data Analysis Methodology

### 6.4.1 Data preprocessing

As stated in section 6.3.2, locomotion trajectory data were collected in the form of (x,y) coordinates every 0.5 seconds for each participant under a unique random ID. The data were further divided into subsequent trajectories. Each trajectory is considered complete either upon passing through the entrance gate or at the end of the main session, excluding the time needed for teleportation or taking breaks.

To analyse the patterns of movement, a decision-making perspective is taken so that two behaviour measures, change of speed $\Delta v$ and change of direction $\Delta \omega$, are extracted from the

position data for every 0.5 second interval. The rationale for the choice of change of direction $\Delta\omega$, over the direction $\omega$ itself, is to exclude the impact of initial positions, as they are randomly decided. The change of speed $\Delta v$, which can also be seen as either acceleration or deceleration, is utilised to reflect responses to other participants.

Both $\Delta v$ and $\Delta\omega$ are extracted from the locomotion trajectory:

$$\Delta v_t = v_t - v_{t-1} \tag{6.1}$$

$$\Delta\omega_t = \omega_t - \omega_{t-1} \tag{6.2}$$

With:

$$v_t = \frac{\sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}}{0.5} \tag{6.3}$$

$$\omega_t = \arctan\left(\frac{y_t - y_{t-1}}{x_t - x_{t-1}}\right) \tag{6.4}$$

$\omega_t$ is measured in the units of radians with respect to the positive x-axis, resulting in positive values when facing towards the entrance gate and negative values when facing away from it. The physical contact data are saved for further analysis in the next chapter.

The trajectory data $T_t$ , which starts from time $t$ with length $2n$, is formatted as follows:

$$T_t := \left[[\Delta v_t, \Delta\omega_t], [\Delta v_{t+0.5}, \Delta\omega_{t+0.5}], \dots, [\Delta v_{t+n}, \Delta\omega_{t+n}]\right] \tag{6.5}$$

All the trajectory data with length shorter than nine are excluded from further analysis due to the limited information they can provide.

Since multiple rounds of experiments were conducted, consistency among different experiments needs to be checked to ensure that participants demonstrate similar behaviour. Additionally, as each participant must perform the same task repetitively, fatigue and boredom may affect their behaviour (Wolff, et al., 2021). Therefore, consistency across different time periods also needs to be examined.

In this thesis, similar behaviour is defined as the trajectory data $T_t$ coming from the same distribution, as it suggests that individuals exhibit comparable patterns of acceleration, deceleration, and turning behaviour, reflecting a consistent scale and style of motion in response to others or the environment. Statistical tests are often used to check if multiple independent groups of data coming from the same distribution (Conover, 1999), and thus are employed in this research to assess consistency.

### 6.4.2 Change point detection

Since this thesis analyses movement patterns within the crowd, trajectory data $T_t$ requires further segmentation to exclude any behaviour that is deemed "outside of the crowd". As introduced in section 2.4.1.1, as the density increases, a sudden transition from laminar to stop-and-go flow is observed for pedestrian movement. Such transition indicates a change of behaviour as pedestrian moves steadily during laminar flow, while a more frequent adjustment during stop-and-go flow. Consequently, an abrupt change of behaviour (in terms of speed and direction) is set as an indicator for entering the crowds.

The arch forming phenomenon, introduced in section 2.3.1, was observed under the jostle scenarios during the database construction in chapter 5, and is often described as emergent. Therefore, a successful reproduction of the jostle scenario should involve the spontaneous emergence of arch forming, as participants only control their avatar's movement at a microscopic level. This is further justified by the structural similarity between the jostle and evacuation scenarios, both of which involve passing through a narrow gate.

On the other hand, the arch formation is often correlated with high density and a change of behaviour pattern (Hoogendoorn & Daamen, 2005). Consequently, upon successful reproduction of arch formation, 'within the crowd' is defined as the region enclosed by the emergent arch, and the radius of this arch is named as the crowd radius. At the periphery of the arch, the most abrupt change of behaviour is expected to occur.

In the field of signal processing, it is often necessary to identify points for abrupt changes of underlying states, referred to as the change point detection (Truong, et al., 2020). This

methodology is applied in this paper to determine the crowd radius by sorting behaviours ($\Delta v$ and $\Delta\omega$) based on the distance from the entrance. The point at which an abrupt change is detected is defined as the crowd radius.

The Python package "ruptures" (Truong, et al., 2020) is imported here to exploit the built-in change point detection algorithm. A window-based search algorithm is chosen due to its ability to focus on local changes (Aminikhanghahi & Cook, 2017). Within each segmentation $[a, b]$, a L2 cost function is utilized due to its emphasis on mean-shifts (Truong, et al., 2020):

$$Cost = \sum_{d=a+1}^{b} \left( \| \Delta v_d - \overline{\Delta v_{a,b}} \|_2^2 + \| \Delta \omega_d - \overline{\Delta \omega_{a,b}} \|_2^2 \right) \tag{6.6}$$

Where $\overline{\Delta v_{a,b}}$ and $\overline{\Delta \omega_{a,b}}$ are empirical mean of $\Delta v$ and $\Delta\omega$. Both $\Delta v$ and $\Delta\omega$ are normalized here to exclude the effect of scale. This cost function quantifies the degree of deviation from the local average behaviour within a segment, enabling the algorithm to pinpoint boundaries where a statistically meaningful change occurs.

### 6.4.3 Cluster analysis

Cluster analysis is utilized in this chapter to distinguish different movement patterns. Since the trajectory data $T_t$ naturally varies in length, common distance metrics (such as L2 norm) cannot be used, as these evaluate pointwise similarity without considering the order (Holder, et al., 2024).

Hence Dynamic Time Warping (DTW) is used in this chapter as it allows the alignment of sequences of different lengths, and is widely used in time series (such as the trajectory data $T_t$) clustering (Holder, et al., 2024). The DTW distance between two trajectory data $T_t$ and $S_s$ $\coloneqq \left[ [\Delta v_s, \Delta\omega_s], [\Delta v_{s+0.5}, \Delta\omega_{s+0.5}], \dots, [\Delta v_{s+m}, \Delta\omega_{s+m}] \right]$ is calculated as follows:

$$DTW(T_t, S_s) = \sqrt{C(2n, 2m)} \tag{6.7}$$

Where $C$ is the cumulative cost matrix:

$$C(i,j) = D(i,j) + min\begin{cases} C(i-1,j) \\ C(i,j-1) \\ C(i-1,j-1) \end{cases} \tag{6.8}$$

Where $1 \leq i \leq n$ and $1 \leq j \leq m$ represent the indices of timestep, and $D(i,j)$ is the Euclidean distance:

$$D(i,j) = \sqrt{\left(\Delta v_i - \Delta v_j\right)^2 + \left(\Delta \omega_i - \Delta \omega_j\right)^2} \tag{6.9}$$

$\Delta v$ and $\Delta \omega$ are normalized here as well to exclude the effect of scale.

Based on the DTW distance, trajectory data can be categorized through agglomerative hierarchical clustering to identify movement patterns, which, as introduced in section 3.2.2.2, assumes each trajectory as a single cluster at first and progressively clustered together based on the similarity (Patel, et al., 2015). Such a method is deemed to be more effective with DTW in comparison to other conventional clustering methods (such as K-means) (AlMahamid & Grolinger, 2022; Niennattrakul & Ratanamahatana, 2007).

## 6.5 Results and discussion

### 6.5.1 Experiment

Data from 70 participants was successfully collected, with 23 participants in the first experiment, 22 participants in the second experiment and 25 participants in the third experiment. All data from the main sessions of both the first experiment and the third experiment were successfully recorded. However, due to a technical issue, data from the last four minutes of the second experiment were not captured. The details of the technical issue are further elaborated in chapter 8.

After data preprocessing, all the position data during teleportation and breaks were removed. It is noteworthy that in each experiment, at least one participant chose to exit without

completion, citing a sense of discomfort due to crowding at the entrance. Likewise, some participants chose to walk away from the entrance, as they believed it reflected their behaviour in real life. All of them agreed to rejoin the crowd upon discussion with the experiment organisers, and the position data from their leaving paths were removed as well.

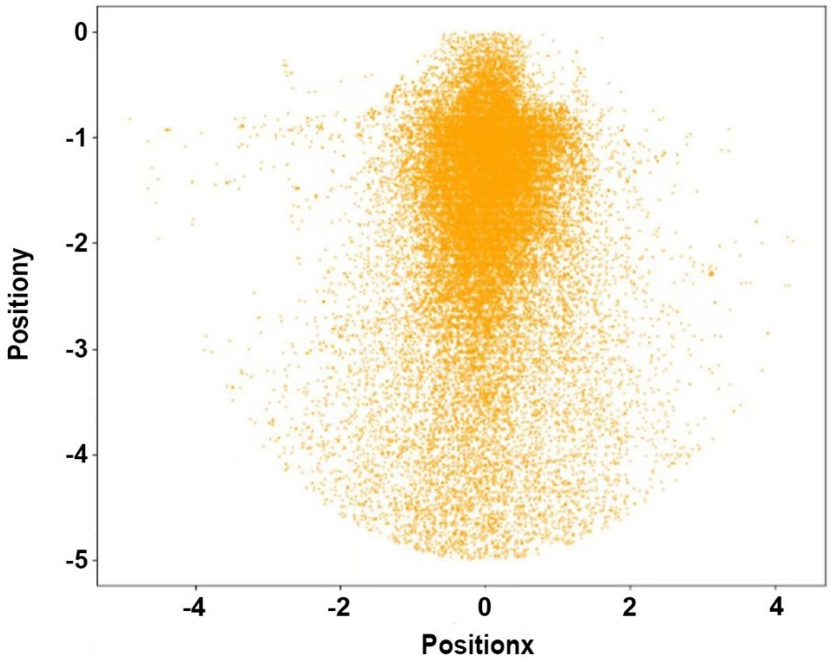



*Figure 6.3 Position data from the experiment*

The remaining position data are presented in Figure 6.3. The midpoint of the entrance gate is located at the origin, and an arch-formation similar to Figure 2.5 can be observed around it. As the emergence of arch formation is successfully achieved, the reproduction of scenario is considered successful. After extracting the behavioural measures, a total of 1032 trajectory data $T_t$ are achieved.

Kruskal-Wallis H test (Kruskal & Wallis., 1952) is used to examine the consistency of behaviour across the different experiments, as the trajectory data in this research does not follow normal distribution and more than three independent groups are involved. The results are presented in Table 6.1, indicating that no statistically significant differences were detected across experiments ($p > 0.05$). To quantify the magnitude of between-experiment differences, pairwise median differences and their 95% bootstrap confidence intervals (Efron & Tibshirani, 1994) were also estimated. For both $\Delta v$ and $\Delta \omega$ the pairwise median differences were zero with degenerate confidence intervals (95% CI [0, 0]), indicating that pairwise sample

medians were identical across all experiment comparisons. Together, these results indicate consistent behaviour across the three experiments. Additionally, this demonstrates that the unforeseeable technical issue in experiment 2 did not lead to statistically significant deviations in participant behaviour.

Similarly, the Mann-Whitney U test (Mann & Whitney, 1947) is used to examine consistency in behaviour between the first five minutes of the experiment and the last five minutes (one minute for experiment 2 due to missing data). Both behaviour measures have p-values greater than 0.05 (see Table 6.1), indicating that no statistically significant temporal differences were detected, i.e., no evidence of learning effect or fatigue. Median differences between time periods were zero with degenerate confidence intervals (95% CI [0, 0]), indicating identical sample medians across the experiment duration.

| Behaviour measure | P-value (across experiments) | P-value (first vs. last half) |
|---|---|---|
| Change of speed $\Delta v$ | 0.735 | 0.264 |
| Change of direction $\Delta\omega$ | 0.241 | 0.689 |

*Table 6.1 Experiment consistency*

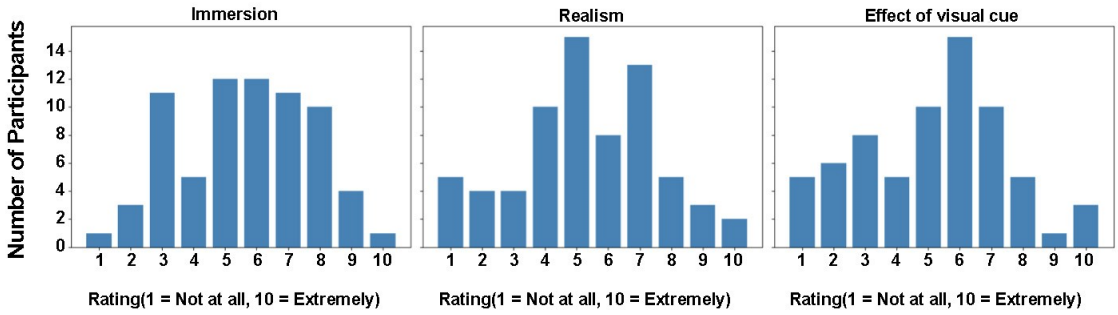



*Figure 6.4 Responses from questionnaire*

Additionally, the results from collected questionnaires are presented in Figure 6.4. The figure demonstrates that most participants gave medium to high ratings to all three questions, with 71.43%, 66.67% and 64.71% of participants giving scores greater than or equal to 5 for "Immersion", 'Realism" and "Effect of visual cue" respectively.

These results are comparable with a previous study by Arias, et al. (2021), where more than 70% participants gave medium-high level of ratings to realism, and 67.74% participants rated the environment as immersive or intermediately immersive. This provides further evidence for the successful reproduction of the jostle scenario. These questionnaire responses are further elaborated in chapter 8, particularly with respect to the utilisation of visual cues.

In summary, the reproduction of the scenario in all three experiments could be considered as valid, both in terms of phenomenon emergence, subjective experience and questionnaire feedback. Additionally, the consistency of participant behaviour across different experiments demonstrates the stability of data collected from designed VR experiment. Finally, the consistency over time suggests the data collection efficiency enabled by the flexibility of the virtual environment does not compromise data quality.

### 6.5.2 Data Analysis

#### *6.5.2.1 Change Point detection*

The results of the analysis of change point detection are presented in Figure 6.5. The y-axis is the normalised magnitude of $\Delta v$ (up) and $\Delta \omega$ (bottom), and the x-axis represents the index of the position where these behaviours occur, sorted in ascending order based on distance.

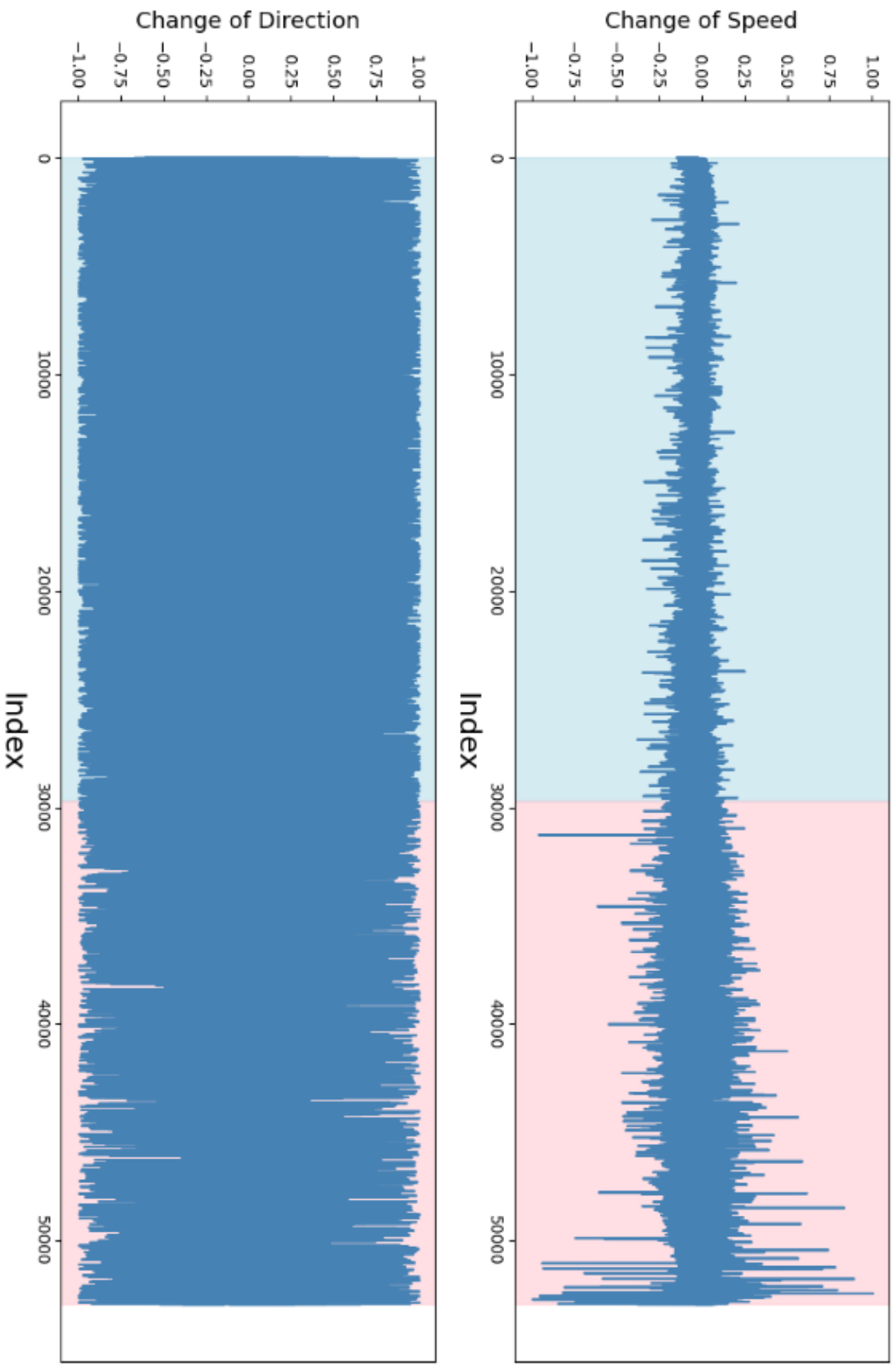



*Figure 6.5 Change point detection*

Two areas separated by the change point are marked with light blue and pink respectively. The index of the separating line corresponds to 2.58 meters from the gate, which is then identified as crowd radius by definition in this experiment. Such a result finds further support by observing the arch in Figure 6.3, where the distance from the origin to the edge of congested area is approximately 2.58 meters in the vertical direction. The length in the horizontal direction is slightly shorter and is believed due to the effect of the initial positions.

Comparing behaviour within and outside crowd radius gave the following results of a p-value of:

i) $3.82 \times 10^{-5}$ for $\Delta v$, indicating a statistically significant difference, and

ii) 0.25 for $\Delta\omega$, suggesting no significant differences.

Such a result is commonly observed as not all covariates undergo changes simultaneously in multivariate data (Si, et al., 2024), and the consistency of $\Delta\omega$ provides valuable insights in the next section's discussion.

For $\Delta v$ , an average absolute value of 0.27 is recorded within the crowd radius and 0.72 outside it, indicating a confinement on the speed change due to the presence of surrounding crowds. The absolute value of $\Delta v$ is used to eliminate the offsetting effects of acceleration and deceleration, ensuring the focus remains on the magnitude of changes.

### *6.5.2.2 Cluster Analysis*

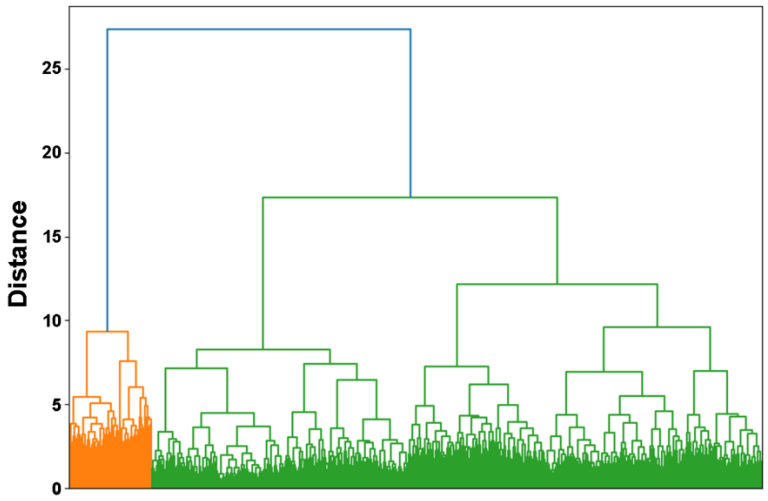



*Figure 6.6 Dendrogram of hierarchical clustering with DTW*

The resulting dendrogram of hierarchical clustering is presented in Figure 6.6, which visualises the entire clustering process. In the diagram, each node represents a cluster and the stem (height) is the distance at which the former two clusters are joined (Everitt, et al., 2011). Intuitively, the optimal number of clusters can be decided by inspecting the longest stem, as this indicates the largest differences (Yim & Ramdeen, 2015), which Figure 6.6 indicates is two. Additionally, a silhouette analysis, which assesses the quality of the clustering by comparing the level of separation among clusters based on the DTW distance (Rousseeuw, 1987), confirms the two clusters.

The two identified clusters are marked with orange and green in Figure 6.6, respectively. Cluster 1 consists of 705 trajectories, while there are 327 trajectories in cluster 2. On further analysis of the proportion of trajectories belonging to each cluster, no overwhelming dominance of one cluster was observed under any unique ID. Hence, such partition reflects different movement strategies rather than personal preferences.

| | Average Speed Change $\mu_{\Delta v}$ | Variance of Speed Change $\sigma^2_{\Delta v}$ | Average Direction Change $\mu_{\Delta\omega}$ | Variance of Direction Change $\sigma^2_{\Delta\omega}$ | Average Walking Distance $\mu_D$ |
|---|---|---|---|---|---|
| Cluster 1 | 0.31 | 0.24 | 0.86 | 1.70 | 3.01 |
| Cluster 2 | 0.20 | 0.13 | 1.08 | 2.40 | 4.47 |

*Table 6.2 Behavioural measures of different clusters (in absolute value)*

Table 6.2 presents the results of a further analysis of each cluster in terms of behavioural measures, with all the values calculated as absolute values as discussed above.

In terms of $\Delta v$, Cluster 1 exhibits greater average speed changes $\mu_{\Delta v}$, indicating sudden acceleration and deceleration, most likely due to interactions with surrounding crowds. This variability in speed is further reflected in its higher variance $\sigma^2_{\Delta v}$, suggesting that the strategy of trajectories in this cluster is to frequently adjust their pace in response to congestion.

On the other hand, Cluster 2 demonstrates a more stable speed pattern, implying less impact from surrounding crowds. The underlying reason for this difference is further explored through $\Delta\omega$. Cluster 2 exhibits both a higher magnitude (measured in radians) and greater frequency of direction changes compared to Cluster 1. This suggests that the strategy of trajectories in Cluster 2 prioritizes maintaining speed, actively adjusting direction more frequently to navigate through available spaces, rather than simply waiting behind others. Consequently, we name the behaviour captured in Cluster 2 as 'Impatient'.

The behaviour captured in Cluster 1 is therefore named as 'Patient', as it exhibits a higher tolerance for speed fluctuations and less frequent direction change. This can be further validated by examining the average walking distance $\mu_D$ of trajectories in Cluster 1, which is 3.01 meters, in comparison to 4.47 meters in Cluster 2. This is only slightly higher than the crowd radius 2.58 meters, which suggests an almost straight path is taken when entering the crowd. Consequently, the behaviour reflected in Cluster 1 is considered similar to daily queuing behaviour, where overtaking is rarely observed, and movement occurs only when the individual in front moves.

| Behaviour measure | P-value (Patient) | P-value (Impatient) |
|---|---|---|
| Change of speed $\Delta v$ | 0.001 | 0.276 |
| Change of direction $\Delta\omega$ | 0.740 | 0.152 |

*Table 6.3 Behavioural consistency for both type*

Additionally, the consistency of behaviour for both types of behaviour within and outside the crowd radius is examined, with results presented in Table 6.3. The significant difference in speed change after entering the crowd radius further emphasises the impact of surrounding crowds on patient behaviour, while the consistency observed in impatient behaviour validates the effectiveness of the frequent direction change strategy.

Similar to the results outlined in section 6.5.1, consistency is observed in the change of direction within and outside the crowd radius for both types of behaviour. This consistency is expected for patient behaviour, as it generally follows a straight path. However, for the impatient behaviour, this result provides additional insight, suggesting that the frequent changes in direction are not only a reaction to entering crowds, but also a strategy employed throughout.

These findings suggest that arch formation under the jostle scenario may be exacerbated by impatient behaviour. Individuals exhibiting such behaviour not only increase the local density by filling available spaces within the crowd, but may also seek positions closer to the gate when outside the crowd. This could effectively lead to more people trying to pass through the gate simultaneously, though, further studies are needed to verify this assumption.

### 6.5.3 Implication

As introduced in section 2.2.1.1, DIM-ICE model assumes that there are two primary modes of crowd behaviour: normal and emergency, and risks should be identified for both modes separately. However, this dichotomy is not suitable for the jostle scenario. Although jostle scenario should be classified as a normal condition due to the absence of perceived threat

(Kuligowski, 2013), the impatient behaviours observed here cannot be characterised as normal behaviour.

Although impatient behaviour represents only 31.69% of trajectories in this experiment, its presence, even as a minority, may be sufficient to generate a distinct type of crowd disaster. In comparison, patient (daily queuing) behaviour alone can induce crowd turbulence and high-density accumulation, which may instead lead to other crowd disaster type (Helbing & Mukerji, 2012; Yang, et al., 2024). This underscores the distinction between impatient behaviour and normal condition behaviour.

In summary, the two types of behaviour identified in this chapter have provided additional insight that heterogeneous behaviours can arise within a single nominal condition. This highlights the importance of understanding crowd behaviour for more effective crowd management strategy.

## 6.6 Summary

This chapter has explored crowd movement patterns under the jostle scenario through three rounds of VR experiments. At least 22 participants are included in each round, with a total of 70 participants' locomotion trajectory data has been recorded. This is the largest scale experiment in the crowd analysis field to the best of our knowledge.

The consistency of participant behaviour across the three experiments, as well as across different time periods, supports the stability of the data collected in the designed VR experiment. The emergence of arch-formation phenomenon has also been observed, indicating a successful reproduction of the scenario, together with feedback from collected questionnaires.

The collected locomotion trajectories have been further analysed to identify the crowd radius where the participants experience an abrupt change of behaviour. Within the crowd radius, two types of behaviour, *patient and impatient*, have been identified through agglomerative hierarchical clustering.

Both types of behaviour have provided valuable insight into understanding jostle scenario. In the next chapter, some of the conclusions drawn in this chapter are cross validated, such as their preferences of different movement strategies. Furthermore, different underlying factors influencing these behaviours are quantified and discussed for a deeper understanding.

# 7. Crowd decision making inference framework

## 7.1 Introduction

On the basis of results from chapter 6, this chapter proposes a decision making inference framework to understand the two distinct behaviours within the jostle scenario, *impatient* and *patient*. Traditional approaches to understanding pedestrian decision making, such as questionnaire and regression-based methods, often rely on subjective judgement or are only able to focus on simple actions. This in turn renders them insufficient in order to fully capture the complexities of real-world behaviour.

A novel framework is proposed that integrates inverse reinforcement learning (IRL) with model interpretability techniques to infer the underlying decision-making processes of pedestrians. By retrieving the reward function encoded in a neural network and interpreting it using model interpretability, the influence of each input feature on behavioural preferences can be quantified. This framework is applied to both impatient and patient behaviour using the datasets collected in the VR experiment.

The inferred results not only reveal key differences and similarities in decision-making between the two behavioural types but also validate findings for their behavioural strategies in chapter 6. Furthermore, the use of derived feature importance to calibrate parameters within an SFM is demonstrated, thereby forming a calibrated SFM. This approach enables interpretable, low-cost simulations that maintain competitive predictive performance while reducing the computational overhead.

Section 7.2 reviews relevant literature including common methods used to infer pedestrian decision making and inverse reinforcement learning. Section 7.3 introduces methodologies employed in this chapter, including the model formation, training algorithm and methods used for model interpretability. Section 7.4 presents the result along with further discussions and section 7.5 summarises the chapter.

The framework proposed in this chapter was presented and discussed at the 26th International Workshop on Multi-Agent-Based Simulation (MABS), held as part of the 2024 International Conference on Autonomous Agents and Multiagent Systems (AAMAS), and was published as:

Yang, X., Yang, L., Majumdar, A., Ochieng, W. (2025). Inferring Pedestrian Decision-Making Through Inverse Reinforcement Learning. In: Thompson, J., Stankov, I. (eds) Multi-Agent-Based Simulation XXV. MABS 2024. Lecture Notes in Computer Science, vol 15583. Springer, Cham. https://doi.org/10.1007/978-3-031-88017-9_7

## 7.2 Literature review

### 7.2.1 Pedestrian decision making

In crowd behavioural research, theory-driven experiments are often used to validate the hypothesised behavioural mechanisms. For instance, Moussaïd, et al. (2009) conducted controlled experiments to validate pedestrian interaction laws and estimate their strength, thereby providing experimental support for the SFM. In the context of decision making inference, one theory being widely tested is bounded rationality (Simon, 1956), which assumes that people employ satisficing strategies, selecting the first option that meets an acceptable level of need.

In this study, another proposed mechanism of pedestrian decision making is utilised, identified by Tong & Bode (2022), which involves a process of making trade-offs between desirable and undesirable attributes in the surrounding environment. Such a process is known as utility theory, which assumes that for pedestrian $n$, with $k$ attributes under consideration, a quantitative value $\beta_k$ (also known as utility) is assigned to each attribute $X_n$. In contrast to bounded rationality, the utilities reflect the degree to which the corresponding attributes contribute to the actions made and are combined in a way commonly known as the utility function $U_n$, formulated as:

$$U_n = \beta_1 X_{n_1} + \beta_2 X_{n_2} + \cdots + \beta_k X_{n_k} \tag{7.1}$$

Under utility theory, each action made by a pedestrian could be viewed as an outcome of optimising the utility function, as it stands for the mechanism of decision making. Prospect theory has been proposed as an alternative for utility theory, in which outcomes are evaluated relative to a reference and perceived as gains or losses rather than absolute utilities (Kahneman & Tversky, 1979). However, this perspective is not adopted in this study, due to the difficulty of setting up the reference.

Two approaches have been widely adopted in order to infer the relative utility (as this thesis is only interested in their comparative contribution) for different attributes encountered during movement: scores given by participants using a questionnaire (Yang & Sun, 2013; Bivina & Parida, 2020; Papadimitriou, et al., 2017) and by means of a regression model based on experimental data (Haghani & Sarvi, 2017; Lovreglio, et al., 2015; Soares, et al., 2021).

However, both approaches suffer from drawbacks: pedestrians do not always perform in the way they declare in the questionnaire (Papadimitriou, et al., 2016) due to respondent bias (Krumpal, 2011), and the regression model can only be used to decide the contribution of each factor towards action choice at a single timestep, often represented as discrete choices (e.g., choice of exit) (Wedagamaa, et al., 2020).

In summary, a better approach than at present for inferring the utility function of the value associated with pedestrian decision making is needed, that is capable of both:

- quantifying contributions of each attribute towards the function decision reflecting pedestrian's movement preferences in real life;

- comprehending the movement preference behind the entire trajectory.

### 7.2.2 Inverse reinforcement learning (IRL)

As introduced in section 3.3.2.4, the reward function $R$ in RL encodes the agent's objectives and behavioural preferences, providing a reward signal during each transition along the trajectory. Accordingly, reward function $R$ can be used to infer an agent's intentions (Fu, et al.,

2018) and the cumulative sum of rewards along the trajectory $\tau$ could be viewed as a utility function (Allievi, et al., 2023):

$$U(\tau) = \sum_{t=1}^{\tau} R_t \tag{7.2}$$

By optimising U, the optimal policy $\pi^*$ is achieved, which stands for the preference of actions over the entire trajectory.

Inverse reinforcement learning (Ng & Russell, 2000), on the other hand, has been proposed to infer the reward function $R$, and is widely regarded as a form of imitation learning. As the name suggests, IRL is considered as a form of reverse engineering of RL, aiming to recover the underlying objectives that best explain an agent's policy $\pi(a \mid s)$, which, as introduced in section 3.3.2.4, represents the probability of choosing action $a$ when in state $s$.

In practice, policy $\pi(a \mid s)$ is typically not directly observable, but rather in the form of expert demonstration, as illustrated in section 3.3.2.5. Given such demonstrations, IRL is able to estimate relative weights of the multi-attribute reward function and construct policy π that is capable of performing specific tasks akin to real-life experts (Ng & Russell, 2000).

In the field of pedestrian dynamics, IRL has been successful in:

- inferring a navigation policy for robots in human crowds (Fahad, et al., 2018; Gonon & Billard, 2023; Chalvatzaki, et al., 2019).

- modelling interactions between pedestrians with vehicles (Nasernejad, et al., 2021; Nasernejad, et al., 2023) and cyclists (Alsaleh & Sayed, 2020).

- predicting future trajectories (Saleh, et al., 2018).

This has exploited the capacity of IRL for constructing policy π, but left the reward function $R$ unexplored.

The Adversarial IRL method (Ho & Ermon, 2016; Fu, et al., 2018) proposed by Chelsea Finn (2016) provides a framework to infer the reward function based on generative adversarial networks. Similar to GAN, introduced in section 3.3.2.3, such a method involves the training of a generator neural network $G$, together with a discriminator neural network $D$. The reward function can then be retrieved from the discriminator $D$ in a form of neural network (Fu, et al., 2018). However, analysing the reward function still poses challenges, as the neural network decision making mechanism is not explicitly accessible (Sheu, 2020).

In the following sections, a new framework is proposed to infer a pedestrian's decision making process by combining IRL learning with model interpretability, which is designed to explain neural network behaviour (He, et al., 2024). This framework is then used to infer the underlying decision making of both types of crowd behaviour identified in the previous chapter. The resulting analysis highlights the relative contribution of each attribute to the decision-making process for each behavioural type.

## 7.3 Methodology

In this section, the methodologies employed in this chapter are reviewed and discussed. It is important to note that since this chapter discusses both IRL and crowd movement, the term trajectory is used in both fields. While they share certain conceptual similarities, they do not refer to the same entity. To distinguish between them, the term used within in the IRL framework is explicitly denoted as trajectory $\tau$.

### 7.3.1 Model framework

#### *7.3.1.1 Model formation*

Following the perspective adopted by both RL and imitation learning, the entire trajectory $\tau$ of pedestrians could be viewed as a Markov Decision Process (MDP) (Sutton & Barto, 2018) denoted by the tuple $M = (S, A, T, R, \gamma)$, where:

i. $S$ is the state space,
ii. $A$ is the action space,
iii. $T(s, s', a)$ is the transition probability from state $s$ to state $s'$ under action $a$ determined by the environment dynamics,
iv. $R$ is the reward function trying to recover,
v. $\gamma$ is the discount factor ranged on $[0,1)$, indicating the level of preference for immediate reward over future reward.

The environment was constructed on the basis of the standard Open AI gym framework (Brockman, et al., 2016) to replicate the experiment settings, with the entrance gate located at the origin and supermarket wall extending along the x-axis. The agent is able to move freely in the third and fourth quadrants. The length of each timestep $\Delta t$ is set at 0.5 seconds.

To reflect the perspective of socio-technical system, the state space $S$ of each pedestrian $i$ is decomposed into three components:

i) Self state

Self state variables are intended to capture the intrinsic properties of a pedestrian. In RL applications for crowd modelling (Vizzari & Cecconello, 2023; Li, et al., 2025), common self state variables include $[x_i, y_i]$ to denote the position of pedestrian $i$, and $[v_i, \omega_i]$, to represent the speed and direction respectively.

The self state of pedestrian $i$ is then denoted as:

$$[x_i, y_i, v_i, \omega_i] \tag{7.3}$$

ii) Surrounding state

The surrounding state aims to capture the state of the surrounding crowds. Since these variables should reflect the information perceived by pedestrian $i$, it is natural to import the concept of eyesight and only focus on the information available in this region. Following Zhang, et al.

(2022), the region of an eyesight is modelled as a sector oriented along the direction of movement $\omega_i$, with a radius of 4 meters and a central angle of 180 degrees.

As summarised by Kim, et al. (2020), key factors of the surrounding crowds that affect pedestrian $i$'s decision making include: their movement state, speed and direction, and total number. However, in the framework of Adversarial IRL, state space is often used as the input of a neural network, which generally requires a fixed size. Consequently, it is impractical to track the movement state of every pedestrian within the sight, as their number varies over time.

To address this limitation, only the movement state $[v_j, \omega_j]$ of the closest pedestrian $j$ within the eyesight is included in the state space. This choice is based on the assumption adopted in the SFM, which states that the magnitude of the interaction between pedestrians is inversely proportional to their distance. Thus, the movement state of the closest pedestrian is considered to exert the greatest influence on the agent's decision-making.

The total number of pedestrians $N$ within the eyesight of an individual is also incorporated into the surrounding state. Since this is the number of pedestrians within a specific region, it can be viewed as a proxy for local density, which exerts a strong impact on pedestrian trajectory (Kim, et al., 2020). Additionally, since physical contact is reproduced in the experiment as demonstrated in section 6.3.1, it is included in the surrounding state as well. To capture perceptual asymmetries, physical contact is divided into two types, the physical contact $c_f$ that is clearly observable, i.e., from the front direction, and the physical contact $c_o$ that is partially observable or entirely unobservable, i.e., from the remaining three directions. Both contact variable $c_f$ and $c_o$ are binary, and a value of 0 indicates no contact and 1 indicates that someone is in contact with pedestrian $i$.

Accordingly, the surrounding state of the pedestrian $i$ is denoted as:

$$[v_j, \omega_j, N, c_f, c_o] \tag{7.4}$$

ii) Environment state

Environment state is designed to capture the interaction between the pedestrian and the external environment.

From the perspective of the physical environment, the environment is relatively simple, with no specific restrictions imposed regarding the degrees of freedom in the direction of movement. The only physical infrastructure present in this environment is the supermarket wall, which extends along the x-axis. The perpendicular distance between the wall and the pedestrian $i$ is thus already captured by the self state variable $y$.

From the social scenario perspective, all the crowds share a common goal: to pass through the entrance gate as quickly as possible. In the RL application to crowd modelling, such a goal is commonly represented by the distance $d$ with the target destination, which is expected to decrease progressively over time. The environment state is then denoted as [ $d$ ].

In summary, the state space $S$ in this chapter is defined as:

$$S = \left[x_i, y_i, v_i, \omega_i, v_j, \omega_j, N, c_f, c_o, d\right] \tag{7.5}$$

Consistent with Chapter 6, the action space $A$ is defined by two behavioural measures:

$$A = [\Delta v, \Delta\omega] \tag{7.6}$$

where $\Delta v$ stands for change of velocity, and $\Delta\omega$ is the change of direction.

At each timestep $t$, the environment dynamics that determine the state transition update the agent's states based on the policy $\pi(a \mid s)$, selecting the corresponding action $a_t \in A$ based on the current state $s_t \in S$.

The speed and direction states are updated as follows:

$$v_{i,t+1} = v_{i,t} + \Delta v_t \tag{7.7}$$

$$\omega_{i,t+1} = \omega_{i,t} + \Delta\omega_t \tag{7.8}$$

With $\omega_{i,t+1}$ normalized to lie within the range $[-\pi, \pi]$ and position states updated subsequently as:

$$x_{\,i,t+1} = x_{i,t} + v_{i,t+1} \cdot \cos\left(\omega_{i,t+1}\right) \cdot \Delta t \tag{7.9}$$

$$y_{\,i,t+1} = y_{i,t} + v_{i,t+1} \cdot \sin\left(\omega_{i,t+1}\right) \cdot \Delta t \tag{7.10}$$

All the remaining states are consequently updated based on agent's new position.

In this chapter, a value of $\gamma = 0.95$ is adopted based on empirical comparison with $\gamma = 0.90$ and $\gamma = 0.99$, representing a balance between short-term reactions and long-term planning. This is due to the assumption that pedestrians are not only reacting to their immediate surroundings, but are also goal-driven and capable of making temporally extended decisions. Such decisions may involve pursuing a task level objective, such as passing through the entrance gate as quickly as possible.

#### *7.3.1.2 Training Algorithm*

*Generator*

Adversarial Inverse Reinforcement Learning (AIRL) proposed by Fu, et al. (2018) is utilised in this chapter due to its capability to recover the reward robust to changes in environment dynamics. Within AIRL's framework**,** the generator $G$ aims to match the occupancy measure, defined as:

$$\rho_\pi = \pi(a|s) \sum_{t=0}^{\infty} \gamma^t \, P(s_t = s|\pi) \tag{7.11}$$

This expression represents the distribution of the state action pair under policy $\pi(a|s)$, i.e., the probability of encountering each state action pair if following policy $\pi(a|s)$. The objective is

for this distribution to closely match that of expert policy $\pi_E$, derived from expert demonstrations.

These expert demonstrations, as introduced in section 3.3.2.5, are often denoted as a collection of trajectories $\mathcal{D} = \{\tau_i\}_{i=1}^{M}$, where $\tau_i = \{(s_t, a_t)\}_{t=0}^{T_i}$ consists of a sequence of state action pairs, $M$ is the total number of trajectories $\tau$, and $t$ indexes each individual timestep within each trajectory $\tau$ (Osa, et al., 2018).

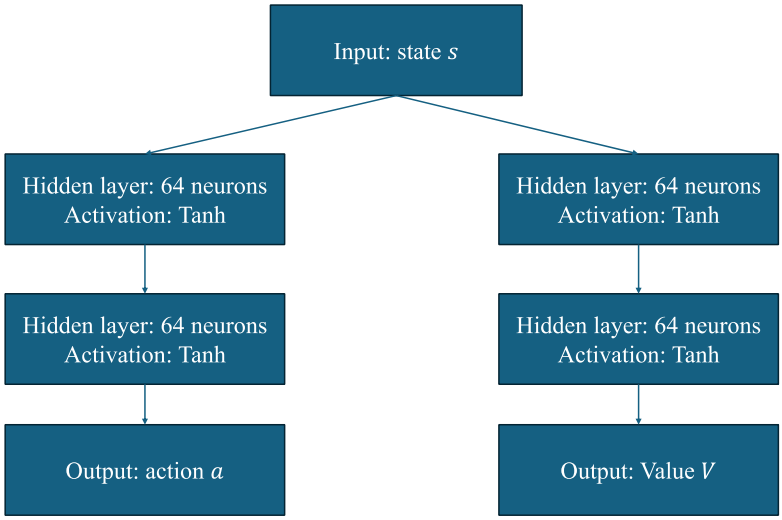



*Figure 7.1 Structure of Generator network*

The structure of the generator $G$ implemented in this chapter is illustrated in Figure 7.1. It consists of two separate subnetworks: one generates action $a$, which corresponds to $\pi(a|s)$ and the other predicts state value $V$, which is used to calculate cumulative reward over time, as introduced in section 3.3.2.4. Each subnetwork comprises two hidden layers, each with 64 neurons, following a common baseline architecture (Srouji, et al., 2018). The stated number of neurons is provided primarily for reproducibility, while varying this number results in less than 2.86 % change in prediction accuracy.

The hyperbolic tangent function tanh is utilised as the activation function for each hidden layer, and it is defined as (Dubey, et al., 2022):

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{7.12}$$

To evaluate the performance of the trained policy $\pi(a|s)$, two metrics that are commonly employed in the field of crowd modelling is imported: average displacement error (ADE) and final displacement error (FDE) (Pellegrini, et al., 2009; Alahi, et al., 2016; Rudenko, et al., 2022; Salzmann, et al., 2021). Both metrics are evaluated over a 4.5 second prediction horizon, similar to the settings adopted in related research (Kretzschmar, et al., 2016; Gonon & Billard, 2023).

ADE measures the mean Euclidean distance between each predicted points and the ground truth positions within the trajectory (Alahi, et al., 2016), calculated as:

$$\text{ADE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\widehat{y_{i,t}} - y_{i,t}|_2 \tag{7.13}$$

Where $N$ is the total number of trajectories, and $T$ is the total number of timesteps. $\widehat{y_{i,t}}$ is the predicted position of the $i$-th trajectory at time $t$, and $y_{i,t}$ is the ground truth position. $|\cdot|_2$ is used to denote the Euclidean distance.

FDE, on the other hand, focuses on the Euclidean distance between the final predicted destination and the true destination (Alahi, et al., 2016), computed as:

$$\text{FDE} = \frac{1}{N}\sum_{i=1}^{N}|\widehat{y_{i,T}} - y_{i,T}|_2 \tag{7.14}$$

Where $\widehat{y_{i,T}}$ is the predicted final destination, and $y_{i,T}$ is the ground truth destination.

In this chapter, methods introduced in chapter 3 are selected as benchmarks, including Social-LSTM, BC and Trajectron ++. These methods are commonly used as benchmarks for crowd modelling models (Gupta, et al., 2018; Manh & Alaghband, 2019; Rudenko, et al., 2022), with BC in particular commonly used as a baseline for evaluating imitation learning algorithms (Ho & Ermon, 2016; Jia, et al., 2024).

The trained policy $\pi(a|s)$ is then compared with benchmarks under ADE and FDE. It is important to note that such comparison is not intended to establish the superiority of the AIRL

algorithm, but rather to demonstrate that the performance of trained policy $\pi(a|s)$ is comparable to that of state-of-the-art models. Unlike RL, which provides a definitive criteria for evaluating training success, IRL lacks such an approach (Wu, et al., 2023). Consequently, the evaluation of the trained policy is often considered as a good alternative (Abbeel & Ng, 2004). Meanwhile, a valid performance of the trained policy $\pi(a|s)$ can also reinforce the validity and relevance of the retrieved reward function $R$.

*Discriminator*

The discriminator $D$, on the other hand, aims to differentiate the generated trajectories $\tau$ with expert demonstration by minimizing cross-entropy loss (Goodfellow, et al., 2016):

$$-E_{\tau\sim gen}\left[\log\big(1-D(\tau)\big)\right]-E_{\tau\sim expert}\left[\log D(\tau)\right] \tag{7.15}$$

Compared with other adversarial IRL methods, such as GAIL (Ho & Ermon, 2016), AIRL gives a special structure to the discriminator $D$:

$$D_{\theta,\phi}(s,a,s')=\frac{\exp\{f_{\theta,\phi}(s,a,s')\}}{\exp\{f_{\theta,\phi}(s,a,s')\}+\ \pi(a|s)} \tag{7.16}$$

Where $f_{\theta,\phi}$ is a neural network parametrized by $\theta$ and $\phi$, and is restricted to the learnt reward network $R_\theta$, which is the reward function trying to retrieve, and a shaping term $h_\phi$:

$$f_{\theta,\phi}(s,a,s')=R_\theta(s,a)+\gamma h_\phi(s')-h_\phi(s) \tag{7.17}$$

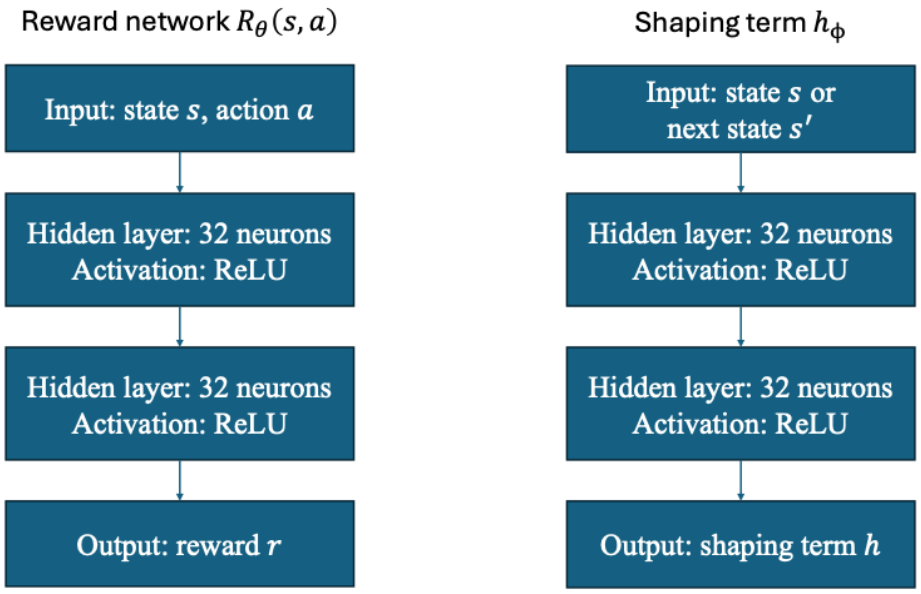



*Figure 7.2 Structure of reward network and shaping term*

The structure of both networks is illustrated in Figure 7.2. Each network consists of two hidden layers with 32 neurons (Srouji, et al., 2018), using ReLU as the activation function, which is defined as (Dubey, et al., 2022):

$$\mathrm{ReLU}(x) = \max(0, x) \tag{7.18}$$

The python package 'imitation" (Gleave, et al., 2022) was used to implement the training algorithm of AIRL and BC. Given the original code for Social-LSTM was unavailable, it was reproduced based on the descriptions provided in the original paper (Alahi, et al., 2016).

Notably, both Trajectron ++ and SoPhie are designed to process datasets that include all pedestrian trajectories within a scene, which makes isolating and training on specific behavioural subsets inherently difficult. Consequently, they were not selected as a benchmark in this evaluation.

### 7.3.2 Model interpretability

Model interpretability methods, as introduced above, aim to understand the internal decision-making of neural networks, providing insight into the attribution of each input feature (He, et al., 2024).

One popular approach to explain the neural network is based on gradients, as they are able to quantify the impact of small changes in input features on the model outputs (Wang, et al., 2024). Integrated Gradient, introduced by Sundararajan, et al. (2017) is a classical gradient-based attribution method. It computes feature attributions by integrating the model's gradients along a straight path from a baseline input $x'$ to the actual input $x$. The method is defined as:

$$\text{Integrated Gradients} = (\mathrm{x} - \mathrm{x}') \cdot \int_0^1 \nabla f\big(\mathrm{x}' + \alpha(\mathrm{x} - \mathrm{x}')\big)\, \mathrm{d}\alpha \tag{7.19}$$

Where $\alpha \in [0,1]$ is the interpolation parameter for the straight path between $x'$ and $x$, while $\nabla f(\cdot)$ denotes the gradient of output function $f$.

Another widely used method to interpret neural networks is named as SHAP (SHapley Additive exPlanations) value, proposed by Lundberg & Lee (2017). This is based on the concept of the Shapley value, originally introduced in cooperative game theory (Shapley, 1953), on ML models. In this framework, each input feature is viewed as a "player" in a game, collaborating with others to produce the model's output. This method ensures a consistent and locally accurate distribution of contribution, by considering all possible combinations of input features.

Gradient SHAP, which integrates ideas from Integrated Gradients and the SHAP framework, is particularly suitable for extracting the SHAP values of neural network with a known structure (Cremades, et al., 2024), and is therefore adopted in this chapter to explain retrieved reward network $R_\theta$. It is defined as:

$$\text{Gradient SHAP} = (x - x') \cdot E_{\alpha, x'}\big[\nabla f\big(x' + \alpha(x - x')\big)\big] \tag{7.20}$$

Where the interpolation parameter $\alpha \sim U(\mathbb{0}, \mathbb{1})$ is uniformly sampled from 0 to 1 to sample points along the path between input $x$ and baseline $x'$. Baseline $x' \sim \mathbb{B}$ is sampled from a baseline distribution $\mathbb{B}$, which provides a set of reference inputs serving as neutral starting points for attribution. The operator $E[\cdot]$ denotes the expectation over these samples.

The Python package "Captum" (Kokhlikyan, et al., 2020) was used to implement the Gradient SHAP algorithms, while the Python package 'SHAP' (Lundberg & Lee, 2017) was used to illustrate the outcome.

## 7.4 Results and discussion

### 7.4.1 AIRL Training

The data collected from the experiment in chapter 6 was divided into two categories, impatient and patient, as the identified from the cluster analysis. Since patient behaviour exhibits significant changes within and outside the crowd radius, and including trajectories outside the crowd radius resulted in reductions of 40.54 % in ADE and 32.79 % in FDE accuracy, only the data within the crowd radius was retained. In contrast, all the data for impatient behaviour was included due to the consistent movement strategy.

Both types of behavioural data were then pre-processed into a set of trajectories $\mathcal{D}$, resulting in 30978 state action pairs for the impatient behaviour, and 7065 state action pairs for the patient behaviour. Following the convention in the fields of ML (Joseph, 2022), each dataset was divided into an 80% training set and a 20% testing set.

The AIRL algorithm was applied to both behavioural types in order to retrieve the desired reward network $R$. Training was conducted on a workstation, equipped with an Intel Core i9-9900K CPU (16 threads, 3.60 GHz), 64 GB of RAM and an NVIDIA T1000 GPU. A total of 1,000,000 training steps were conducted for both behavioural types.

The training results of the $\pi(a|s)$ are presented in Table 7.1 based on ADE and FDE in the unit of metres, where value closer to 0 indicating a better prediction performance. The

comparison with selected benchmarks is also included, with BC trained on the same model formation. For Social-LSTM, six input steps are used, corresponding to three seconds, closely aligning with its original settings of 3.2 seconds (Alahi, et al., 2016). Both ADE and FDE are computed over a prediction horizon of 4.5 seconds, equivalent to nine timesteps.

| Methods | Behavioural Type | ADE | FDE |
|---|---|---|---|
| AIRL | Impatient | 0.78 | 1.19 |
| | Patient | 0.35 | 0.57 |
| Social-LSTM | Impatient | 0.77 | 1.21 |
| | Patient | 0.41 | 0.64 |
| BC | Impatient | 1.32 | 2.19 |
| | Patient | 0.60 | 1.01 |

*Table 7.1 Policy evaluation and benchmarks comparison*

It can be concluded that the policy $\pi(a|s)$ trained through AIRL using the proposed model formation achieved performances that are comparable to or exceeding that of Social-LSTM and BC. In general, all methods achieved a statistically significant better performance (at $\alpha = 0.05$), both in terms of ADE and FDE, for predicting patient behaviour than impatient behaviour, reflecting the greater unpredictability of the latter and highlighting again the importance of understanding their decision making.

As discussed above, the training of AIRL based on the proposed model formation is successful, given the results presented in Table 7.1. This also validates the subsequent investigation of the retrieved reward network $R$, discussed in the next section.

### 7.4.2 Decision making inference

Gradient SHAP was subsequently used to interpret the reward network $R$ for both types of behaviour. All the trajectories $\tau$ in the testing set are used as the inputs, while the baselines are sampled from the training set, similar to previous studies (Yap, et al., 2021; Liu, et al., 2022).

The results for interpretating impatient behaviour and patient behaviour are presented in Figure 7.3 and Figure 7.4 respectively. Two types of graphs are used in this chapter to illustrate the SHAP value. The graph on the left is a beeswarm plot, which shows a clear mapping between each input data point with its corresponding SHAP value. In this graph, each dot represents a single data point, with its colour indicating the feature value: red denotes a relatively high value, blue indicates a relatively low value, and purple represents an intermediate value.

The graph on the right is a bar plot, which displays the mean absolute SHAP value for each feature, providing a summary of overall feature importance. This contrasts with the beeswarm plot, which visualises the raw SHAP values for individual data points to capture both impact and variability. In both plots, features listed in the y-axis are in descending order based on their mean absolute SHAP value. However, as the SHAP values reflect the relative attributions of features in a specific model, it cannot be used to compare across different models, i.e., a SHAP value of 3.5 for a certain feature in Figure 7.3 does not necessarily indicate greater influence than a SHAP value of 3 in Figure 7.4.

The SHAP value for all the features within the state space $S$ and action space $A$ are listed. For features within $S = [x, y, v_i, \omega_i, v_j, \omega_j, N, c_f, c_o, d]$each of them are denoted as: self position $x$, self position $y$, self speed (the distance moved over the last time interval), self direction, front speed, front direction, density, front contact and other contact respectively. Additionally, features within $A = [\Delta v, \Delta\omega]$ are denoted as speed change and direction change.

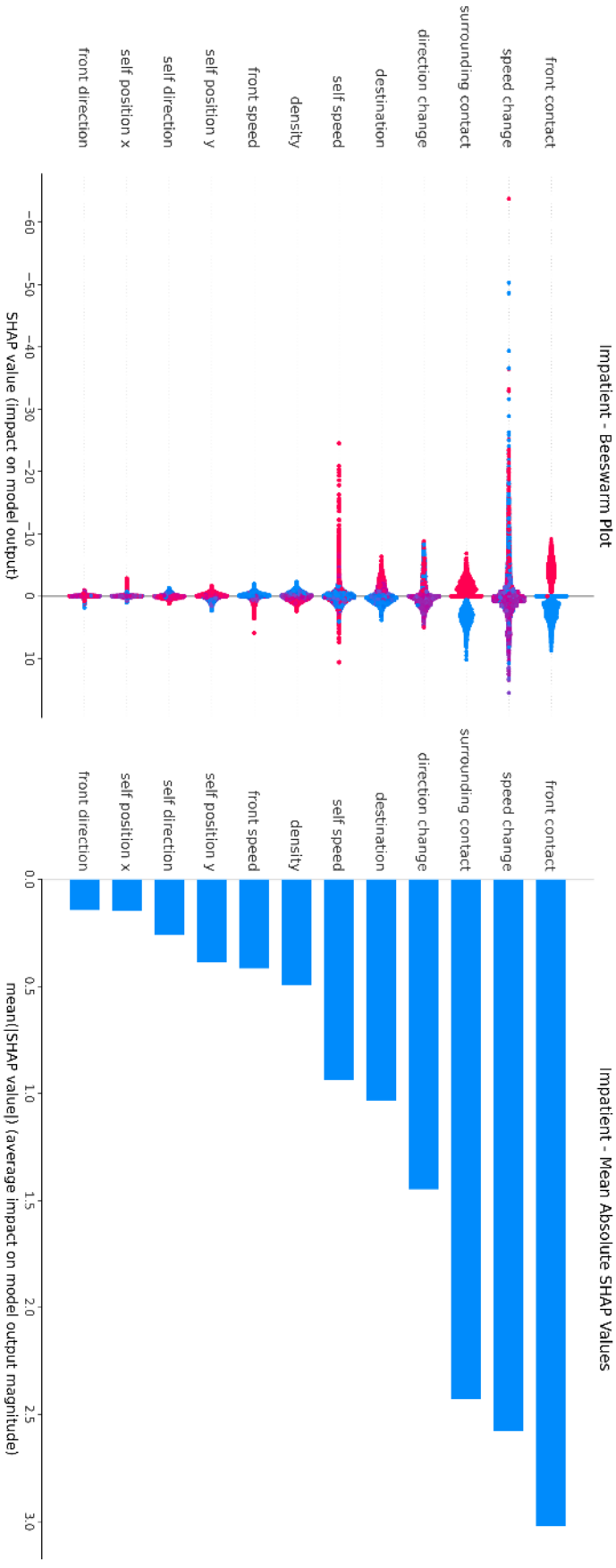



*Figure 7.3 Impatient behaviour decision making*

Based on Figure 7.3, the results validate the finding from chapter 6 that speed change has a considerable influence on the decision making of impatient crowd as the second influential feature with a mean absolute value of 2.58. The beeswarm plot shows that both high or low speed change, corresponding to a sharp acceleration or deceleration, are associated with negative SHAP values, while most intermediate values have a positive SHAP value. The pattern indicates that the impatient crowd tends to avoid such abrupt changes, which is consistent with the results in chapter 6.

Additionally, self speed is more important than self direction to the impatient crowd, with a mean absolute value of 0.94, ranking sixth, compared with 0.26 for self direction, which ranks tenth. Interestingly, some instances of high self speed are associated with a negative impact on the predicted reward. This is assumed to be the case because such high speeds are often observed when individuals are still far from the entrance gate, which is a less undesirable state.

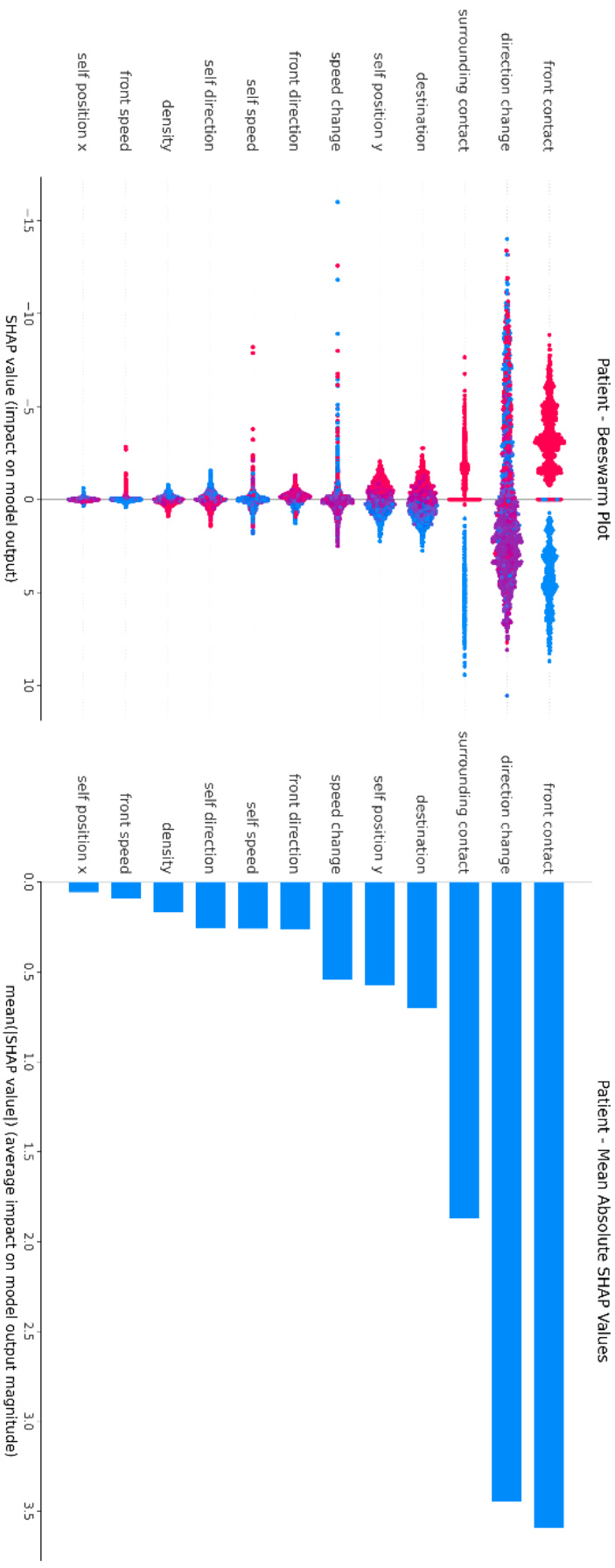



*Figure 7.4 Patient behaviour decision making*

On the other hand, the results from Figure 7.4 are also consistent with the behavioural pattern identified for patient crowd in chapter 6, i.e. they tend to maintain a straight path and avoid abrupt direction changes. This is reflected through the beeswarm that both high or low direction changes, corresponding to a sharp turn either counterclockwise or clockwise, are mostly associated with negative SHAP value, as indicated by their position on the positive x-axis.

Furthermore, the patient crowd receives a much smaller impact from the speed change than the impatient crowd, with speed change only ranking as the sixth influential feature with a mean absolute value of 0.54. This suggests that the patient crowd is more tolerant of frequent speed fluctuations while tending to avoid abrupt directional shifts, reinforcing the previously observed preference for queuing behaviour.

In terms of impact from surrounding crowds, both impatient and patient behaviour demonstrate a tendency to avoid physical contact with the surrounding crowds, especially with the people within their eyesight as front contact is ranked the first influential features for both, with mean absolute values of 3.02 and 3.60 respectively. This tendency may be partly attributed to the collider design in the VR experiment: since pushing between avatars was not permitted, pedestrians ahead were effectively perceived as obstacles.

For movement state of surrounding crowds, patient crowds focus more on the front direction, aligning with their queuing behaviour, while impatient crowds value more their front speed, favouring higher front speed. These preferences reflect their distinct behavioural characteristics.

In terms of impact from the social external environment, destination shows a moderate level of influence, ranking fifth for impatient and fourth for patient respectively. The results demonstrated in the beeswarm plots aligns with intuitive expectations that the shorter distance with the entrance gate is desirable. This consistency could be seen as a further validation of our decision making inference framework. Additionally, both types of behaviours see higher density as a sign of progress, as shown in the beeswarm plots, due to higher density indicate closer to the entrance gate.

In summary, both types of behaviour prioritize avoiding physical contact with surrounding crowds, while rank goals of scenario in the middle. These similarities underscore the existence of causal mechanisms that operate independently of human perception, as discussed in section

4.2. More importantly, features that contribute to the distinct differences between these two behaviours are mainly movement states, reflecting different preferences in movement strategies. This further emphasises the importance of understanding individual goals.

### 7.4.3 Implications

In the sections above, the proposed framework, which combines AIRL and model interpretability, have been introduced and discussed to infer the decision making behind two types of crowd behaviours. The results can be used as a valuable source for a better crowd management strategy, further discussed in the next chapter. On the other hand, this framework holds broader potential beyond crowd management and could contribute to other pedestrian related fields. This is also be demonstrated through a case study in urban planning presented in the next chapter.

In this section, we demonstrate another implication of the proposed framework, utilising the quantified importance of each feature on decision making as a foundation for crowd modelling models that capture intentions, with SFM serving as an illustrative example.

As discussed in section 3.3.2.1, calibrating strength factors remains a major challenge when utilizing SFM, especially since these factors cannot be measured directly (Kretz, et al., 2018). To address this, this thesis proposes to use the SHAP values retrieved from our framework as the strength factor for each individual factor in SFM.

Since SFM only captures impacts from physical entities, only front contact $c_f$, other contact $c_o$, destination $d$ and density $N$ are incorporated. The resulting SHAP-calibrated SFM is expressed as:

$$\vec{F} = SHAP_d\overrightarrow{f_d}(d) + \alpha \cdot \left\{(SHAP\}_{c_f}\overrightarrow{f_{c_f}}(c_f) + SHAP_{c_o}\overrightarrow{f_{c_o}}(c_o)\right) + \beta \cdot SHAP_N\overrightarrow{f_N}(N) \quad (7.21)$$

Where $\vec{F}$ is the total social force and $\vec{f}$ denoting each individual force. Based on the beeswarm plots, $\overrightarrow{f_d}$ and $\overrightarrow{f_N}$ are attractive forces, while $\overrightarrow{f_{c_f}}$ and $\overrightarrow{f_{c_o}}$ are repulsive forces. The SHAP term represents the mean absolute value of corresponding SHAP values in the bar plots. The

parameter $\alpha$ and $\beta$ are scaling factors that ensure all the forces operate at a comparable magnitude.

| Behavioural type | Distance range | ADE | FDE |
|---|---|---|---|
| Impatient | 9 | 1.56 | 2.78 |
| | 2.58 | 0.76 | 1.29 |
| Patient | 2.58 | 0.63 | 1.30 |

*Table 7.2 Simulation results of SFM*

The SHAP-calibrated SFM was evaluated on the same datasets by randomly sampling 20% of all available initial points drawn from trajectories within each behavioural type. From these starting conditions, the model was simulated forward over a 4.5-second prediction horizon, with its performance assessed using ADE and FDE. The results are presented in Table 7.2.

For impatient behaviour, two conditions were tested:

i) one using initial points within the range of 9 meters, consistent with settings in Table 7.1, and

ii) another constrained within the range of 2.58 meters, which is the crowd radius.

Although the results using the 9-meter range yielded a slightly lower prediction accuracy, performance within the crowd radius was comparable to other advanced models. This underperformance at longer ranges may be attributed to the frequent absence of $c_f$ and $c_o$ outside the crowd radius, which leads to a lack of repulsive forces. For patient behaviour, the model has also achieved comparable performance, further validating the utilisation of SHAP values as a strength factor.

Furthermore, SHAP-calibrated SFM can be directly implemented in a similar scenario, with much less computational cost and no requirement for data inputs. Given that SFM still underpins many modern methods, e.g., Zhang, et al. (2022) use SFM as the ground-truth physics model, the SHAP-calibrated SFM can similarly support the development of crowd

modelling frameworks. In summary, outcomes of the proposed decision-making inference framework can be used as a starting point for crowd modelling models.

## 7.5 Summary

This chapter introduced a novel framework for inferring pedestrian decision-making, combining AIRL with Gradient SHAP. This approach was applied to understand the decision making underlying the two types of crowd behaviours, impatient and patient, within the jostle scenario, identified through cluster analysis in chapter 6.

The proposed framework uses AIRL to retrieve a reward neural network, which is then interpreted using Gradient SHAP to quantify the contribution of each input feature. This enables a nuanced understanding of the factors that influence pedestrian movement decisions, offering deeper insights into the decision-making patterns that underlie each type of behaviour as discussed in section 7.4.2.

Trained policy $\pi(a|s)$ was evaluated using metrics ADE and FDE, which demonstrated competitive performance against the established benchmarks introduced in chapter 3, including Social-LSTM and BC. The interpretability analysis revealed consistent movement preferences for each crowd type as identified in chapter 6, highlighting avoidance of abrupt speed changes for impatient crowds and patient crowds' queuing behaviour, while also confirming shared tendencies, such as avoiding physical contacts.

Importantly, the chapter also explored the practical implications of the retrieved feature importance by incorporating them as strength factors into a SHAP-calibrated SFM. The resulting model demonstrated comparable predictive accuracy with lower computational cost, even under varying initial conditions.

The previous three chapters have identified nine types of crowd disasters and investigated in considerable depth with a VR-based experiment and ML methods to identify and understand distinct behaviour patterns. This investigation provides a deeper understanding of crowd behaviour, incorporating the influence of the social external environment. Moreover, the findings of the past three chapters offer broader implications for both crowd management strategies and the study of pedestrian dynamics, which is further explored in the next chapter.

# 8. Implications

## 8.1 Introduction

Having presented three studies in chapter 5, chapter 6 and chapter 7, this chapter discusses the implications of this research with three main aims in mind. The first is to broaden the impact of this work beyond theoretical contributions and examine how research findings may inform real-world crowd management practices by engaging with key stakeholders. This is discussed in Section 8.2. The second, given the innovative use of VR in crowd analysis, both the challenges encountered and the advantages gained throughout the process from design to implementation are discussed in section 8.3. Finally, the scalability of the proposed decision making inference framework is explored by presenting a case study involving urban planning in section 8.4.

## 8.2 Implication for crowd management

In order to understand the real-world applicability of the proposed contributions to crowd management practices, an interview was conducted with a senior stakeholder from the Night Time Industry Association, a trade association representing the interest of nightclubs, live music venues and other nightlife sectors. Given the high exposure of these industries to crowd management challenges, the interviewee 1 brought substantial expertise in developing and implementing crowd management plans.

After presenting the main findings of this research, including the nine types of crowd disaster and the investigation of two behavioural types within the jostle scenario, interviewee 1 acknowledged the relevance of our research findings and described them as “insightful”. Meanwhile, interviewee 1 has also confirmed the wide adoption of two operational guides in the UK: the “Safer Crowd, Safer Venues’’ (Chebib, et al., 2025) and the “Purple Guide” (Events Industry Forum, 2022). The Green Guide (Sports Grounds Safety Authority, 2018) was also considered, however, as it primarily addresses safety at sports grounds rather than the types of events examined in this study, it was not carried forward in the subsequent analysis. Both

selected guides were subsequently reviewed and following three potential contributions of the research findings of this thesis were identified, as illustrated in Figure 8.1:

i. The FIST model is one of the most utilized risk analysis models for pre-event assessment. In the research presented in chapter 5, an extension of FIST model was introduced which incorporated the social scenario and physical environment in order to better capture the impact from the external environment. This extended FIST model could therefore lead to more a comprehensive risk analysis outcome, if implemented in place of the original model.

ii. As both documents repeatedly emphasise the importance of understanding crowd behaviour, especially with the U.K. Cabinet Office document "Understanding crowd behaviour" (Challenger, et al., 2010) remaining highly relevant over a decade after publication. The patterns of impatient and patient behaviour identified in both chapters 6 and chapter 7 can be incorporated in the crowd management plans.

iii. Both documents also highlight the critical role of crowd analysis, noting that understanding the likely routes taken by people aids not only in calculating venue capacity but also in developing more robust incident prevention strategies. Consequently, the SHAP-calibrated SFM from chapter 7 can be used as an alternative that is easier to implement and less computationally intensive.

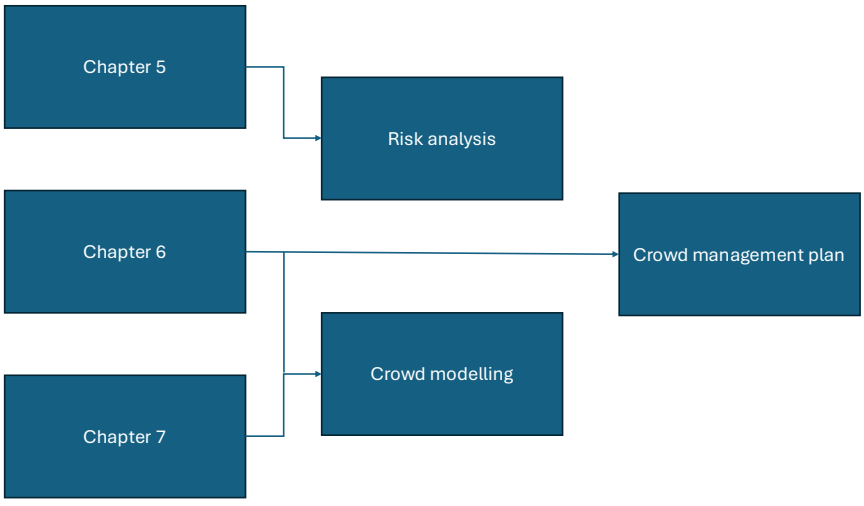



*Figure 8.1 Potential contributions of research findings*

In order to understand the real-world applicability of the proposed contributions to crowd management practices, an interview with a senior stakeholder from the UKCMA was conducted, which represents the interest of commercial organisations focusing on providing crowd management and event security services. The main research findings were also presented to interviewee 2 as well, along with the proposed potential contributions. No further interviews were conducted due to the scarcity of specialists in this domain (as elaborated below), as well as the difficulty of accessing them.

In the following sections, the responses from the interviewee 2 for these contributions are presented with further implications discussed.

### 8.2.1 Extended FIST model

The interviewee 2 did not see potential for applying the extended FIST model directly in current crowd management practices, noting that the DIM-ICE model, elaborated in section 2.2.1.1, is more widely known and commonly used in the industry. However, the interviewee 2 acknowledged that the jostle scenario, identified through the analysis of past crowd disasters using the extended FIST model, has been recognised by the industry, while typically not having a crowd management plan in place. In particular, events involving less than 5000 people typically do not have a formal crowd management plan. This situation has started to change following the Brixton crowd disaster in 2022, in which two people died when fans without tickets tried to force entry into a sold-out concert (Warren, 2024).

Although the framework used to analyse crowd disasters may not be directly applicable to current crowd management practices, the behavioural pattern identified has been recognised by operational personnel and highlights a significant gap. As the industry begins to consider developing crowd management plans for such types of events, the research presented in this thesis can be further incorporated as a valuable source of data and a reference point in the absence of precedent. This also emphasises the importance of continued research in this field.

### 8.2.2 Patient and impatient behaviours

On presenting the findings on the two types of behaviours identified within the jostle scenario, the interviewee 2 expressed a strong interest in understanding their individual details. When discussing the potential application of these findings within crowd management practices, the interviewee 2 explained that crowd management is typically undocumented as a standalone plan, but is instead integrated with other documents into a broader event plan. Therefore, the understanding of these two behavioural types could serve as a valuable data source for describing audience profiles in a more narrative and heuristic manner, for example by outlining "five tangible things that could happen."

Consequently, in the following section presents 'Tangible crowd behaviours during the jostle scenario' based on the findings in chapter 6 and chapter 7.

#### *8.2.2.1 Tangible crowd behaviours during jostle scenario*

The jostle scenario typically refers to events where most people attempt to obtain limited items before others in the crowd. Such events often include sales promotions, charity distributions, or gatherings that operate on a "first come, first served" basis without issuing tickets. This list is not exhaustive, and any event fitting the description of a jostle scenario may use this section as a point of reference.

The locations most likely to experience potential crowd disasters are typically at the entrance of a venue or at any bottleneck where crowd flow becomes restricted during ingress. Therefore, additional resources should be allocated to these areas. However, other types of crowd disaster may also occur, for example during counterflow associated with egress or when perceived threats emerge.

The main cause of crowd disasters in this scenario is the presence of people attempting to move at a high speed. Although they may not always achieve their desired speed due to various constraints, this primarily reflects a mental rather than a physical state. To better understand the presence of a high desired speed, one key indicator is the "arch formation" phenomenon, where an arch-shaped congestion pattern develops in front of a bottleneck.

Two types of behaviour are commonly observed during entry phase of the jostle scenario: patient and impatient behaviour. Although the former is generally more prevalent within crowds, impatient behaviour is the primary cause of the arch formation phenomenon and therefore warrants particular attention. These two behavioural types reflect different strategies adopted by individuals to manoeuvre within the crowd. Notably, both types of behaviour would typically avoid physical contact with others, especially with those in front of them. Therefore, any widespread physical contact observed, provides a concerning sign that the crowd may be starting to lose control due to its high density.

For the patient behaviour, individuals act much like that seen during a daily queuing situation, where they do not easily overtake others or change their directions. Consequently, they tend to follow directions taken by people in front of them in a straight path, creating a ripple-like pattern in the crowd's movement, with people moving forward and then pausing in succession as the flow naturally slows and picks up again.

For impatient behaviour, individuals' movements are generally harder to predict because they are more sensitive to short-term incentives, especially a slowdown in the flow speed. People displaying impatient behaviour dislike changes in their speed and prefer to maintain a steady pace. Consequently, they often change direction, especially when those in front of them are moving more slowly, seek to keep moving and occupy any available space.

Furthermore, impatient behaviour tends to have higher level of tolerance for higher crowd density than patient behaviour, making them more likely to cause extreme local density, which may result in injuries or death. It is therefore crucial to intervene at an early stage prior to conditions becoming severe. A warning signal is a slowdown in the crowd flow rate, as this can heighten impatience and cause individuals to push into already dense areas of the crowd. When such signs are observed, it is important to increase monitoring and take appropriate precautionary measures before the density builds up. Possible measures include deploying staff to guide movement, or introducing physical barriers to limit aggressive overtaking. The use of barriers should be carefully evaluated, as they may also introduce risks of collision related injury.

### 8.2.3 Application of the SHAP-calibrated SFM

The interviewee 2 described the concept behind the SHAP-calibrated SFM maybe as “a game changer for the industry, given it is easier to implement and less computationally intensive than current methods. The interviewee 2 also noted that modelling is generally absent for small crowd events due to both cost concerns and the lack of legal requirements. This has become more important since small crowd events have begun to develop crowd management plans in recent years, where modelling can be crucial.

Furthermore, another key objective of the proposed SHAP-calibrated SFM, i.e. to reduce the need for subjective expert input, aligns well with the nature of the operational industry, which the interviewee 2 characterised as having “a lack of specialists.” This shortage has led to a reliance on commercially developed software, which is often too costly for small crowd events.

However, it should be acknowledged that the proposed SHAP-calibrated SFM still requires both further testing to evaluate its generalizability to other scenarios and refinement prior to being directly integrated into crowd management practices. Moreover, it addresses only a subset of crowd management considerations and is not applicable to all planning requirements, e.g., it cannot address high-level route choice. Finally, while the model itself does not require new data inputs for implementation, its initial development depends on accurate and extensive datasets, which may be challenging to obtain in real-world crowd management contexts.

In summary, the discussion with the interviewee 2 not only verified the relevance of pursuing a less computationally intensive approach but also highlighted the importance of minimizing the need for specialist expertise in crowd management solutions. Furthermore, as reviewed in section 3.3, recent research in crowd analysis has increasingly focused on neural network-based models due to their higher accuracy. However, these approaches are significantly more computationally demanding and complex in terms of runtime operations compared to conventional models such as the SFM (Kolivand, et al., 2021). Consequently, the proposed strategy of informing conventional models with insights derived from neural network-based methods appears to be better aligned with the current needs of the industry and should be explored further in future research. The proposed SHAP-calibrated SFM concept is also examined under different conditions in section 8.4.

## 8.3 Implication for usage of virtual reality in crowd analysis

The VR experiment serves as an essential element for this research, both as the data source and as the method of investigation. As discussed in chapter 6, VR remains an innovative tool in scientific research, while providing major advantages for exploring scenarios that are hard to reproduce in a physical experiment. While the results of chapter 6 are solid and insightful, it is equally important to reflect on the broader experience of conducting this VR-based investigation. This includes considerations from the design and development stages to implementation and data interpretation. Such reflection not only highlights the strengths of using VR in crowd analysis but also reveals practical challenges and considerations for future studies.

The author participated in a project named WeWalk (WeWalk, 2025) during the first year of PhD study. The project aims to develop a smart cane with navigation guidance for visually impaired person and involves a set of physical experiments. The experience gained from assisting these physical experiments is referenced here as a point of comparison to highlight the differences between physical and virtual experiments. As the WeWalk project specifically addressed visually impaired users, it should be noted that the present thesis focuses on typical pedestrian populations and does not explicitly model functional limitations such as visual impairment.

### 8.3.1 Environment development

In contrast to field physical experiments, which can make use of a pre-existing space as the experimental environment, the virtual environment has to be developed entirely from scratch typically. This development process was the most time-consuming phase of this investigation, taking over three months to complete before all the desired functionalities were fully implemented and tested.

By comparison, the preparation time of WeWalk experiment, as a field experiment, only took a morning. A large classroom has been used as the experimental space, with paths marked with tapes on the floor and data collection sensors calibrated on-site. This demonstrates that the

preparation for physical experiment can be much shorter than VR experiment in some scenarios. In the VR experiment, many aspects of "real-life" need to be digitally replicated, e.g., rules that determine an individual's movement and perception, aspects that are inherently present and readily available in real-world physical environments. However, in lab physical experiment, different layouts may still need to be deployed to artificially reconstruct specific scenarios.

The software used for developing the virtual environment is Unity, upon which VIRSE is built. Unity is also commonly used in the field of Architecture, Media and Education. However, these fields often focus on the creation of interactive experiences, while the experiment in this thesis, implementing features, e.g. a scoreboard and teleportation mechanism, are more closely aligned to game development. The skillset required here can be seen as a combination of experiential design and coding, an interdisciplinary expertise that is relatively scarce among researchers in the field of crowd analysis, despite a growing body of crowd dynamics research using VR (Haghani, 2020). Consequently, the virtual environment was developed in collaboration with researchers from Royal College of Art, who develop VR to create interactive contents that fit with their specific designs.

Furthermore, a major challenge in conducting multi-participant VR experiments is to ensure that each participant can perceive the real-time actions of others, which in turn requires a reliable network synchronization so that actions performed by one individual are accurately reflected to all other participants. Fortunately, this challenge was mitigated in this thesis through the use of the VIRSE platform (Imperial College London, 2025), which has the built-in online interaction functionality. However, it should be noted that such a platform may not be accessible to all researchers in the field of crowd analysis. Developing a comparable system independently would be both time-consuming and resource-intensive, underlining the significant technical barriers to conducting multi-participant VR experiments.

More importantly, the developed virtual environment is difficult to reuse in future studies, as its configuration, including the selected scenarios and implemented features, is highly tailored to fit with the specific objectives of the current experiment. This limitation further increases the cost and effort associated with preparing VR-based experiments.

Despite these challenges, VR-based experiments offer unique advantages that are difficult to achieve through traditional physical setups. For instance, obtaining ethical approval is often

considerably easier when exploring dangerous scenarios, as the virtual environment avoids physical risk to participants. In the ethical approval granted by Science, Engineering and Technology Research Ethics Committee of Imperial College London, this VR-based experiment was classified as low risk.

### 8.3.2 Experiment design

As discussed above, aspects of the real-world need to be digitally replicated in the virtual environment so that the actions of participants can reflect their behaviours in real-life. The two most essential aspects replicated in our experiment were movement and the feel of physical contact.

As demonstrated in chapter 6, the movement mechanism of each avatar is enabled through the drag-move of the controller. This is a built-in functionality of VIRSE to replace the more commonly used thumb stick-based movement. According to the developers of VIRSE, this replacement was motivated by two primary considerations:

1) The common use of the thumb stick is more likely to cause motion sickness, due to the mismatch between the user's relatively stationary physical body and the visually perceived motion in the virtual environment.

2) The drag-move mechanism more closely mimics movement in real-life, as it allows the users to control their speed more subtly, based on the pace and direction of arm movement.

These two considerations were deemed to be effective during the experiment process, with an additional advantage emerging, which are discussed in the next section.

Similar experiments that use VR to explore crowd movement may adopt alternative approaches, allowing the participants to physically walk within a confined space while wearing the headsets, so that the information they received is from the established virtual environment. This approach automatically solves the realism problem, but significant ethical concerns arise especially at a multi-participant level.

Physical experiment requires careful design to prevent participants from colliding with physical objects, especially immovable structures such as the walls and pillars. In the WeWalk experiment, additional safety measures were taken to prevent any visually impaired person participating from colliding with the pillar. An experiment assistant was also at hand nearby to intervene if potential harm arose.

However, this challenge becomes even more complex in virtual environments. While virtual markers can be added to represent the positions of physical objects, doing so requires a precise 1:1 replication of the real-world space. It is undeniable that even minor discrepancies may exist between the actual locations of physical objects and their virtual counterparts. Consequently, human supervision remains necessary to ensure participant safety. Where these safety measures can be useful for scenarios that only involve slow movement, they are inadequate for scenarios where rushing or sudden high-speed movement is expected, especially for dangerous scenarios that are commonly explored through VR.

Introducing multiple participants makes it impractical to allow for real movement. Similar to the issue with physical objects, even if the positions of other participants are projected accurately in virtual environment, collisions are still likely to occur, particularly because detailed movements of limbs are difficult to capture. Furthermore, it is unrealistic to rely on human supervision given the number of participants involved. A potential solution would be to place each participant in a separate physical space while projecting their avatars into a shared virtual environment. However, this approach is highly resource-intensive and may be infeasible for most research settings.

Another concern of conducting multi-participant scenarios is related to how to reproduce their interactions, especially in the physical aspect. In our experiment, the sensation of physical contact is substituted with a visual cue to simulate the feeling of compression. Other alternatives, such as haptics from the controllers and playing specific music from the headset, were considered but ultimately dismissed due to a potential bad user experience. This concern is primarily due to the expected long-time duration that participants are in physical contact with others. While visual substitution enabled scalable multi-participant interaction, it cannot fully replicate the mechanical and tactile characteristics of real human contact, which remain a known limitation of current VR crowd simulations.

Feedback for the effectiveness of visual cues are also collected through the questionnaire with the results presented in Figure 6.4. As demonstrated in section 6.5.1, over 60% participants gave a medium-high level of ratings to the usage of the visual cue, which is comparable with ratings for realism and immersion. Notably, this question received the highest standard deviation among the three questions, indicating a wide range of opinions and demonstrating the need for further research in addressing the challenge of substituting any physical sensation. Prior research demonstrates that incorporating haptic feedback can enhance perceived crowding, emotional response, and behavioural realism in virtual crowd environments (Montecchiari, et al., 2017; Berton, et al., 2022; Venkatesan, et al., 2023). More broadly, perceived crowd realism in VR depends not only on visual density but also on embodied interaction constraints such as proximity and collision response (Pelechano & Allbeck, 2016). Future work should therefore investigate hybrid interaction approaches that combine visual, physical, and haptic cues to improve the fidelity of interpersonal contact in high-density VR crowd scenarios.

Furthermore, the participants are unable to exert actual forces on others, commonly seen in crowded scenarios, in the experiment outlined in this thesis. Such an omission is mainly due to ethical concerns, as participants may be conscious that they can push freely without causing any physical harm. Additionally, accurately replicating physical force interactions would require a delicately designed environment and significantly greater computational resources. This limitation highlights the heavy reliance of VR-based studies on infrastructure, which is further discussed in the next section.

Besides the ability to explore dangerous scenario, data are another strength of conducting a VR-based study. Since the virtual environment is fully controllable, all desired data can be collected directly without the process of inference. For example, locomotion trajectory is often a fundamental component in crowd analysis and can be captured as ground truth in the VR-based experiment presented in this thesis. In comparison, the WeWalk physical experiment required additional calibration for sensor-based data and faced challenges related to measurement accuracy.

Finally, video recording is available in the virtual environment, where replays from multiple angles can be generated without physical constraints. On the other hand, taking recordings

during physical experiment is often restricted by the layout of the experimental space, while also subject to stringent ethical scrutiny.

### 8.3.3 Experiment execution

All types of experiment have certain reliance on equipment, but any VR-based study has considerable higher reliance, making it more vulnerable to technical issues. As stated in section 6.5.1, a technical issue had caused data loss in the second experiment. Although the data quality remained unaffected, the incident has demonstrated the disruptive consequences of a technical failure for the research process, as nearly half of the usable data was lost.

Upon investigation, the issue was traced to a version update of the VIRSE platform. The first experiment was conducted in April 2024, while the second experiment conducted in July 2024. In the intervening period between these two experiments, a new functionality was introduced in VIRSE that is incompatible with the developed virtual environment. This new functionality increased the computational load on the virtual environment, ultimately resulting in data loss. This has again highlighted the challenges faced when developing a custom virtual environment. It is difficult for an individual researcher to build such virtual environment from scratch, while tools developed at the organisational level, such as VIRSE, often lack features like a long-term-support version, which is a common practice in a large-scale software development to ensure stability.

It is important to clarify that this is not intended as a critique of VIRSE. Given the scarcity of developed tools for creating customized multi-participant environment, VIRSE has provided substantial convenience, and its developer offered essential help throughout the experimental journey. Given the current resources available, maintaining a long-term support version is understandably difficult, so a practical approach might be to conduct the experiment as soon as the development is complete. However, this approach can further reduce the reusability of the developed virtual environment and is heavily dependent on the timely recruitment participants.

Participant recruitment though remains a challenge in conducting multi-participant experiments at scale. The size of each round of experiments is limited by the number of

available devices. Although the VIRSE lab is capable of accommodating at most 33 participants simultaneously, a precautionary measure adopted was to cap the participant size to 25, due to the possibility of unexpected technical issues that might make some devices unavailable. Even with this reduced number, considerable coordination was required to identify suitable timeslots, and unforeseen participant dropouts before the experiment session were observed in both the first and second experiments.

As stated in section 6.3.2, bare minimum inclusion criteria were applied. Based on experience in other VR-based studies, participants' familiarity with video games has demonstrated impacts on their proficiency in controlling avatars when the thumb stick movement control is adopted, due to its prevalence in video games. However, the drag-move implemented by VIRSE is rarely encountered in video games, placing all the participants at a comparable level of proficiency. Consequently, familiarity with video games was not included as an inclusion criterion for recruitment, and the potential influence of prior experience on participant behaviour was effectively minimised.

This section has reflected upon the experience conducting the VR experiment. Despite various challenges, valuable insights have been achieved with essential data collected. However, difficulties persist, largely due to the current stage of technological development and the still-unexplored dimensions of employing such innovative tools. The hope is that the reflection presented in this section can serve both as a practical guidance for those considering the use of VR, and a reference point for future research directions.

## 8.4 Implication for utilising proposed decision making inference framework

The decision making inference framework proposed in chapter 7 has served as the crucial tool in this research, enabling a deeper understanding of crowd behaviour. Notably, this framework demonstrates high scalability, making it applicable beyond the domain of crowd management.

In this section, a case study in collaboration with the School of Architecture, Southeast University, China is presented. The case study focused on exploiting the proposed framework in the context of urban planning, based upon the urban commercial street Xinjiekou in Nanjing.

### 8.4.1 Nanjing Xinjiekou case study

This case study quantifies the influence from elements of the built environment, such as locations of walls and green spaces, i.e., trees planted alongside the sidewalk, on the decision making of pedestrians. These factors can be used for future design plans that are more human centric, as well as locating the comfort zone of the pedestrians.

Furthermore, different time periods are taken into account, as time plays a vital role in shaping the relationship between the built environment and human behaviour (Saarloos, et al., 2009). This case study investigates three different time periods of the day: morning, noon and evening.

#### *8.4.1.1 Data*

The area of interest in this case study is Nanjing's commercial centre, Xinjiekou. As shown in Figure 8.2, the research area features eight-story commercial buildings on the east side and a sidewalk greening strip on the west. This sidewalk is 19 meters in width and 146 meters in length.

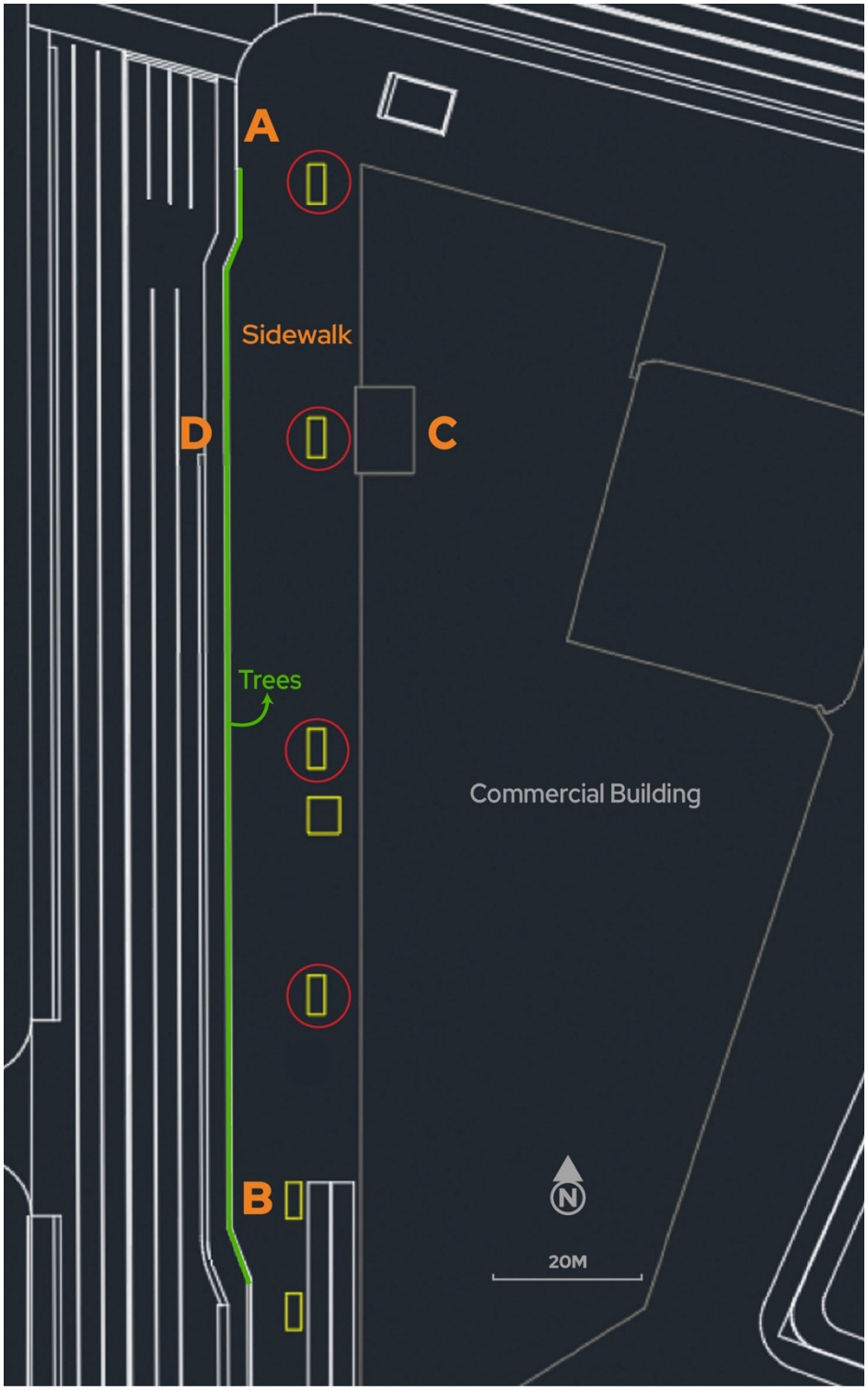



*Figure 8.2 Xinjiekou commercial street*

The data used for this case study were collected on two consecutive working days in May 2024, comprising of a 15-minute video recorded at the same time in the morning, at noon and in the evening, denoted as AM, NOON and PM. The footage was recorded using an aerial drone at an altitude of 7m, with a 30° downward tilt angle in 1080p resolution at 60 frames per second.

The pedestrian trajectories were subsequently derived from these videos using a pre-developed computer vision algorithm (Yang, et al., 2025). At the initial stage, a perspective transformation was performed using OpenCV (Bradski, 2000) by selecting four corresponding points between video frames and ground plane to compute a transformation matrix for coordinate conversion. Afterwards, the pedestrian identification was conducted employing the YOLOv5 (You Only Look Once) object detection framework (Ultralytics, 2020) with model weights trained on the open-source CrowdHuman dataset (Shao, et al., 2018). Then, the StrongSORT method (Du, et al., 2023) was applied to continuously track the detected pedestrians throughout the video sequence. Overall, the detection achieved over 95% accuracy.

Among the processed trajectories, two primary types of travel patterns were identified: vertical travel, primarily between points A and B as marked in Figure 8.2 and horizontal travel, mainly between points C and D, where point C denotes an entrance to the shopping centre and point D corresponds to a bus station. In the subsequent analysis, all trajectories between C and D were excluded, due to their limited exposure to the built environment along that street. Furthermore, trajectories travelling from B to A, which are in the northbound direction, were also excluded due to their scarcity, accounting for less than 10% of the processed data, and the additional model dimension required to capture their distinct behavioural patterns.

#### *8.4.1.2 Model formation*

Based on previous research (Ewing, et al., 2015), the features within the area of interest that directly influence pedestrian movement include the following: street wall, green space, transparency of street wall, and the presence of physical obstacles along the walking path. More specifically, the obstacles refer to ventilation shafts rising from the ground, indicated by the white rectangles marked with red circles in Figure 8.2.

These elements form the foundation of agent state space $S$, which includes: the shortest distance to both street wall $D_w$ and green space $D_g$, transparency of the street wall $T$ (denoted by a binary number where 1 indicates transparency), and horizontal distance with the closest sharp corners of the obstacles $D_o$. Scenario-specific information, such as slope and floor material is not included due to lack of comparison within the scene.

The choice of horizontal over the Euclidean distance is motivated by the nature of the vertical journeys observed in all the selected trajectories. Due to the street layout, the Euclidean distance from obstacles inevitably decreases over time, regardless of the pedestrian's intention or behaviour. However, studies (Easa, 2016; Nikmanesh, et al., 2025) have demonstrated that lateral clearance, the horizontal distance from obstacles, has a significant influence on pedestrians' collision avoidance behaviour, making it a more meaningful metric in this context.

In addition to the built environment features, the current position $[x, y]$ and destination $[x_d, y_d]$ were also included in the state space $S$ as navigation features. In summary, the state space $S$ is expressed as:

$$S = \left[D_w, y, x_d, y_d, D_g, D_o, T\right] \tag{8.1}$$

Here, the x-coordinate of current position $x$ is replaced by wall distance $D_w$, as the wall spans alongside the y-axis after data preprocessing, making these two values equivalent.

In contrast to chapter 7, the action space $A$ is defined as:

$$A = [\, v, \omega] \tag{8.2}$$

Where $v$ denotes the speed and $\omega$ the direction. This simplification is adopted because changes in speed and direction are not the primary focus of this case study.

Aside from this modification, the environment dynamics, training algorithm, and model interpretability framework remain consistent with those outlined in section 7.3. BC was still utilized as the benchmarks, with Social-LSTM replaced by SoPhie and Trajectron ++ given their better ability in extracting features from the physical environment. The implementation of Trajectron ++ [3] and SoPhie [4] were accessed via their official GitHub repository.

[3] https://github.com/StanfordASL/Trajectron-plus-plus
[4] https://github.com/coolsunxu/sophie

#### *8.4.1.3 Results and discussion*

| Methods | Time period | ADE | FDE |
|---|---|---|---|
| AIRL | AM | 0.84 | 1.59 |
| | NOON | 0.80 | 1.52 |
| | PM | 0.94 | 1.68 |
| BC | AM | 1.14 | 2.08 |
| | NOON | 0.83 | 1.47 |
| | PM | 0.86 | 1.60 |
| SoPhie | AM | 1.67 | 2.44 |
| | NOON | 0.96 | 1.30 |
| | PM | 1.05 | 1.34 |
| Trajectron ++ | AM | 0.89 | 1.50 |
| | NOON | 0.94 | 1.58 |
| | PM | 1.41 | 2.40 |

*Table 8.1 Policy evaluation and benchmarks comparison*

The training of AIRL was conducted on the same workstation as described in chapter 7, and a total of 800000, 1000000 and 1300000 training steps were conducted for AM, NOON and PM respectively. The training results, along with comparisons to selected benchmarks, are presented in Table 8.1. Both ADE and FDE are evaluated over a horizon of 4.8 seconds by convention (Alahi, et al., 2016), which is equivalent to 12 timesteps. For SoPhie and Trajectron ++, the input length is set as eight timesteps, corresponding to 3.2 seconds.

It can be concluded that the trained policy achieved comparable or better performance than all other selected benchmarks across all three time periods. Notably, all methods, besides Trajectron ++, demonstrate their best performance during NOON compared to the AM and PM. This is likely due to the relatively less complex nature of pedestrian behaviour at noon, which is further discussed below. As stated in section 7.4, the main purpose of this comparison is to validate the subsequent investigation of the retrieved reward network, which is again considered as successful, based on these results.

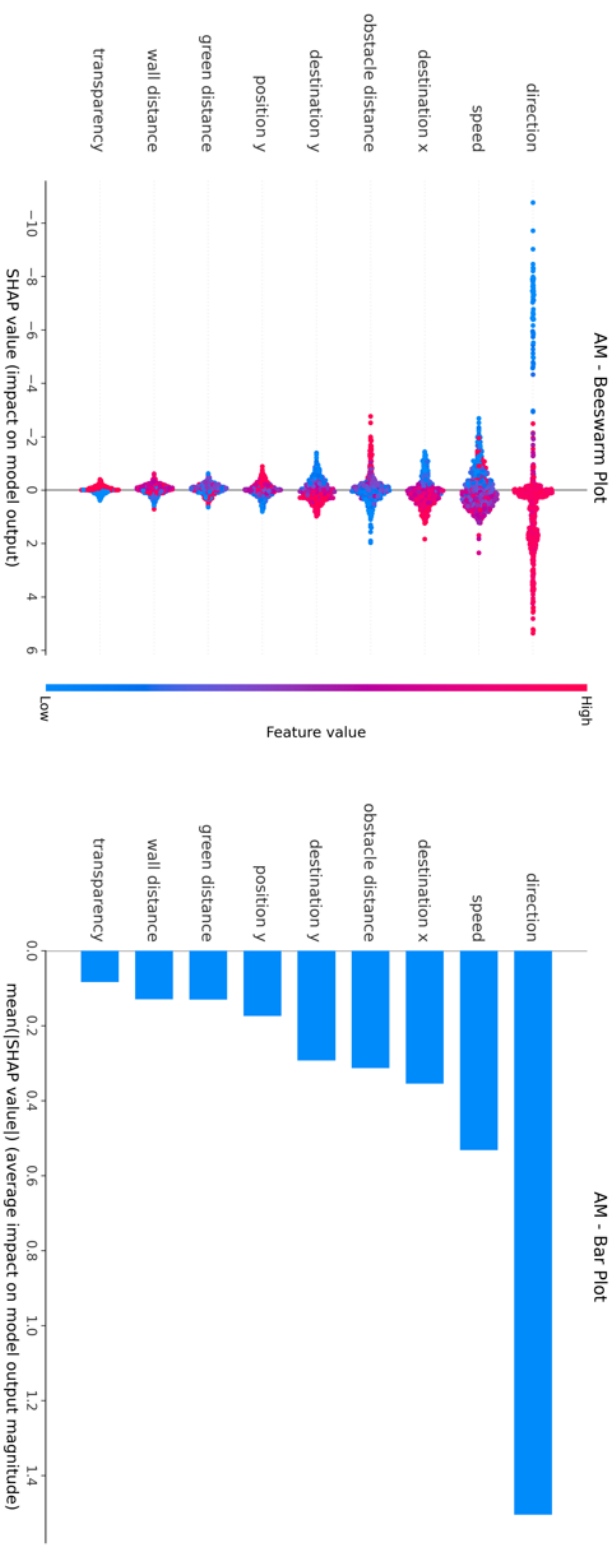



*Figure 8.3 Pedestrian decision making during morning time*

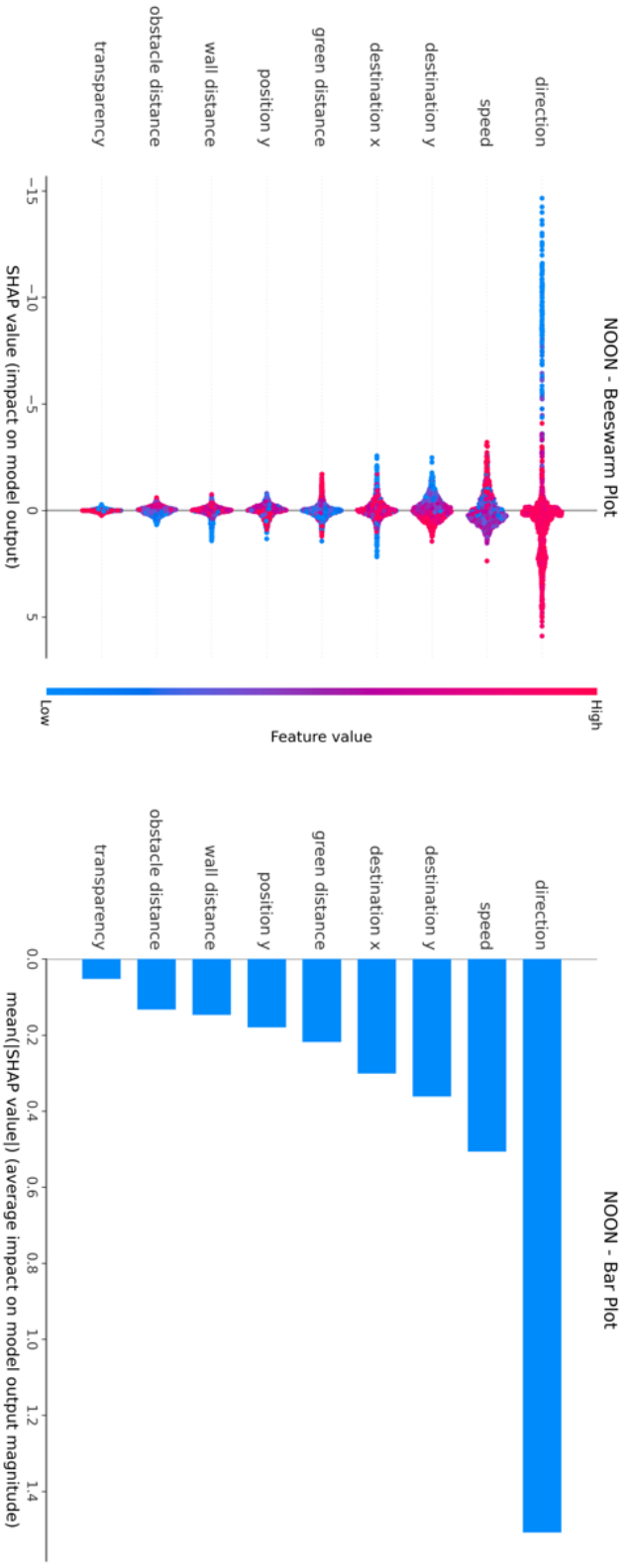



*Figure 8.4 Pedestrian decision making during noon time*

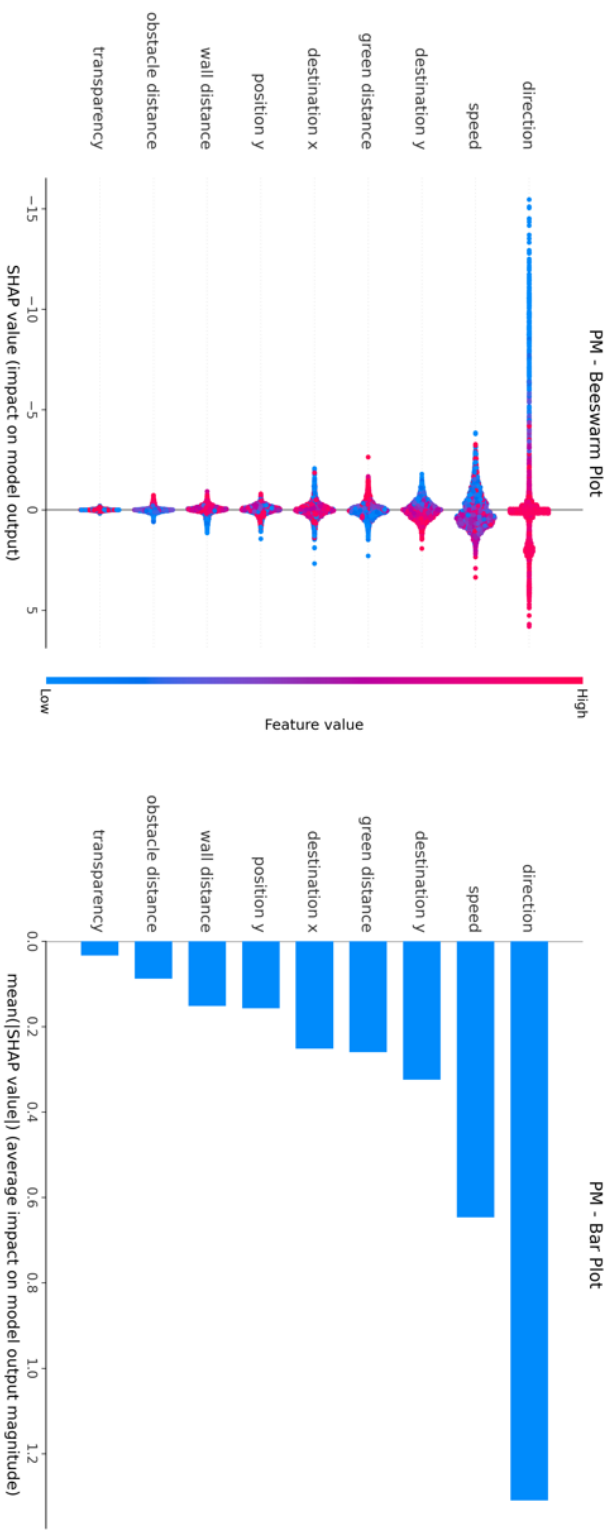



*Figure 8.5 Pedestrian decision making during evening time*

The results of gradient SHAP are also presented in both beeswarm and bar plot, with Figure 8.3, Figure 8.4 and Figure 8.5 demonstrating results for AM, NOON and PM respectively. In general, navigation features, especially direction and speed, demonstrate a higher impact on pedestrians' decision making than built environment features across three time periods. This pattern reflects both the data selection criteria, which only includes pedestrians travelling southbound from point A to point B, and also their tendency to maintain a preferred walking speed.

As this case study aims to investigate the impact of built environment, the corresponding SHAP values for built environment features are extracted and discussed for each time period in the following section. In particular, the relationship between each built environment feature value, e.g., the exact wall distance, and corresponding SHAP values are also plotted to locate the "comfort zone 'for pedestrians within the scene, following similar approach taken by Li (2024).

i) AM

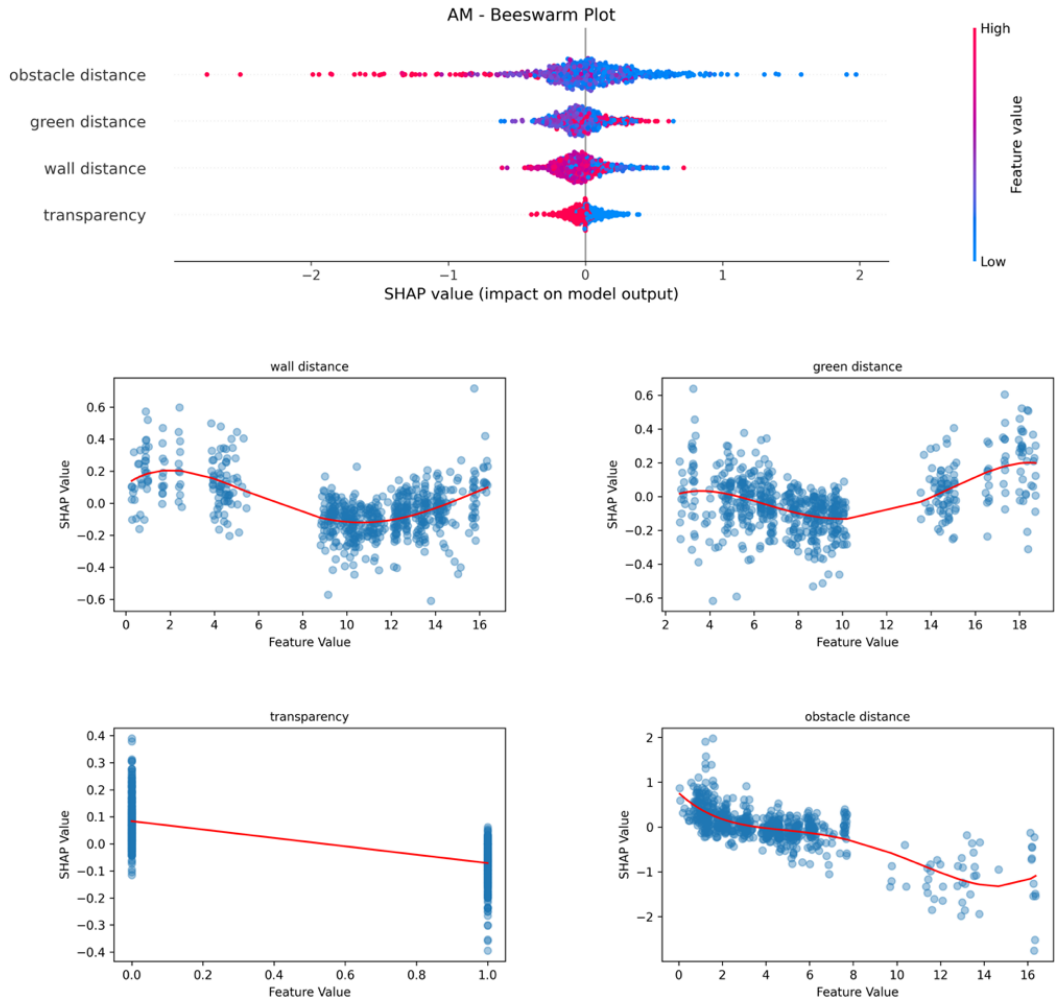



*Figure 8.6 Built environment features analysis during morning time*

The beeswarm plot for SHAP values of built environment features and their corresponding feature values are presented in Figure 8.6. It is evident that the horizontal distance to the obstacles is the most influential environmental feature for pedestrians during morning time, with a shorter distance associated with higher rewards. This is assumed to be due to the physical layout of the obstacles, which spread alongside the pedestrians' desired path, thus keeping a close distance to these can be seen as way of navigation guidance, consistent with the way finding theory proposed by Lynch (1960):

“Edges are the linear elements not considered as paths: they are usually, but not quite always, the boundaries between two kinds of areas. They act as lateral references.”

The prioritisation of obstacle-adjacent paths suggests a preference for maintaining a straight trajectory, possibly driven by the urgency associated with commuting during morning time.

Furthermore, although both the green distance and the wall distance show comparable influence on pedestrian behaviour, with a tendency to maintain around 2 metres from these features, their overall impact is considerably smaller than that of obstacle distance. Consequently, most pedestrians during the morning period are likely to favour paths that run alongside the edge of the obstacles, with the comfort zone likely falling within a 3-metres distance from them, corresponding to the threshold where SHAP values begin to turn negative based on Figure 8.6.

Notably, while the transparency of the walls shows minimal impact, its effect consistently negative for transparent walls and positive for opaque ones. This contrasts with the findings of Kesici & Erkan (2023), which observed greater pedestrian engagement near transparent walls. A possible explanation lies in the local lighting conditions: with the sun positioned in the west during the morning hours, most direct sunlight was blocked by the building itself, resulting in insufficient illumination in the area of interest. Under these conditions, the transparent wall appears darker and more visually ambiguous than the predominantly white solid wall, which is brighter. This contrast may reduce the perceived attractiveness of the transparent wall during the morning period. Future studies can examine how varying illumination conditions influence the attractiveness of transparent walls.

ii) NOON

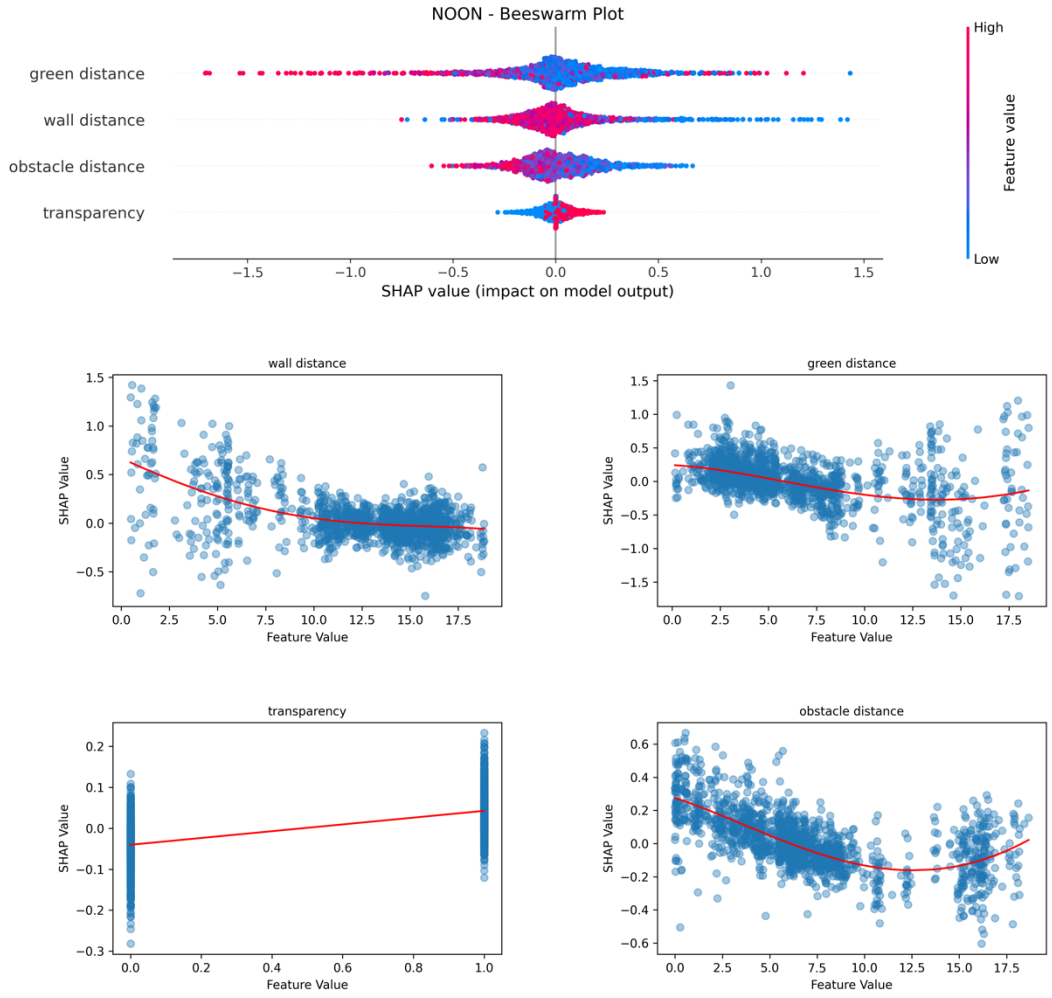



*Figure 8.7 Built environment features analysis during noon time*

Figure 8.7 presents the analysis of built environment features during noon. Similar to the morning behaviour, pedestrians demonstrate a clear tendency to stay closer to the green space, though with a shift toward preferring immediate proximity during noon period. In addition, green space distance becomes the most influential feature during this period, which is most likely to be attributed to the sun's position directly overhead at noon, resulting in intense direct sunlight. Consequently, the shaded area provided by the green space becomes highly desirable,

making proximity to it a top priority for pedestrians, consistent with the shade-seeking behaviour identified by Melnikov, et al. (2022).

Hence, the comfort zone for pedestrians during noon is located along the space beneath the green space, with a width of 5 metres according to Figure 8.7. The influence of the sun is also reflected in other features, as pedestrians show a preference for walking closer to the street wall, which offers shade as well, while the importance of obstacle distance diminishes in comparison.

Furthermore, with the sun having shifted away from the west, the illumination condition is improved for the area of interest, resulting in the transparent wall regaining its usual level of attractiveness to pedestrians. This is consistent with findings from Kesici & Erkan (2023) that more transparent walls attract people to look into it, leading to interactive engagement with pedestrians.

iii) PM

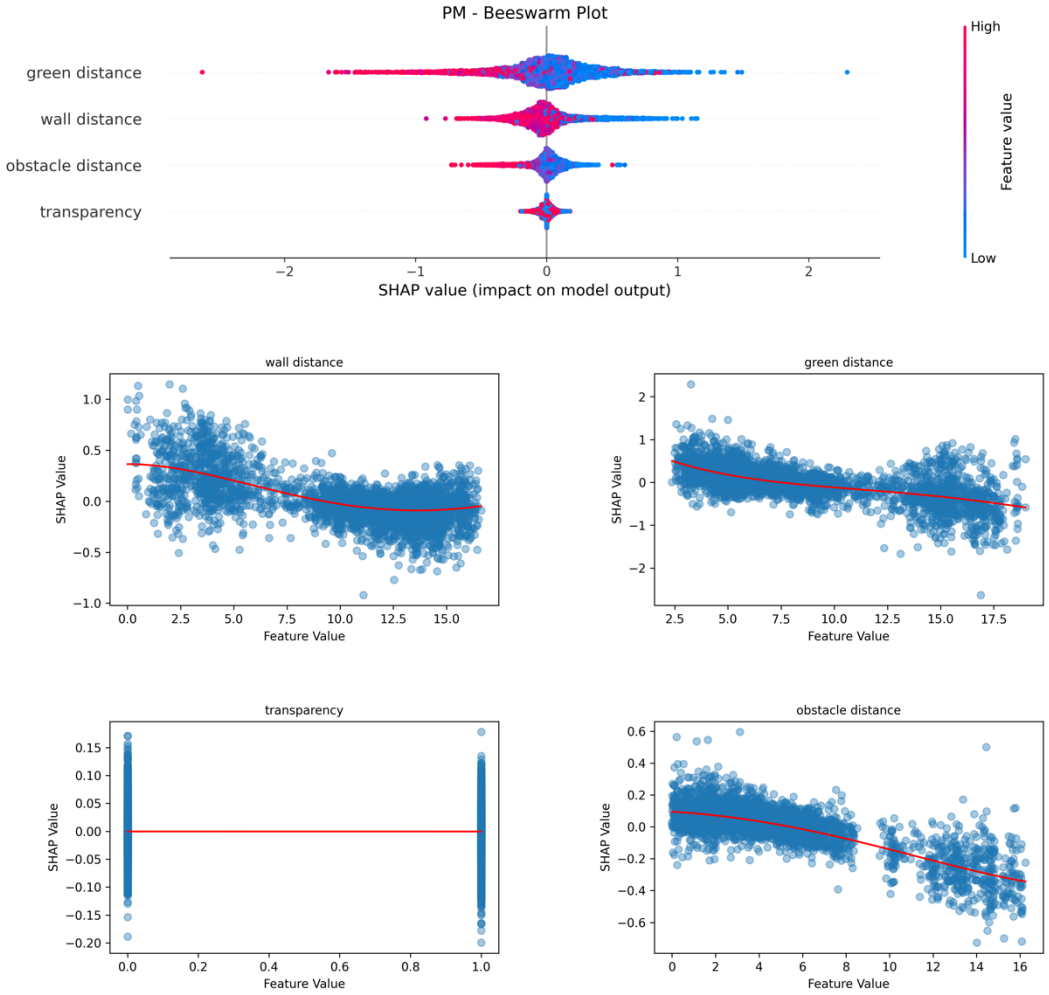



*Figure 8.8 Built environment features analysis during evening time*

According to Figure 8.8, built environment features demonstrate similar impact on pedestrians during evening time as observed at noon, both in terms of relative importance and SHAP correlation, except for the transparency. The comfort zone can be identified similar to the noon time as well, with a 5 metres wide path underneath the green space. The difference of comfort zone between AM and the later periods (NOON and PM) is likely to be attributed to walking preferences during leisure.

Based on Ge, et al. (2023), green landscapes and moderately enclosed spaces are preferred by pedestrians when walking for leisure, which typically occurs during the evening time (Mondal, et al., 2021). This is consistent with the findings in Figure 8.8, where getting closer to green spaces and street wall ranks highly among built environment features influencing pedestrian decision making. For transparent walls, no clear relationship with SHAP values has been identified, which may be due to additional visual cues, such as the greater prominence of interior lighting during the evening.

Another notable difference between noon and evening periods is each feature's relationship with their SHAP value. During evening, features such as green distance and obstacle distance exhibit a more linear relationship, suggesting that their influence on pedestrian behaviour is more direct. In contrast, the more nonlinear patterns observed at noon may reflect the presence of additional factors, such as the shape of shadows, contributing to decision making.

Similar to the work presented in section 7.4.2, the resulting SHAP value has been used to develop a SHAP-calibrated SFM for this scenario as well. In addition to the built environment features, i.e. wall distance $D_w$, green distance $D_g$, transparency of the street wall $T$ and horizontal distance with obstacles $D_o$, also included are the navigation features direction $\omega$ and destination $d$ due to their strong impact on pedestrian behaviour. The total social force $\vec{F}$ is then expressed as:

$$\vec{F} = \mathrm{SHAP}_{\omega}\overrightarrow{f_{\omega}} + \alpha \cdot \mathrm{SHAP}_{d}\overrightarrow{f_{d}}(d) + \beta \cdot \left((SHAP)_{D_w}\overrightarrow{f_{D_w}}(D_w) + SHAP_{D_g}\overrightarrow{f_{D_g}}(D_g) + SHAP_{D_o}\overrightarrow{f_{D_o}}(D_o) + SHAP_{T}\overrightarrow{f_{T}}(T)\right) \quad (8.3)$$

Where $\overrightarrow{f_{\omega}}$ is the direction of southbound to reflect their general movement tendency, and $\alpha$ and $\beta$ are scaling factors. For each time period, 20% of available initial points are randomly sampled. The SHAP-calibrated SFM was simulated forward over a 4.8-second prediction horizon, with its performance assessed using ADE and FDE, as presented in Table 8.2.

| Time period | ADE | FDE |
|---|---|---|
| AM | 1.48 | 2.96 |
| NOON | 1.40 | 2.72 |
| PM | 1.42 | 2.80 |

*Table 8.2 Simulation results of SFM*

By comparing with Table 8.1, the results have demonstrated again that the SHAP-calibrated SFM can achieve comparable results with state-of-art methods, highlighting the potential application of the proposed decision making inference framework as a less computationally intensive modelling approach.

#### *8.4.1.4 Summary*

In this case study, the proposed decision making inference framework was applied to an urban planning study, underscoring its potential scalability.

This case study can be viewed as an ideal extension of the research presented in chapter 7, as it demonstrates the framework's capability to capture the impacts of a broader range of environmental features, which contrasts with the jostle scenario that focused on interactions near an entrance gate.

At the same time, it also provides a valuable point of comparison, since the data used in this case study are more conventional, i.e., more widely used in the field of crowd analysis, extracted from video footage with a considerably larger spatial range of movement. This further highlights the versatility and applicability of the proposed framework in pedestrian-related studies. However, it should be noted that the proposed framework requires accurate pedestrian positional data and therefore depends on the reliability of data extraction methods. For example, in extremely dense crowds, severe occlusion can reduce the accuracy of video-based detectors such as YOLO.

# 9. Conclusion and future work

## 9.1 Revisiting the research aim and objectives

As stated in chapter 1, the aim of this research is to contribute to a more comprehensive understanding of crowd disasters by incorporating social aspects, and to develop methods for analysing crowd movement and behaviour. To accomplish this aim, the following objectives are set in details:

1. Define qualitative characteristics of crowd disaster.

2. Construct a database of past crowd disasters.

3. Validate the usage of virtual reality (VR) in crowd analysis study.

4. Identify movement patterns under specific scenario settings.

5. Propose an approach to infer decision making behind crowd behaviour.

6. Propose improvement in the practical crowd management based on the research results.

In this section, each research objective is revisited to evaluate how it was addressed and its main findings.

### 9.1.1 Define qualitative characteristics of crowd disaster

The qualitative characteristics of crowd disasters incorporating the socio-technical system perspective were defined in chapters 2 and 5, including three causal factors from the established FIST model: Time, Information and Physical space, together with two subsystems of the

external environment: Social scenario and Physical environment and Casualty to measure the severity of the consequences.

Within this objective, a working definition of crowd disasters was established as “An incident in which people are injured or die caused by a high density of surrounding people.” The importance of considering the impact from social scenario on crowd behaviour was validated through the 2014 Shanghai crowd disaster case study, emphasising the need of incorporating a socio-technical system perspective. In addition, a set of indicators was summarised for identifying the presence of three causal factors through a review of the relevant literature, and a taxonomy was proposed for each identified characteristic, accompanied by the development of a codebook to facilitate the construction of the crowd disaster database.

### 9.1.2 Construct a database of past crowd disasters

The construction process of a crowd disaster database based on defined characteristics was discussed in chapter 5, with news articles from the subscribed database Factiva as the primary data source, and other Internet websites as supplementary source. The database includes all identified crowd disaster events occurring between 1989 and 2021, each of which was systematically coded using the developed codebook.

The resulting database demonstrated high consistency, completeness, and inter-rater reliability, supporting the robustness of the taxonomy and coding procedure. The study also showed that news articles serve as a valid source for crowd disaster analysis, given the lack of official reports. However, language can be a barrier to making data collection more inclusive. For example, some incidents in the database were only documented in Chinese, suggesting that similar gaps may exist for other languages. Nine distinct types of crowd disasters were identified through K-Modes cluster analysis, each representing unique combinations of characteristics, highlighting the necessity to investigate further if a refined crowd management plan is desired.

### 9.1.3 Assess the feasibility and reliability of using virtual reality (VR) for investigating crowd behaviour in specific scenario.

A major objective of this research was to validate VR as a reliable scientific tool and trustworthy data source in selected scenario. This validation exercise is detailed in chapter 6, and was addressed in two parts. First, the selected jostle scenario was successfully reproduced in the virtual environment, both in terms of phenomena emergence and participant feedback. Second, a consistency analysis was conducted on the collected behavioural data, revealing that participants exhibited stable patterns of movement both across different experimental rounds and within different time intervals of the same round. These results collectively support the use of VR as both a scientifically rigorous and ethically viable approach to investigating dangerous crowd scenarios.

Further discussions presented in section 8.3 demonstrated that the development of a virtual environment requires interdisciplinary skillsets, combining interactive content design and coding. This combination is rarely possessed by a single researcher within the field of crowd analysis. In addition, the environmental space of VR-based study requires much more delicate design and a much longer preparation time than physical experiment.

Conducting a multi-participant VR-based study also remains challenging. While there is a lack of industry-developed tools, organizational-level solutions may face undesirable technical issues. Individually developed tools, on the other hand, are time-consuming and ineffective due to their lack of being reusable. Consequently, VR-based studies are typically more vulnerable to technical issues than any conventional physical experiment.

Several additional considerations were identified. Sensory substitution remains a challenging topic for a VR-based study, and special consideration must be given to participant recruitment criteria, especially regarding proficiency in avatar control. Nevertheless, when scenarios are successfully reproduced, VR can provide a highly controlled and accurate source of behavioural data.

### 9.1.4 Identify movement patterns under specific scenario settings

In chapter 6, two types of behaviour in the selected jostle scenario were identified through agglomerative hierarchical clustering: an ‘impatient’ behaviour, due to their strong tendency to maintain their speed, and a ‘patient’ behaviour, as they follow daily queuing behaviour.

Several additional findings were identified. First, change point detection algorithms from the signal processing field can be used to identify the spatial range of the crowd by locating the point of an abrupt behavioural change. Secondly, none of the trajectories from the individual participants exhibited a clear dominance of a single behavioural type, suggesting that the identified patterns are better interpreted as movement strategies rather than stable personal preferences. Finally, arch formation is a characteristic emergent phenomenon of the jostle scenario, primarily driven by the gap-seeking of impatient behaviour.

### 9.1.5 Propose an approach to infer decision making behind crowd behaviour across various scenarios.

In chapter 7, a framework combining IRL and model interpretability was proposed, utilizing the retrieved reward function as an approach to quantitatively understand the decision making of certain crowd behaviours. Within this framework, crowd behaviour within the jostle scenario can be effectively modelled using a MDP with a state space comprising self-state, surrounding-state, and environmental-state variables, and an action space defined by change of speed and change of direction. The policy trained through an AIRL algorithm based on the model formulated achieved comparable performance against common benchmarks, both in terms of ADE and FDE.

In addition, factors within the defined state and action spaces were assessed quantitatively for their influence on the decision-making processes of the two identified behavioural types. These results not only validate the behavioural strategies outlined in chapter 6, but also provide a deeper understanding into interactions with surrounding crowds and the external environment.

Building on the proposed framework, the quantified impact of each factor can serve as a starting point for a SHAP-calibrated SFM, which demonstrates competitive performance relative to

advanced ML-based methods, while requiring significantly less computational resources and expert intervention.

Finally, the proposed approach can also be applied to other pedestrian-related fields, such as urban planning, by using more conventional video-extracted data and demonstrating its ability to capture the impacts of various external environmental factors.

### 9.1.6 Proposed improvements in practical crowd management

In chapter 8, three contributions to real-world crowd management practices were proposed after a review of operational guides adopted by the crowd management industry and these were further examined through interviews with two major event industry practitioners. Consequently, a section on tangible crowd behaviours was developed based on the research findings to serve as a valuable data source for the crowd management industry.

Two additional insights were identified during the interviews with stakeholders. First, small scale events in the UK, i.e. those with fewer than 5,000 participants, typically do not have a crowd management plan in place. This includes the jostle scenarios identified in this research. However, such events have started to consider developing these plans following the 2022 Brixton crowd disaster. Therefore, the outcomes of this research can provide a useful starting point for these organisers, particularly given their lack of prior experience. Second, due to both cost concerns and no legal requirements, crowd modelling, e.g. using simulation modelling, is not commonly used for small events. This highlights a critical gap, i.e. the limited availability of experts in the industry.

## 9.2 Future work

Future research in three areas is recommended based on the discussion points detailed throughout this thesis, especially in chapter 8.

The experimental findings were derived from a single jostle scenario and a relatively homogeneous participant sample, which may limit transferability to other crowd contexts and

populations. Additional demographic variables, including age, gender, cultural background, and substance use, may influence behavioural responses and should be systematically examined in future studies. Extending the experimental design to a broader range of crowd situations, such as the remaining disaster types identified in the clustering analysis, would further support the generalisation of the behavioural taxonomy proposed in this research.

In the VR experiment, although visual cues substitution enabled scalable multi-participant interaction, physical contact and sensory feedback were simplified relative to real crowds. Future work should therefore investigate improved multimodal interaction techniques, including haptic augmentation and alternative locomotion control schemes, as well as methods for streamlining the construction and reuse of high-fidelity virtual environments. These developments would strengthen ecological validity and enable more realistic simulation of dense crowd conditions.

Finally, the proposed decision making inference framework was validated primarily within the jostle scenario and using the modelling approaches available at the time of study. Further validation across diverse pedestrian domains, including real-world trajectory datasets and operational crowd management contexts, is required to establish robustness and scalability. Integrating physiological sensing (e.g., heart rate, respiration) and exploring real-time behaviour classification methods would support in-event monitoring applications. In addition, incorporating the quantified decision-making factors into conventional crowd models may help bridge the gap between advanced crowd modelling techniques and industry needs for reduced computational cost and expert input.

## 9.3 Final thoughts

The aim of this research was to contribute to a new, comprehensive understanding of crowd disasters by incorporating to a greater extent than at present, the social aspects involved, and to explore novel methods for understanding crowd movement and behaviour within this new framework. Upon revisiting each set objective, it is reasonable to conclude that this aim has been well achieved in this thesis.

The underlying motivation for this research lies in fully incorporating the socio-technical systems perspective, which emphasises socially shaped behaviour and the influence of the external social environment. These factors have long been neglected due to a prevailing culture of technocracy. While technically advanced methodologies such as VR and ML were employed in this study to investigate movement patterns and understand behavioural decision making, careful consideration was given to their capacity to capture social aspects. This was addressed both during the methodological design process and in the interpretation of the results, reflecting the strengths of the pragmatic worldview adopted in this research.

It is hoped that this thesis will contribute to a more refined real-life crowd management practices, ultimately leading to more effective prevention of crowd disaster. Furthermore, the strategy adopted in this research is intended to inform and inspire future studies, both in terms of theoretical development and methodological innovation, within the broader field of crowd analysis and beyond.

# Bibliography

4028mdk09, 2010. *Crowd control barrier (Absperrgitter der Firma Doublet GmbH)..* [Online]
Available at: https://commons.wikimedia.org/wiki/File:Absperrgitter.JPG
[Accessed 9 March 2026].

Abbeel, P. & Ng, A. Y., 2004. *Apprenticeship learning via inverse reinforcement learning.* New York, NY, the twenty-first international conference on Machine learning (ICML '04).

Adobe, 2025. *Mixamo.* [Online]
Available at: https://www.mixamo.com/#/
[Accessed 27 Jan 2025].

Adrian, J. et al., 2025. Glossary for Research on Human Crowd Dynamics - 2nd Edition. *Collective Dynamics,* Volume 10, pp. 1-32.

Alahi, A. et al., 2016. *Social LSTM: Human Trajectory Prediction in Crowded Spaces.* Las Vegas, NV, USA, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Alasmari, A. M., Farooqi, N. S. & Alotaibi, Y. A., 2024. Recent trends in crowd management using deep learning techniques: a systematic literature review. *Journal of Umm Al-Qura University for Engineering and Architecture,* Volume 15, pp. 355-383.

Aldayri, A. & Albattah, W., 2022. Taxonomy of Anomaly Detection Techniques in Crowd Scenes. *Sensors,* 22(16).

Alkhadim, M., Gidado, K. & Painting, N., 2018. Risk management: The effect of FIST on perceived safety in crowded large space buildings. *Safety Science,* Volume 108, pp. 29-38.

Allievi, W. B. K., Banzhaf, H. & Stone, F. S., 2023. Reward (Mis)design for autonomous driving. *Artificial Intelligence,* Volume 316.

AlMahamid, F. & Grolinger, K., 2022. Agglomerative Hierarchical Clustering with Dynamic Time Warping for Household Load Curve Clustering. *arXiv:2210.09523.*

Almukhtar, S. & Watkins, D., 2016. *How One of the Deadliest Hajj Accidents Unfolded.* [Online]
Available at: https://www.nytimes.com/interactive/2016/09/06/world/middleeast/2015-hajj-stampede.html
[Accessed 12 July 2025].

Alsaleh, R. & Sayed, T., 2020. Modeling pedestrian-cyclist interactions in shared space using inverse reinforcement learning. *Transportation Research Part F: Traffic Psychology and Behaviour,* Volume 70, pp. 37-57.

Ameen, M. & Stone, R., 2023. Advancements In Crowd-Monitoring System: A Comprehensive Analysis of Systematic Approaches and Automation Algorithms: StaADVANCEMENTS IN CROWD-MONITORING SYSTEM: A COMPREHENSIVE

ANALYSIS OF SYSTEMATIC APPROACHES AND AUTOMATION ALGORITHMS: STATE-OF-THE-ART. *International Journal of Network Security & Its Applications,* 15(3).

Aminikhanghahi, S. & Cook, D. J., 2017. A survey of methods for time series change point detection. *Knowledge and Information Systems,* Volume 51, pp. 339-367.

Ando, K., Ota, H. & Oki, T., 1988. Forecasting the flow of people. *Railway Research Review,* 45(8), pp. 8-14.

Arias, S., Mossberg, A., Nilsson, D. & Wahlqvist, J., 2021. A Study on Evacuation Behavior in Physical and Virtual Reality Experiments. *Fire Technology,* Volume 58, pp. 817-849.

Arias, S. et al., 2021. Pursuing behavioral realism in Virtual Reality for fire evacuation research. *Fire and Materials,* Volume 45, pp. 462-472.

Artime, O. & Domenico, M. D., 2022. From the origin of life to pandemics: emergent phenomena in complex systems. *Phil. Trans. R. Soc. A.,* Volume 380.

Banuri, S., Danková, K. & Keefer, P., 2024. An experimental test of gaming incentive. *Journal of Economic Behavior & Organization,* Volume 224, pp. 825-850.

Barr, D. et al., 2024. Explaining a collective false alarm: Context and cognition in the Oxford Street crowd flight incident. *European Journal of Social Psychology,* Volume 54, pp. 1515-1530.

Basu, N. et al., 2022. A systematic review of the factors associated with pedestrian route choice. *Transport Reviews,* 42(5), pp. 672-594.

Bateman, G. & Majumdar, A., 2020. Characteristics of emergency evacuations in airport terminal buildings: A. *Safety Science.*

Bazior, G., Wąs, J. & Pałka, D., 2025. Pedestrian dynamics model for high densities. *Expert Systems with Applications,* Volume 272.

BBC, 2022. *Hillsborough: Timeline of the 1989 stadium disaster.* [Online]
Available at: https://www.bbc.co.uk/news/uk-england-merseyside-47697569
[Accessed 12 July 2025].

Beek, A. v., Duives, D. C., Feng, Y. & Hoogendoorn, S. P., 2024. Comparison of pedestrian wayfinding behavior between a real and a virtual multi-story building – A validation study. *Transportation Research Part C: Emerging Technologies,* Volume 163.

Bell, E., Bryman, A. & Harley, B., 2018. *Business Research Methods.* Oxford: Oxford University Press.

Belotti, M. & Martins, F., 2021. Analysis and verification of the social forces model in pedestrian lane formation scenarios.. *J Braz. Soc. Mech. Sci. Eng.,* Volume 43.

Bendali-Braham, M. et al., 2021. Recent trends in crowd analysis: A review. *Machine Learning with Applications,* Volume 4.

Bertalanffy, L. V., 1972. The History and Status of General Systems Theory. *The Academy of Management Journal,,* 15(4), pp. 407-426.

Berton, F. et al., 2022. Crowd navigation in VR: Exploring haptic rendering of collisions. *IEEE Transactions on Visualization and Computer Graphics,* 28(7), pp. 2589-2601.

Bhaskar, R., 2008. *A Realist Theory of Science.* 1st Edition ed. s.l.:Routeledge.

Bhuvaneswari, K. & Rauf, H. A., 2009. *Edgelet based human detection and tracking by combined segmentation and soft decision.* Perundurai, India, International Conference on Control, Automation, Communication and Energy Conservation.

Bivina, G. R. & Parida, M., 2020. Prioritizing pedestrian needs using a multi-criteria decision approach for a sustainable built environment in the Indian context. *Environment, development and sustainability,* Volume 22, pp. 4929-4950.

Boccara, N., 2010. *Modeling Complex Systems.* s.l.:Springer.

Bradski, G., 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools,* Volume 120, pp. 122-125.

Brockman, G. et al., 2016. OpenAI Gym. *arXiv:1606.01540.*

Cavallo, V., Dommes, A., Dang, N.-T. & Vienne, F., 2019. A street-crossing simulator for studying and training pedestrians. *Transportation Research Part F: Traffic Psychology and Behaviour,* Volume 61, pp. 217-228.

Challenger, R., Clegg, C. W. & Robinson, M. A., 2010. *Understanding crowd behaviours, Vol. 1 Practical guidance and lessons identified ,* London: The Stationery Office (TSO) ..

Chalvatzaki, G., Papageorgiou, X. S., Maragos, P. & Tzafestas, C. S., 2019. Learn to Adapt to Human Walking: A Model-Based Reinforcement Learning Approach for a Robotic Assistant Rollator. *Robotics and Automation Letters,* 4(4), pp. 3774-3781.

Chan, A. B., Liang, Z.-S. J. & Vasconcelos, N., 2008. *Privacy preserving crowd monitoring: Counting people without people models or tracking.* Anchorage, AK, USA, IEEE Conference on Computer Vision and Pattern Recognition.

Chang, A. H., Wang, S.-J., Anandjiwala, A. & Hsu, E. B., 2024. A Crowd Disaster Study: The Itaewon Seoul Crush. *Cureus,* 16(9).

Chebib, A. M. et al., 2025. *UKCMA.* [Online]
Available at: https://www.ukcma.com/
[Accessed 29 April 2025].

Chelsea Finn, P. C. ,. P. A. S. L., 2016. A Connection Between Generative Adversarial Networks, Inverse Reinforcement Learning, and Energy-Based Models. *abs/1611.03852.*

Chen, J. & Li, N., 2025. Crowdsourced virtual reality experimental approach in pedestrian and evacuation dynamics research: Part I. Process model and digital platform. *Advanced Engineering Informatics,* Volume 65.

Chen, Y. et al., 2019. Mocycle-GAN: Unpaired Video-to-Video Translation. *arXiv:1908.09514* .

Chesney, T., 2006. An empirical examination of Wikpedia's credibility.. *First Monday,* 11(11).

Coates, W. C., Jordan, J. & Clarke, S. O., 2021. A practical guide for conducting qualitative research in medical education: Part 2—Coding and thematic analysis. *AEM Educ Train.,* 5(4), p. 10645.

Conover, W. J., 1999. *Practical Nonparametric Statistics.* 3rd Edition ed. New York: John Wiley & Sons.

Corbetta, A. et al., 2017. Fluctuations around mean walking behaviors in diluted pedestrian flows. *Phys. Rev. E,* Volume 95.

Couto, R., Leal, J., Costa, P. M. & Galvão, T., 2015. *Exploring Ticketing Approaches Using Mobile Technologies: QR Codes, NFC and BLE.* Gran Canaria, Spain, IEEE 18th International Conference on Intelligent Transportation Systems.

Cremades, A., Hoyas, S. & Vinuesa, R., 2024. Additive-feature-attribution methods: a review on explainable artificial intelligence for fluid dynamics and heat transfer. *arXiv:2409.11992* .

Creswell, J. W., 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* 4th Edition ed. s.l.:SAGE Publications.

CY, C. & WHK, L., 1998. Pedestrian Route Choices between Escalator and Stairway in MTR Stations.. *J Transp Eng,* 124(3), pp. 277-285.

Dündar, S., Gokkurt, B. & Soylu, Y., 2012. Mathematical Modelling at a Glance: A Theoretical Study.. *Procedia - Social and Behavioral Sciences.,* Volume 46.

Daamen, W., Hoogendoorn, S. P. & Piet H. L. BovyDepartment of Transport and Planning, F. o. C. E. a. G. U. o. T. S. 1. P. B. 5. 2. G. D. R. R., 2005. First-Order Pedestrian Traffic Flow Theory. *Transportation Research Record,* 1934(1), pp. 43 - 52.

Dai, Z. et al., 2024. A study of pedestrian wayfinding behavior based on desktop VR considering both spatial knowledge and visual information. *Transportation Research Part C: Emerging Technologies,* Volume 163.

Dalal, N. & Triggs, B., 2005. *Histograms of oriented gradients for human detection.* San Diego, CA, USA, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).

DeAngeles, D., Schurr, M., Birnbaum, M. & Harms, B., 1998. Traumatic asphyxia following stadium crowd surge: stadium factors affecting outcome. *WMJ,* 97(9), pp. 42-45.

Demrozi, F. et al., 2023. A Comprehensive Review of Automated Data Annotation Techniques in Human Activity Recognition. *arXiv:2307.05988.*

Dickie, J., 1995. Major crowd catastrophes. *Safety Science,* Volume 18, pp. 309-320.

Direkoglu, C., 2020. Abnormal Crowd Behavior Detection Using Motion Information Images and Convolutional Neural Networks. *IEEE Access,* 8(80408-80416).

Dong, Y.-H.et al., 2017. Emergency preparedness for mass gatherings: Lessons of "12.31" stampede in Shanghai Bund. *Chin J Traumatol,* 20(4), pp. 240-242.

Dow Jones & Reuters, 2023. *Factiva.* [Online]
Available at: https://visit.dowjones.com/factiva/lp/dg/factiva-1/?LS=Search&utm_medium=cpc&utm_source=google&utm_campaign=EMEA-UK[EN]_GGL-Brand[GEN]-FA-Factiva_MT-Exact_EXP-20221115&CID=7015Y000003wt4UQAQ&utm_term=factiva_(e)&utm_content=&gclid=CjwKCAjw4P6oBhBsEiwAKYVk
[Accessed 6 October 2023].

Drăgulinescu, A. M. C., Marcu, I., Halunga, S. & Fratu, O., 2018. *Persons Counting and Monitoring System Based on Passive Infrared Sensors and Ultrasonic Sensors (PIRUS)..* Cham, Pervasive Computing Paradigms for Mental Health, N. Oliver et al., Eds., Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 207, Springer.

Drury, J. et al., 2019. Facilitating Collective Psychosocial Resilience in the Public in Emergencies: Twelve Recommendations Based on the Social Identity Approach. *Front Public Health,* 7(141).

Drury, J. & Reicher, S., 2020. Crowds and Collective Behavior. In: *Oxford Research Encyclopedia of Psychology.* s.l.:Oxford University Press.

Duan, J., Zhai, W. & Cheng, C., 2020. Crowd Detection in Mass Gatherings Based on Social Media Data: A Case Study of the 2014 Shanghai New Year's Eve Stampede. *International Journal of environmental research and public health,* 17(22).

Dubey, S. R., Singh, S. K. & Chaudhuri, B. B., 2022. Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark. *arXiv:2109.14545.*

Dupuy, M., 2012. *Framework for the Analysis of Separation-related Incidents in Aviation.,* s.l.: Imperial College London.

Du, Y. et al., 2023. Strongsort: Make deepsort great again. *IEEE Transactions on Multimedia.*

Easa, S. M., 2016. Pedestrian Crossing Sight Distance: Lateral Clearance Guidelines for Roadways. *Transportation Research Record: Journal of the Transportation Research Board,* 2588(1).

Efron, B. & Tibshirani, R., 1994. *An Introduction to the Bootstrap.* 1st Edition ed. New York: Chapman and Hall/CRC.

Ehsanpour, M. et al., 2021. JRDB-Act: A Large-scale Dataset for Spatio-temporal Action, Social Group and Activity Detection. *arXiv:2106.08827* .

Elaiw, A., Al-Turki, Y. & Alghamdi, M., 2019. A Critical Analysis of Behavioural Crowd Dynamics—From a Modelling Strategy to Kinetic Theory Methods. *Symmetry,* 11(7).

Elliott, D. & Smith, D., 1993. Football stadia disasters in the United Kingdom: learning from tragedy?. *Industrial & Environmental Crisis Quarterly,* 7(3), pp. 205-229.

Events Industry Forum, 2022. *The purple guide.* [Online]
Available at: https://www.thepurpleguide.co.uk/index.php
[Accessed 18 June 2025].

Everitt, B. S., 1972. Cluster analysis: A brief discussion of some of the problems.. *The British Journal of Psychiatry,* 120(555), pp. 143-145.

Everitt, B. S., Landau, S., Leese, M. & Stahl, D., 2011. *Cluster analysis.* 5th Edition ed. s.l.:Wiley.

Ewing, R. et al., 2015. Streetscape Features Related to Pedestrian Activity. *Journal of Planning Education and Research*, 36(1), pp. 5-15.

Fahad, M., Chen, Z. & Guo, Y., 2018. *Learning How Pedestrians Navigate: A Deep Inverse Reinforcement Learning Approach.* Madrid, IEEE.

Fang, Z. M. et al., 2019. Experimental study on the movement characteristics of 5–6 years old Chinese children when egressing from a pre-school building. *Safety Science,* Volume 113, pp. 264-275.

Feliciani, C., 2023. *List of crowd accidents from 1900 to 2019.* [Online]
Available at: https://doi.org/10.5281/zenodo.7523480
[Accessed 13 February 2026].

Feliciani, C., Corbetta, A., Haghani, M. & Nishinari, K., 2023. Trends in crowd accidents based on an analysis of press reports. *Safety Science,* Volume 164, p. 106174.

Feng, Y., Duives, D. C. & Hoogendoorn, S. P., 2021. Using virtual reality to study pedestrian exit choice behaviour during evacuations. *Safety Science,* Volume 137.

Feng, Y., Duives, D. & Hoogendoorn, S., 2021. Development of a VR tool to study pedestrian route and exit choice behaviour in a multi-story building. *arXiv:2103.05560.*

Fiandeiro, M., Nguyen, T. T., Wong, H. & Hsu, E. B., 2023. Modernized Crowd Counting Strategies for Mass Gatherings—A Review. *Journal of Acute Medicine,* 13(1), pp. 4-11.

Flötteröd, G. & Lämmel, G., 2015. Bidirectional pedestrian fundamental diagram. *Transportation Research Part B: Methodological,* pp. 194-212.

Foster, D. J., Block, A. & Misra, D., 2024. Is Behavior Cloning All You Need? Understanding Horizon in Imitation Learning. *arXiv:2407.15007* .

Fourati, J., Issaoui, B. & Zidi, K., 2017. *Literature Review of Crowd Management: A Hajj Case Study.* s.l., ICINCO.

Fruin, J., 1993. *The causes and prevention of crowd disasters.* London, International Conference on Engineering for Crowd Safety.

Fruin, J. J., 1971. *Designing for pedestrians: a level-of-service concept.* Washington District of Columbia, 50th Annual Meeting of the Highway Research Board.

Fu, J., Luo, K. & Levine, S., 2018. *Learning robust rewards with adversarial inverse reinforcement learning.* Vancouver, International Conference on Learning Representations.

Fu, L. et al., 2024. The effect of the ticket gate layout on mixed-flow movement including luggage-laden pedestrians in metro stations. *Safety Science,* Volume 179, p. 106625.

Galea, E. et al., 2017. An international survey and full-scale evacuation trial demonstrating the effectiveness of the active dynamic signage system concept. *Fire Mater,* Volume 41, pp. 493-513.

Garcimartín, A. et al., 2016. Flow of pedestrians through narrow doors with different competitiveness. *Journal of Statistical Mechanics: Theory and Experiment,* Volume 2016.

Ge, Y., He, Z. & Shang, K., 2023. Influence of the Built Environment on Pedestrians' Route Choice in Leisure Walking. *ISPRS Int. J. Geo-Inf,* 12(9).

Gill, J. R. M. & Landi, K. M., 2004. Traumatic Asphyxial Deaths Due to an Uncontrolled Crowd. *The American Journal of Forensic Medicine and Pathology,* 25(4), pp. 358-361.

Gleave, A. et al., 2022. imitation: Clean Imitation Learning Implementations. *arXiv:2211.11972.*

Gonon, D. & Billard, A., 2023. Inverse Reinforcement Learning of Pedestrian–Robot Coordination. *Robotics and Automation Letters,* 8(8), pp. 4815-4822.

Goodfellow, I., Bengio, Y. & Courville, A., 2016. *Deep Learning.* s.l.:MIT Press.

Goodfellow, I. J. et al., 2014. Generative Adversarial Networks. *arXiv:1406.2661.*

Gordon, C. C. et al., 1989. *Anthropometric Survey of U.S. Army Personnel: Summary Statistics, Interim Report for 1988,* Yellow Springs, Ohio: ANTHROPOLOGY RESEARCH PROJECT INC.

Gupta, A. et al., 2018. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. *arXiv:1803.10892.*

Haberfellner, R., de Weck, O., Fricke, E. & V¨ossner, S., 2019. Systems Engineering - Fundamentals and Application.. *Springer Nature..*

Haghani, M., 2020. Empirical methods in pedestrian, crowd and evacuation dynamics: Part I. Experimental methods and emerging topics. *Safety Science,* Volume 129.

Haghani, M. et al., 2023. A roadmap for the future of crowd safety research and practice: Introducing the Swiss Cheese Model of Crowd Safety and the imperative of a Vision Zero target. *Safety Science,* Volume 168, p. 106292.

Haghani, M. et al., 2023. A roadmap for the future of crowd safety research and practice: Introducing the Swiss Cheese Model of Crowd Safety and the imperative of a Vision Zero target. *Safety Science,* Volume 168, p. 106292.

Haghani, M. et al., 2019. Panic, Irrationality, and Herding: Three Ambiguous Terms in Crowd Dynamics Research. *Journal of Advanced Transportation.*

Haghani, M. & Ronchi, E., 2024. Revisiting the paper ”Simulating dynamical features of escape panic”: What have we learnt since then?. *Collective Dynamics,* Volume 9, pp. 1-11.

Haghani, M. & Sarvi, M., 2017. Stated and revealed exit choices of pedestrian crowd evacuees. *Transportation Research Part B: Methodological,* Volume 95, pp. 238-259.

Hassani, H. et al., 2022. Classical dynamic consensus and opinion dynamics models: A survey of recent trends and methodologies. *Information Fusion,* Volume 88, pp. 22-40.

Hassner, T., Itcher, Y. & Kliper-Gross, O., 2012. *Violent flows: Real-time detection of violent crowd behavior.* Providence, RI, USA, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.

Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction..* 2nd Edition ed. s.l.:Springer.

Helbing, D., 1992. Models for Pedestrian Behavior. In: *Natural Structures. Principles, Strategies, and Models in Architecture and Nature.* Stuttgart: s.n.

Helbing, D., Buzna, L., Johansson, A. & Werner, T., 2005. Self-Organized Pedestrian Crowd Dynamics: Experiments, Simulations, and Design Solutions. *Transportation Science,* 39(1), pp. 1-24.

Helbing, D., Farkas, I. J. & Vicsek, T., 2000. Freezing by Heating in a Driven Mesoscopic System. *Phys. Rev. Lett.,* Volume 84, p. 1240.

Helbing, D., Farkas, I. & Vicsek, T., 2000. Simulating dynamical features of escape panic. *Nature,* Volume 407, pp. 487-490.

Helbing, D., Farkas, I. & Vicsek, T., 2002. Crowd Disasters and Simulation of Panic Situations. In: *The science of disasters.* s.l.:s.n., pp. 330-350.

Helbing, D. & Johansson, A., 2010. Pedestrian, Crowd and Evacuation Dynamics. In: *Extreme Environmental Events.* s.l.:s.n., pp. 697-716.

Helbing, D., Johansson, A. & Al-Abideen, H. Z., 2007. Dynamics of crowd disasters: an empirical study. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics,* 75(4).

Helbing, D. & Molnár, P., 1995. Social force model for pedestrian dynamics. *Phys. Rev. E,* Volume 51.

Helbing, D. & Mukerji, P., 2012. Crowd disasters as systemic failures: analysis of the Love Parade disaster. *EPJ Data Science,* 1(7).

He, Z. et al., 2024. Multilevel Interpretability Of Artificial Neural Networks: Leveraging Framework And Methods From Neuroscience. *arXiv:2408.12664.*

Hochreiter, S. & Schmidhuber, J., 1997. LONG SHORT-TERM MEMORY. *Neural Compution,* 9(8), pp. 1735-1780.

Ho, J. & Ermon, S., 2016. *Generative adversarial imitation learning.* Montreal, MIT.

Holder, C., Middlehurst, M. & Bagnall, A., 2024. A review and evaluation of elastic distance functions for time series clustering. *Knowledge and Information Systems,* Volume 66, p. 765–809.

Hoogendoorn, S. & Bovy, P., 2000. Gas-Kinetic Modeling and Simulation of Pedestrian Flows.. *Transportation Research Record,* 1710(1), pp. 28-36.

Hoogendoorn, S. P. & Daamen, W., 2005. Pedestrian Behavior at Bottlenecks. *Transportation Science,* 39(2), pp. 147-159.

Hoogendoorn, S. P. & Daamen, W., 2005. Pedestrian Behavior at Bottlenecks. *Transportation Science,* 39(2).

Hsieh, Y.-H., Ngai, K. M., Jr, F. M. B. & Hsu, E. B., 2009. Epidemiological Characteristics of Human Stampedes. *Disaster Medicine and Public Health Preparedness,* 3(4), pp. 217-223.

Huang, L. et al., 2021. STI-GAN: Multimodal Pedestrian Trajectory Prediction Using Spatiotemporal Interactions and a Generative Adversarial Network. *IEEE Access,* Volume 9, pp. 50846-50856.

Huang, Z., 1998. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values.. *Data Mining and Knowledge Discovery,* Volume 2, pp. 283-304.

Husman, M. A. et al., 2021. Unmanned Aerial Vehicles for Crowd Monitoring and Analysis. *Electronics,* 10(23), p. 2974.

Imperial College London, 2025. *Virtual Reality Student Experience.* [Online]
Available at: https://www.imperial.ac.uk/virse/
[Accessed 27 Jan 2025].

International Organization for Standardization, 2020. *ISO 20414:2020 — Fire safety engineering — Verification and validation protocol for building fire evacuation models.,* Geneva: ISO.

Irfan, M., Marcenaro, L. & Tokarchuk, L., 2016. *Crowd analysis using visual and non-visual sensors, a survey.* Washington, DC, USA, IEEE Global Conference on Signal and Information Processing (GlobalSIP).

Jain, A. K., Murty, M. N. & Flynn, P. J., 1999. Data clustering: a review. *ACM Computing Surveys,* 31(3), pp. 264-323.

Jain, K., Prawesh, S., Dalal, I. S. & Tripathi, A., 2025. Benchmarking Pedestrian Dynamics Models for Common Scenarios: An Evaluation of Force-Based Models. *arXiv:2501.05106* .

Jia, X. et al., 2024. Towards Diverse Behaviors: A Benchmark for Imitation Learning with Human Demonstrations. *arXiv:2402.14606.*

Jing, F. et al., 2022. A Novel Named Entity Recognition Algorithm for Hot Strip Rolling Based on BERT-Imseq2seq-CRF Model. *Appl. Sci.,* 12(22).

Johansson, A., Helbing, D., Al-Abideen, H. Z. & Al-Bosta, S., 2008. From crowd dynamics to crowd safety: A video-based analysis.. *Advances in Complex Systems,* 11(4), pp. 497-527.

Joseph, V. R., 2022. Optimal ratio for data splitting. *Statistical Analysis and Data Mining,* 15(4), pp. 531-538.

Juřík, V. et al., 2023. Analysis of the use of behavioral data from virtual reality for calibration of agent-based evacuation models. *Heliyon,* 9(3).

K.Haavik, T., 2020. Societal resilience – Clarifying the concept and upscaling the scope. *Safety Science,* Volume 132.

Kahneman, D. & Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica,* 47(2), pp. 263-292.

Kaiser, M. S. et al., 2018. Advances in Crowd Analysis for Urban Applications Through Urban Event Detection. *IEEE Transactions on Intelligent Transportation Systems,* 19(10), pp. 3092-3112.

Kalatian, A., Sobhani, A. & Farooq, B., 2019. Analysis of distracted pedestrians' waiting time: Head-Mounted Immersive Virtual Reality application. *arXiv:1903.11812* .

Kelley, H. H., Jr., J. C. C., Dahlke, A. E. & Hill, A. H., 1965. Collective behavior in a simulated panic situation. *Experimental social psychology,* 1(1), pp. 20-54.

Kesici, N. & Erkan, N. Ç., 2023. The effect of public facade characteristics on changing pedestrian behaviors. *Journal of Architecture and Urbanism,* 47(1), pp. 68-76.

Khan, K. et al., 2020. Advances and Trends in Real Time Visual Crowd Analysis. *Sensors,* 20(18).

Kim, J., Tak, S., Bierlaire, M. & Yeo, H., 2020. Trajectory Data Analysis on the Spatial and Temporal Influence of Pedestrian Flow on Path Planning Decision. *Sustainability,* 12(24).

Kinateder, M., Wirth, T. D. & Warren, W. H., 2018. Crowd dynamics in virtual reality. In: *Crowd Dynamics, Volume 1: Theory, Models, and Safety Problems..* s.l.:Springer, pp. 15-36.

Kirik, E., Yurgel'yan, T. & Krouglov, D., 2007. *An intelligent floor field cellular automation model for pedestrian dynamics.* San Diego, CA, USA, SCSC '07: Proceedings of the 2007 Summer Computer Simulation Conference .

Kitzinger, J., 2009. *The media and public risk,* Cardiff: Risk & Regulation Advisory Council.

Kokhlikyan, N. et al., 2020. Captum: A unified and generic model interpretability library for PyTorch. *arXiv eprint,* Volume 2009.07896.

Kolivand, H. et al., 2021. An integration of enhanced social force and crowd control models for high-density crowd simulation. *Neural Comput & Applic,* Volume 33, p. 6095–6117.

Kretzschmar, H., Spies, M., Sprunk, C. & Burgard, W., 2016. Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research,* 35(11).

Kretz, T., Grünebohm, A. & Schreckenberg, M., 2006. Experimental study of pedestrian flow through a bottleneck. *Journal of Statistical Mechanics: Theory and Experiment,* Volume 2006.

Kretz, T., Lohmiller, J. & Sukennik, P., 2018. Some Indications on How to Calibrate the Social Force Model of Pedestrian Dynamics. *Transportation Research Record: Journal of the Transportation Research Board,* 2672(20).

Krumpal, I., 2011. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity ,* Volume 47, pp. 2025-2047.

Kruskal, W. H. & Wallis., W. A., 1952. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association ,* pp. 583-621.

Kugunavar, S. & Prabhakar, C. J., 2021. Convolutional neural networks for the diagnosis and prognosis of the coronavirus disease pandemic. *Vis Comput Ind Biomed Art,* 4(1).

Kui, Q., Li, J. & Zhang, B., 2009. *A Quantitative Model to Describe the Crowd Massing Risk Based on FIST Model.* Wuhan, IEEE.

Kuligowski, E., 2013. Predicting Human Behavior During Fires. *Fire Technology,* Volume 49, pp. 101-120.

Labhiri, A., Vaiciulyte, S., Kuligowski, E. & Ronchi, E., 2024. Evacuation decisions of tourists in wildfire scenarios. *International Journal of Disaster Risk Reduction,* Volume 113, p. 104836.

Lamba, S. & Nain, N., 2017. Crowd Monitoring and Classification: A Survey. In: S. Bhatia, K. Mishra, S. Tiwari & V. Singh, eds. *Advances in Computer and Computational Sciences.* Singapore: Springer.

Lamba, S. & Nain, N., 2017. *Crowd Monitoring and Classification: A Survey..* Singapore, Bhatia, S., Mishra, K., Tiwari, S., Singh, V. (eds) Advances in Computer and Computational Sciences. Advances in Intelligent Systems and Computing, vol 553. Springer.

Lamba, S. & Nain, N., 2020. Segmentation of crowd flow by trajectory clustering in active contours. *Vis Comput ,* Volume 36, pp. 989-1000.

Landis, J. R. & Koch., G. G., 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics,* 33(1), pp. 159-174.

Larsson, A. et al., 2021. The impact of crowd composition on egress performance. *Fire Safety Journal,* Volume 120, p. 103040.

Lawson, G. et al., 2019. Immersive virtual worlds: Multi-sensory virtual environments for health and safety training. *arXiv:1910.04697.*

LeCun, Y., Bengio, Y. & Hinton, G., 2015. Deep learning. *Nature,* Volume 521, pp. 436-444.

Lempitsky, V. & Zisserman, A., 2010. *Learning To Count Objects in Images.* Red Hook, NY, USA, Curran Associates Inc..

Lewin, K., 1951. *Field theory in social science: selected theoretical papers.* (Edited by Dorwin Cartwright.) ed. s.l.:Harpers..

Lighthill, M. J. & Whitham, G. B., 1955. On kinematic waves II: A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences,* 229(1178), pp. 317-345.

Li, J. et al., 2013. *Designing for crowd well-being: Current designs, strategies and future design suggestions.* Tokyo, Japan, 5th IASDR conference: Consilience and innovation in design.

Li, L., Liu, H. & Han, Y., 2019. Arch formation-based congestion alleviation for crowd evacuation. *Transportation Research Part C: Emerging Technologies,* Volume 100, pp. 88-106.

Li, M., Gao, Q. & Yu, T., 2023. Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters. *BMC Cancer,* Volume 23.

Lin, W. et al., 2023. HiEve: A Large-Scale Benchmark for Human-Centric Video Analysis in Complex Events. *International Journal of Computer Vision ,* Volume 131, pp. 2994-3018.

Li, R. et al., 2025. Virtual experimental study on pedestrian evacuation dynamics in view-limited environment. *Travel Behaviour and Society,* Volume 39.

Liu, M. et al., 2022. Balanced background and explanation data are needed in explaining deep learning models with SHAP: An empirical study on clinical decision making. *arXiv:2206.04050.*

Liu, W. et al., 2017. *Characterizing the relationship between environment layout and crowd movement using machine learning.* New York, 10th International Conference on Motion in Games .

Li, X., Liang, Y., Yang, Z. & Li, J., 2025. Pedestrian Trajectory Prediction Based on Dual Social Graph Attention Network. *Appl. Sci.,* 15(8).

Li, X., Wu, J. & Li, X., 2018. *Theory of practical cellular automaton.* Berlin/Heidelberg, Germany: Springer.

Li, Y. et al., 2019. A review of cellular automata models for crowd evacuation. *Physica A: Statistical Mechanics and its Applications,* Volume 526.

Li, Y., Chen, Y., Liu, J. & Huang, T., 2025. Efficient crowd simulation in complex environment using deep reinforcement learning. *Scientific Reports,* Volume 15, p. 5403.

Li, Z., 2024. GeoShapley: A Game Theory Approach to Measuring Spatial Effects in Machine Learning Models. *Annals of the American Association of Geographers,* 114(7), pp. 1365-1385.

Lovreglio, R., Ronchi, E. & Borri, D., 2014. The validation of evacuation simulation models through the analysis of behavioural uncertainty. *Reliability Engineering & System Safety,* Volume 131, pp. 166-174.

Lovreglio, R., Ronchi, E. & Nilsson, D., 2015. A model of the decision-making process during pre-evacuation. *Fire Safety,* Volume 78, pp. 168-179.

Lovreglio, R., Ronchi, E. & Nilsson, D., 2016. An Evacuation Decision Model based on perceived risk, social influence and behavioural uncertainty. *Simulation Modelling Practice and Theory,* Volume 66.

Lundberg, S. M. & Lee, S.-I., 2017. *A unified approach to interpreting model predictions.* Red Hook, NY, USA, the 31st International Conference on Neural Information Processing Systems (NIPS'17).

López-Marín, L. M., Rivera, A. L., Fernández, F. & Loske, A. M., 2018. Shock wave-induced permeabilization of mammalian cells. *Physics of Life Reviews,* Volume 26-27, pp. 1-38.

Lu, Y., Qiu, L., Lyu, X. & Jiang, X., 2021. Human stampede causative factors and cluster risk: A multi-dimensional analysis based on ISODATA and Fuzzy Theory. *International Journal of Disaster Risk Reduction,* Volume 66.

Lynch, K., 1960. *The image of the city.* Cambridge: MIT Press.

Müseler, A., 2010. [Online]
Available at: https://arne-mueseler.com/
[Accessed 12 July 2025].

Macal, C. & North, M., 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation ,* Volume 4, pp. 151-162.

MacQueen, J., 1967. *Some Methods for Classification and Analysis of Multivariate Observations.* Berkeley, Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, University of California Press.

Manh, H. & Alaghband, G., 2019. Scene-LSTM: A Model for Human Trajectory Prediction. *arXiv:1808.04018.*

Mann, H. & Whitney, D., 1947. On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. *Annals of Mathematical Statistics,* pp. 50-60.

Marciani, S. et al., 2026. The role of leaderboards for gamification in virtual reality simulation mastoidectomy training: a controlled trial. *Eur Arch Otorhinolaryngol* .

Martella, C., Li, J., Conrado, C. & Vermeeren, A., 2017. On current crowd management practices and the need for increased situation awareness, prediction, and intervention. *Safety Science,* Volume 91, pp. 381-393.

Ma, T., Wang, Q. & Larrañaga, M. D., 2011. *THE UTILITY OF A PANIC MODEL ON SIMULATING CROWD DISASTERS.* Baltimore, s.n.

Mayr, U. et al., 2011. Competitiveness across the Life Span: The Feisty Fifties. *Psychol Aging,* 27(2), pp. 278-285.

McFadden, R. D. & Macropoulos, A., 2008. *Wal-Mart Employee Trampled to Death.* [Online]
Available at: https://www.nytimes.com/2008/11/29/business/29walmart.html
[Accessed 28 May 2025].

McHugh, M. L., 2012. Interrater reliability: the kappa statistic. *Biochemia Medica,* 22(3), pp. 276-282.

Mekler, E. D., Brühlmann, F., Opwis, K. & Tuch, A. N., 2013. *Do points, levels and leaderboards harm intrinsic motivation?: an empirical analysis of common gamification elements.* New York, USA, First International Conference on Gameful Design, Research, and Applications.

Melnikov, V. R. et al., 2022. Behavioural thermal regulation explains pedestrian path choices in hot urban environments. *Scientific Reports,* Volume 12.

Mondal, A. et al., 2021. How do people feel while walking? A multivariate analysis of emotional well-being for utilitarian and recreational walking episodes. *International Journal of Sustainable Transportation ,* 15(6).

Montecchiari, G., Gallina, P. & Bulian, G., 2017. *An experiment using immersive virtual reality and a haptic interface to study human behaviour in evacuation.* London, Safety and Reliability – Theory and Applications: Proceedings of ESREL 2017.

Mori, M. & Tsukaguchi, H., 1987. A new method for evaluation of level of service in pedestrian facilities. *Transport Research A,* 21(3).

Moussaïd, M. et al., 2009. Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proc Biol Sci ,* 276(1668), pp. 2755-2762.

Moussaïd, M., Helbing, D. & Theraulaz, G., 2011. How simple rules determine pedestrian behavior and crowd disasters. *PNAS,* 108(17), pp. 6884-6888.

Moussaïd, M. et al., 2016. Crowd behaviour during high-stress evacuations in an immersive virtual environment. *J. R. Soc. Interface.,* Volume 13.

Murtagh, E. M. et al., 2021. Outdoor Walking Speeds of Apparently Healthy Adults: A Systematic Review and Meta-analysis. *Sports Med,* 51(1), pp. 125-141.

Nanay, B., 2018. Multimodal mental imagery. *Cortex,* Volume 105, pp. 125-134.

Narang, S., Best, A., Curtis, S. & Manocha, D., 2015. Generating Pedestrian Trajectories Consistent with the Fundamental Diagram Based on Physiological and Psychological Factors. *PLOS ONE,* 10(4).

Nasernejad, P., Sayed, T. & Alsaleh, R., 2021. Modeling pedestrian behavior in pedestrian-vehicle near misses: A continuous Gaussian Process Inverse Reinforcement Learning (GP-IRL) approach. *Accident Analysis & Prevention,* Volume 161.

Nasernejad, P., Sayed, T. & Alsaleh, R., 2023. Multiagent modeling of pedestrian-vehicle conflicts using Adversarial Inverse Reinforcement Learning. *Transportmetrica A: Transport Science,* 19(3).

Newman, M. E. J., 2011. Complex Systems: A Survey. *arXiv:1112.1440* .

Ngai, K., Burkle, F., Hsu, A. & Hsu, E., 2009. Human Stampedes: A Systematic Review of Historical and Peer-Reviewed Sources. *Disaster Medicine and Public Health Preparedness,* pp. 191-195.

Ng, A. Y. & Russell, S. J., 2000. *Algorithms for Inverse Reinforcement Learning.* San Francisco, Morgan Kaufmann Publishers.

Ng, M., Li, M. J., Huang, J. & He, Z., 2007. On the Impact of Dissimilarity Measure in k-Modes Clustering Algorithm.. *IEEE transactions on pattern analysis and machine intelligence.,* Volume 29, pp. 503-507.

Nguyen, L. N. et al., 2024. Non-contact multimodal indoor human monitoring systems: A survey. *Information Fusion,* Volume 110.

Niennattrakul, V. & Ratanamahatana, C., 2007. *On Clustering Multimedia Time Series Data Using K-Means and Dynamic Time Warping.* Seoul, South Korea, International Conference on Multimedia and Ubiquitous Engineering .

Nikmanesh, M., Cinelli, M. E. & Marigold, D. S., 2025. Identifying factors that contribute to collision avoidance behaviours while walking in a natural environment. *Scientific Reports,* Volume 15.

Nong, Y., Sun, J.-Q. & Wang, K., 2025. AI-Driven Procedural Content Generation for VR/AR Environments. *Current Research in Next Generation Materials Engineering,* 1(2).

Olivier, A.-H., Bruneau, J., Kulpa, R. & Pettré, J., 2018. Walking with Virtual People: Evaluation of Locomotion Interfaces in Dynamic Environments. *Transactions on Visualization and Computer Graphics,* 24(7), pp. 2251-2263.

Olsen, O. E., Kruke, B. I. & Hovden, J., 2007. Societal safety: concept, borders and dilemmas. *J. Contingencies Crisis Manage.,* 15(2), pp. 69-79.

Osa, T. et al., 2018. An Algorithmic Perspective on Imitation Learning. *arXiv:1811.06711.*

Ourdas, C. & Ponis, S., 2023. Evaluating the Effects of Gamification in Behavioural Change: A Proposed SEM-Based Approach. *Sustainability ,* 15(6), p. 5442.

Papadimitriou, E., Lassarre, S. & Yannis, G., 2016. Pedestrian risk taking while road crossing: a comparison of. *Transportation Research Procedia,* Volume 14, pp. 4354-4363.

Papadimitriou, E., Lassarre, S. & Yannis, G., 2017. Human factors of pedestrian walking and crossing behaviour. *Transportation Research Procedia,* Volume 25, pp. 2002-2015.

Pasmore, W., Winby, S., Mohrman, S. A. & Vanasse, R., 2019. Reflections: Sociotechnical Systems Design and Organization Change. *Journal of Change Management ,* 19(2).

Patel, S., Sihmar, S. & Jatain, A., 2015. *A study of hierarchical clustering algorithms.* New Delhi, India, 2nd International Conference on Computing for Sustainable Global Development (INDIACom).

Pelechano, N. & Allbeck, J. M., 2016. *Feeling crowded yet? Crowd simulations for VR.* Greenville, SC, USA, 2016 IEEE Virtual Humans and Crowds for Immersive Environments (VHCIE).

Pellegrini, S., Ess, A., Schindler, K. & Gool, L. v., 2009. *You'll never walk alone: Modeling social behavior for multi-target tracking.* Kyoto, Japan, 2009 IEEE 12th International Conference on Computer Vision,.

Peng, L., Jian, M. & Siuming., L., 2016. Discrete element crowd model for pedestrian evacuation through an exit. *Chinese Physics B,* 25(3).

Pietro, C., Chicchi, G. I. A., Alcañiz, R. M. & Giuseppe, R., 2018. The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature. *Frontiers in Psychology,* Volume 9.

Pitafi, S., Anwar, T. & Sharif, Z., 2023. A Taxonomy of Machine Learning Clustering Algorithms, Challenges, and Future Realms. *Applied Sciences,* 13(6).

Polus, A., Schofer, J. L. & A. Ushpiz, 1983. *Journal of Trans-portation Engineering,* Volume 109, p. 46.

Pretorius, M., Gwynne, S. & Galea, E. R., 2013. Large crowd modelling: an analysis of the Duisburg Love Parade disaster. *Fire and materials,* 39(4), pp. 301-322.

Qiao, G. et al., 2022. An Information-Theoretic Approach for Estimating Scenario Generalization in Crowd Motion Prediction. *arXiv:2211.00817* .

Qiao, G. et al., 2019. Scenario Generalization of Data-driven Imitation Models in Crowd Simulation. *arXiv:1910.00738.*

Quarantelli, E., 1957. Panic Behavior: Some Empirical Observations. *Sociology and Social Research,* Volume 41, pp. 187-194.

Radha, K. & Bansal, M., 2023. Feature Fusion and Ablation Analysis in Gender Identification of Preschool Children from Spontaneous Speech. *Circuits Syst Signal Process ,* Volume 42, pp. 6228-6252.

Reicher, S. D., 1996. "The Battle of Westminster": Developing the social identity model of crowd behaviour in order to explain the initiation and development of collective conflict.. *European Journal of Social Psychology,* 26(1), pp. 115-134.

Ren, X., Zhang, J., Song, W. & Cao, S., 2019. The fundamental diagrams of elderly pedestrian flow in straight corridors under different densities. *Journal of Statistical Mechanics: Theory and Experiment,,* Volume 2019.

Reuters, 2023. *South Korean police blame fatal Halloween crush on lack of preparation, poor response.* [Online]
Available at: https://www.reuters.com/world/asia-pacific/south-korean-police-blame-fatal-halloween-crush-lack-preparation-poor-response-2023-01-13/?utm_source=chatgpt.com
[Accessed 12 July 2025].

Rodriguez, M., Laptev, I., Sivic, J. & Audibert, J. Y., 2011. *Density-aware person detection and tracking in crowds.* Piscataway, NJ, 2011 International Conference on Computer Vision.

Rodriguez, M. Z. et al., 2019. Clustering algorithms: A comparative approach. *PloS one,* 14(1).

Roggen, D., Wirz, M., Tröster, G. & Helbing, D., 2011. Recognition of crowd behavior from mobile sensors with pattern analysis and graph clustering methods. *Networks and Heterogeneous Media,* 6(3), pp. 521-544.

Ronchi, E., 2021. Developing and validating evacuation models for fire safety engineering. *Fire Safety Journal,* Volume 120.

Rong, L. & Li, C., 2020. A Strong Baseline for Crowd Counting and Unsupervised People Localization. *arXiv:2011.03725.*

Ropohl, G., 1999. PHILOSOPHY OF SOCIO-TECHNICAL SYSTEMS. *Society for Philosophy and Technology Quarterly Electronic Journal,* 4(3).

Rousseeuw, P. J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics,* Volume 20, pp. 53-65.

Rudenko, A. et al., 2022. The Atlas Benchmark: an Automated Evaluation Framework for Human Motion Prediction. *arXiv:2207.09830.*

Saarloos, D., Kim, J.-E. & Timmermans, H., 2009. The Built Environment and Health: Introducing Individual Space-Time Behavior. *Int J Environ Res Public Health.,* 6(6), pp. 1724-1743.

Sadeghian, A. et al., 2018. SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints. *arXiv:1806.01482.*

Saleh, K., Hossny, M. & Nahavandi, S., 2018. *Long-Term Recurrent Predictive Model for Intent Prediction of Pedestrians via Inverse Reinforcement Learning.* Canberra, IEEE.

Salzmann, T., Ivanovic, B., Chakravarty, P. & Pavone, M., 2021. Trajectron++: Dynamically-Feasible Trajectory Forecasting With Heterogeneous Data. *arXiv:2001.03093.*

Saunders, M., Lewis, P. & Thornhill, A., 2019. *Research Methods for Business Students.* 8th Edition ed. s.l.:Pearson Education Limited. .

Sazlı, M. H., 2006. A brief review of feed-forward neural networks. *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering,* 50(01).

Schadschneider, A. et al., 2009. *Evacuation Dynamics: Empirical Results, Modeling and Applications..* s.l., In R. A. Meyers (Ed.), Encyclopedia of Complexity and Systems Science, Springer..

Schwandt, T. A., 1994. Constructivist, interpretivist approaches to human inquiry. In: N. K. Denzin & Y. S. Lincoln (Eds.), ed. *Handbook of qualitative research.* s.l.:Sage Publications, Inc., pp. 118-137.

Sedaghati, N., Ardebili, S. & Ghaffari, A., 2025. Application of human activity/action recognition: a review. *Multimed Tools Appl* .

SGSA, 2018. *Guide to Safety at Sports Grounds 'Green Guide'.* [Online]
Available at: https://sgsa.org.uk/wp-content/uploads/2019/02/Worked-Example-A-Football-and-Rugby.pdf
[Accessed 5 October 2023].

Shao, C.-H., Shao, P.-C. & Kuo, F.-M., 2019. Stampede Events and Strategies for Crowd Management. *Disaster Research,* 14(7), pp. 949-958.

Shao, S. et al., 2018. CrowdHuman: A Benchmark for Detecting Human in a Crowd. *arXiv:1805.00123.*

Shapley, L., 1953. A Value for n-Person Games. . In: H. a. T. A. Kuhn, ed. *Contributions to the Theory of Games II.* Princeton: Princeton University Press, pp. 307-317.

Sharbini, H. & Bade, A., 2009. *Analysis of Crowd Behaviour Theories in Panic Situation.* Jeju, Korea, Information and Multimedia Technology.

Sharma, D., Bhondekar, A. P. & Ghanshyam, A. K. S. &. C., 2018. A review on technological advancements in crowd management. *Journal of Ambient Intelligence and Humanized Computing,* Volume 9, pp. 485-495.

Shepherd, R. & Clegg, C., 2011. Crowd disasters: A socio-technical systems perspective.. *Contemporary Social Science.,* Volume 6, pp. 343-360.

Sheu, Y.-h., 2020. Illuminating the Black Box: Interpreting Deep Neural Network Models for Psychiatric Research. *Frontiers in Psychiatry ,* Volume 11.

Shipman, A. & Majumdar, A., 2018. Fear in Humans: A Glimpse into the Crowd-Modeling Perspective. *Transportation Research Record: Journal of the Transportation Research Board,* 2672(1).

Shipman, A. et al., 2022. *Novel trajectory prediction algorithm using a full dataset: comparison and ablation studies.* Macau, China, IEEE 25th International Conference on Intelligent Transportation Systems (ITSC).

Sieben, A., Schumann, J. & Seyfried, A., 2017. Collective phenomena in crowds—Where pedestrian dynamics need social psychology. *PLoS ONE ,* 12(6).

Simon, H. A., 1956. Rational choice and the structure of the environment. *Psychological review,* 63(2), pp. 129-38.

Sindagi, V. A. & Patel, V. M., 2017. *CNN-Based cascaded multi-task learning of high-level prior and density estimation for crowd counting.* Lecce, Italy, 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS).

Si, T. et al., 2024. Multivariate Time Series Change-Point Detection with a Novel Pearson-like Scaled Bregman Divergence. *Stats,* 7(2), pp. 462-480.

Soares, F. et al., 2021. To cross or not to cross: Impact of visual and auditory cues on pedestrians' crossing decision-making. *Transportation Research Part F: Traffic Psychology and Behaviour,* Volume 82, pp. 202-220.

Song, X. et al., 2021. Pedestrian Trajectory Prediction Based on Deep Convolutional LSTM Network. *IEEE Transactions on Intelligent Transportation Systems,* 22(6), pp. 3285-3302.

Sotnik, G., Fischer, A. P., Ibáñez, I. & Cousins, S. J. M., 2021. A transdisciplinary typology of change identifies new categories of adaptations and forms of co-adaptation in coupled human and natural systems. *Sustain Sci,* Volume 16, pp. 1609-1623.

Sports Grounds Safety Authority, 2018. *Guide to Safety at Sports Grounds,* London: Sports Grounds Safety Authority.

Srouji, M., Zhang, J. & Salakhutdinov, R., 2018. Structured Control Nets for Deep Reinforcement Learning. *arXiv:1802.08311* .

Still, G. K., 2000. *Crowd Dynamics.* [Online]
Available at: https://www.gkstill.com/CV/PhD/Chapter1.html
[Accessed 8 Feburary 2022].

Still, G. K., 2021. RAMP analysis. In: *Applied Crowd Science.* s.l.:CRC Press.

Still, G. K., 2025. *Crowd Safety and Crowd Risk Analysis.* [Online]
Available at: https://www.gkstill.com/Support/WhyModel/dimice.html
[Accessed 17 June 2025].

Still, G., Papalexi, M., Fan, Y. & Bamford, D., 2020. Place crowd safety, crowd science? Case studies and application. *Journal of Place Management and Development,* 13(3), pp. 261-278.

Still, K., Papalexi, M., Fan, Y. & Bamford, D., 2020. Place crowd safety, crowd science? Case studies and application. *Place Management and Development,* 13(4).

Still, K., Papalexi, M., Fan, Y. & Bamford, D., 2020. Place crowd safety, crowd science? Case studies and application. *Journal of Place Management and Development,* 13(4).

Stott, C. & Radburn, M., 2020. Understanding crowd conflict: social context, psychology and policing. *Current Opinion in Psychology,* Volume 35, pp. 76-80.

Strumbelj, E. & Kononenko, I., 2010. An Efficient Explanation of Individual Classifications using Game Theory. *Journal of Machine Learning Research,* Volume 11, pp. 1-18.

Sundararajan, M., Taly, A. & Yan, Q., 2017. Axiomatic Attribution for Deep Networks. *arXiv:1703.01365* .

Sutton, R. S. & Barto, A. G., 2018. *Reinforcement Learning: An Introduction.* 2nd Edition ed. Cambridge, MA: MIT Press.

T, G. & HR, M., 2024. How to Improve Video Analytics with Action Recognition: A Survey. *ACM Computing Surveys,* 57(1), pp. 1-36.

Thida, M. et al., 2013. A Literature Review on Video Analytics of Crowded Scenes. In: P. Atrey, M. Kankanhalli & A. Cavallaro, eds. *Intelligent Multimedia Surveillance.* Heidelberg: Springer.

Titus, M., Hagstrom, G. & Watson, J. R., 2021. Unsupervised manifold learning of collective behavior. *PLoS Comput Biol,* 17(2).

Tomar, A., Kumar, S. & Pant, B., 2022. *Crowd Analysis in Video Surveillance: A Review.* Chiangrai, Thailand, International Conference on Decision Aid Sciences and Applications (DASA),.

Tong, Y. & Bode, N. W. F., 2022. The principles of pedestrian route choice. *Journal of the Royal Society Interface,,* 19(189).

Trist, E., 1981. *The Evolution of Socio-Thecnical Systems.,* Toronto: Ontario Ministry of Labour, Ontario Quality of Working Life Centre..

Trist, E. L. & Bamforth, K. W., 1951. S Some Social and Psychological Consequences of the Longwall Method of Coal-Getting: An Examination of the Psychological Situation and Defences of a Work Group in Relation to the Social Structure and Technological Content of the Work System. *Human Relations,* 4(1), pp. 3-38.

Truong, C., Oudre, L. & Vayatis, N., 2020. Selective review of offline change point detection methods. *Signal Processing,* Volume 167.

Tudor-Locke, C. et al., 2018. How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. *British Journal of Sports Medicine,* Volume 52, pp. 776-788.

Ullah, H. et al., 2019. *Two Stream Model for Crowd Video Classification.* Rome, Italy, 8th European Workshop on Visual Information Processing (EUVIP).

Ultralytics, 2020. *YOLOv5.* [Online]
Available at: https://github.com/ultralytics/yolov5
[Accessed 19 June 2025].

Unity Technologies, 2025. *Unity Game Engine.* [Online]
Available at: https://unity.com
[Accessed 27 Jan 2025].

Vahdani, E. & Tian, Y., 2023. Deep Learning-Based Action Detection in Untrimmed Videos: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 45(4), pp. 4302-4320.

Valmorisco, S., Raya, L. & Sanchez, A., 2024. Enabling personalized VR experiences: a framework for real-time adaptation and recommendations in VR environments. *Virtual Reality,* 28(128).

Vanumu, L. D., Rao, K. R. & Tiwari, G., 2017. Fundamental diagrams of pedestrian flow characteristics: A review. *European Transport Research Review,* Volume 9.

Venkatesan, R. K., Banakou, D., Slater, M. & Steed, A., 2023. Haptic feedback in a virtual crowd scenario improves the emotional response. *Frontiers in Virtual Reality,* Volume 4.

Verschaffel, L., Greer, B. & De Corte, E., 2002. Everyday Knowledge and Mathematical Modeling of School Word Problems.. In: K. L. R. O. B. v. a. V. L. Gravemeijer, ed. *Symbolizing, Modeling and Tool Use in Mathematics Education.* Netherlands: Kluwer Academic, pp. 171-195.

Viola, P. & Jones, M., 2004. Robust Real-Time Face Detection. *International Journal of Computer Vision,* Volume 57, pp. 137-154.

Viterale, F. D. C., 2025. An Examination of the Complexity of the Large Sociotechnical Space System. *Aerospace,* 12(6).

Vizzari, G., 2020. *Virtual Reality to Study Pedestrian Wayfinding: Motivations and an Experiment on Usability.* Utrecht, Netherlands, IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR).

Vizzari, G. & Cecconello, T., 2023. Pedestrian Simulation with Reinforcement Learning: A Curriculum-Based Approach. *Future Internet,* 15(1).

Von Neumann, J., 1966. *Theory of Self-Reproducing Automata.* Urbana and London: University of Illinois Press..

Wang, C. et al., 2025. Immersive virtual reality experiments for emergency evacuation response in deep underground space. *Tunnelling and Underground Space Technology,* Volume 163.

Wang, J. et al., 2019. Propagation characteristics of the pedestrian shockwave in dense crowd: Experiment and simulation. *International Journal of Disaster Risk Reduction,* Volume 40.

Wang, J., Lv, W., Cao, S. & Guo, C., 2024. CosForce: A Force-Based General Pedestrian Model with Anticipation and Reaction Mechanisms. *arXiv:2410.10746.*

Wang, M., Zhou, X. & Chen, Y., 2025. A comprehensive survey of crowd density estimation and counting. *IET Image Process,* Volume 19, p. e13328.

Wang, P. et al., 2021. Pedestrian interaction with automated vehicles at uncontrolled intersections. *Transportation Research Part F: Traffic Psychology and Behaviour,* pp. 10-25.

Wang, Y. & Li, X., 2021. Reward function shape exploration in adversarial imitation learning: an empirical study. *arXiv:2104.06687.*

Wang, Y., Srinivasan, A. R., Lee, Y. M. & Markkula, G., 2024. Modeling Pedestrian Crossing Behavior: A Reinforcement Learning Approach with Sensory Motor Constraints. *arXiv:2409.14522.*

Wang, Y., Zhang, T., Guo, X. & Shen, Z., 2024. Gradient based Feature Attribution in Explainable AI: A Technical Review. *arXiv:2403.10415* .

Warren, J., 2024. *Brixton Academy: Crush victim's family taking legal action.* [Online]
Available at: https://www.bbc.co.uk/news/uk-england-london-68619308#:~:text=Gaby%2C%2023%2C%20and%20Rebecca%20Ikumelo,the%20night%20of%20the%20crush.
[Accessed 29 June 2025].

Wedagamaa, D. M. P., Bennettb, S. & Dissanayake, D., 2020. Analyzing Pedestrian Perceptions towards Traffic Safety Using Discrete Choice Models. *International Journal on Advanced Science Engineering Information Technology,* 10(6).

Weppner, J., Bischke, B. & Lukowicz, P., 2016. *Monitoring crowd condition in public spaces by tracking mobile consumer devices with wifi interface.* New York, NY, USA, Association for Computing Machinery.

WeWalk, 2025. *Introducing Brand New AI Integrated Smart Cane.* [Online]
Available at: https://wewalk.io/en/
[Accessed 12 July 2025].

Wolff, W., Bieleke, M., Martarelli, C. S. & Danckert, J., 2021. A Primer on the Role of Boredom in Self-Controlled Sports and Exercise Behavior. *Front Psychol. ,* Volume 12.

Wong, V. W. H. & Law, K. H., 2023. Fusion of CCTV Video and Spatial Information for Automated Crowd Congestion Monitoring in Public Urban Spaces. *Algorithms,* 16(3).

Working with crowd, 2023. *Crowd Disaster and Incidents.* [Online]
Available at: https://www.workingwithcrowds.com/crowd-disasters-and-incidents/
[Accessed 2 July 2023].

Wu, F., Ke, J. & Wu, A., 2023. Inverse Reinforcement Learning with the Average Reward Criterion. *arXiv:2305.14608.*

Xiao, L. et al., 2025. *Wide Field-of-View Mixed Reality.* New York, NY, USA, SIGGRAPH '25: ACM SIGGRAPH 2025 Emerging Technologies .

Xia, X., Chen, J., Zhang, J. & Li, N., 2024. How the strength of social relationship affects pedestrian evacuation behavior: A multi-participant fire evacuation experiment in a virtual metro station. *Transportation Research Part C: Emerging Technologies,* Volume 167.

Xing, S. et al., 2024. A fine discrete floor field cellular automaton model with natural step length for pedestrian dynamics. *Simulation Modelling Practice and Theory,* Volume 130.

Xinhuanet, 2015. *上海外滩跨年夜发生踩踏事故 已致 35 人遇难 43 人伤 (A stampede occurred on New Year's Eve at the Waitai in Shanghai, leaving 35 people dead and 43 injured.).* [Online]
Available at: https://www.mhwmm.com/zonghezixun/8664.html
[Accessed 12 July 2025].

Xinjiang Television, 2015. *Shanghai Waitan Stampede investigation report.* [Online]
Available at: http://politics.people.com.cn/n/2015/0121/c1001-26424342.html
[Accessed 26 October 2023].

Xue, H., Huynh, D. Q. & Reynolds, M., 2018. SS-LSTM: A Hierarchical LSTM Model for Pedestrian Trajectory Prediction. *2018 IEEE Winter Conference on Applications of Computer Vision,* pp. 1186-1194.

Xu, Z. et al., 2024. Analyzing scenario criticality and rider's intervention behavior during high-level autonomous driving: A VR-enabled approach and empirical insights. *Transportation Research Part C: Emerging Technologies,* Volume 158.

Yang, L. et al., 2025. Data-driven agent-based modeling to assess public space design quality at transport stations considering infectious disease outbreaks and emergencies. *Under review.*

Yang, M. et al., 2018. *Cluster-Based Crowd Movement Behavior Detection.* Canberra, ACT, Australia, 2018 Digital Image Computing: Techniques and Applications (DICTA).

Yang, X. et al., 2024. Characteristics of crowd disaster: Database construction and pattern identification. *International Journal of Disaster Risk Reduction,* Volume 111.

Yang, Y. & Sun, J., 2013. Study on Pedestrian Red-Time Crossing Behavior: Integrated Field Observation and Questionnaire Data. *Transportation Research Record: Journal of the Transportation Research Board,* 2393(1).

Yan, L., Zhu, M. & Yu, C., 2019. Crowd Video Captioning. *arXiv:1911.05449.*

Yap, M. et al., 2021. Verifying explainability of a deep learning tissue classifier trained on RNA-seq data. *Scientific Reports,* Volume 11, p. 2641.

Ye, R. et al., 2024. A study on the arch mechanism of pedestrian evacuation and congestion alleviation strategies at building exits. *Journal of Building Engineering,* Volume 88.

Yim, O. & Ramdeen, K. T., 2015. Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data. *The Quantitative Methods for Psychology ,* 11(1), pp. 8-21.

Yin, T. et al., 2022. The One-Man-Crowd: Single User Generation of Crowd Motions Using Virtual Reality. *Transactions on Visualization and Computer Graphics,* 28(5), pp. 2245-2255.

Yuan, Y., Fang, J. & Wang, Q., 2015. Online Anomaly Detection in Crowd Scenes via Structure Analysis. *IEEE Transactions on Cybernetics,* 45(3), pp. 548-561.

Yuen, J. & Lee, E., 2012. The effect of overtaking behavior on unidirectional pedestrian flow. *Safety Science,* 50(8), pp. 1704-1714.

Zhang, G., Yu, Z., Jin, D. & Li, Y., 2022. *Physics-infused Machine Learning for Crowd Simulation.* New York, 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

Zhang, H., Goodfellow, I., Metaxas, D. & Odena, A., 2018. Self-Attention Generative Adversarial Networks. *arXiv:1805.08318* .

Zhang, J., Klingsch, W., Schadschneider, A. & Seyfried, A., 2012. Ordering in bidirectional pedestrian flows and its influence on the fundamental diagram. *J. Stat. Mech.*.

Zhang, J., Klingsch, W., Schadschneider, A. & Seyfried, A., 2013. *Experimental study of pedestrian flow through a T-junction.* Berlin, Traffic and Granular Flow '11, Springer.

Zhang, L. et al., 2024. A Crowd Movement Analysis Method Based on Radar Particle Flow. *Sensors,* 24(6).

Zhang, M., Xu, R., Siu, M. F. F. & Luo, X., 2023. Human decision change in crowd evacuation: A virtual reality-based study. *Journal of Building Engineering,* 68(1).

Zhang, Z. et al., 2025. Adaptive exit choices of pedestrians during emergency evacuation: A study combining virtual experiments, survey and modelling. *Advanced Engineering Informatics,* 65(C).

Zhao, R. et al., 2024. A Review of Abnormal Crowd Behavior Recognition Technology Based on Computer Vision. *Applied Sciences,* 14(21).

Zhou, R., Cui, Y., Wang, Y. & Jiang, J., 2021. A modified social force model with different categories of pedestrians for subway station evacuation. *Tunnelling and Underground Space Technology,* Volume 110.

Zhou, S. et al., 2010. Crowd modeling and simulation technologies. *ACM Transactions on Modeling and Computer Simulation,* 20(4), pp. 1-35.

Zhou, Y. et al., 2020. Understanding Crowd Behaviors in a Social Event by Passive WiFi Sensing and Data Mining. *IEEE Internet of Things Journal,* 7(5), pp. 4442-4454.

# Appendix I

# 2014 Shanghai crowd disaster official report (translated)

At 23:35 on December 31, 2014, there was a crowded stampede on the stairs of the pedestrian passage leading to the Huangpu River Observation Platform at the southeast corner of Chenyi Square on the Bund, Huangpu District, Shanghai, resulting in 36 deaths and 49 injuries.

After the incident, General Secretary Xi Jinping, Premier Li Keqiang and other leading comrades from the central government issued important instructions respectively, requiring Shanghai to go all out to treat the wounded, do a good job in the aftermath, investigate the cause of the incident, learn lessons from the incident, and release it to the public in a timely and accurate manner. information.

Comrade Han Zheng, Secretary of the Municipal Party Committee, and Comrade Yang Xiong, the Mayor, attached great importance to the investigation of the incident, and demanded a thorough investigation of the cause, serious accountability, and learned lessons. In accordance with relevant laws and regulations such as the Emergency Response Law of the People's Republic of China and the Measures of Shanghai Municipality for Implementing the Emergency Response Law of the People's Republic of China, Shanghai quickly established a joint investigation team of the municipal government, a member of the Standing Committee of the Municipal Party Committee and an executive deputy municipality. Chief Tu Guangshao served as the team leader, Deputy Mayor Zhou Bo served as the deputy team leader, the relevant deputy secretary-general of the municipal government, the Municipal Supervision Bureau, the Municipal Safety Supervision Bureau, the Municipal Public Security Bureau Disciplinary Commission, the Municipal Emergency Office, the Municipal Government Legislative Affairs Office, the Municipal Health and Family Planning Commission, The relevant responsible comrades of the Municipal Tourism Bureau and other departments are

members. The joint investigation team invited national and municipal experts in emergency management, public safety management, and law to analyse and demonstrate the incident.

The joint investigation team carried out incident investigation through on-site investigation, investigation and evidence collection, expert demonstration, and comprehensive analysis. After the incident, the joint investigation team quickly organized forces to investigate the scene of the incident, and successively investigated and visited the injured, policemen on duty, citizens and tourists, and subway staff, to understand the incident and to check the relevant circumstances of the incident site. The joint investigation team also retrieved and viewed 36 surveillance probes in the Bund area with a cumulative duration of about 70 hours of video recordings, systematically sorted out relevant laws and regulations, and consulted the "three plans", normative documents, meeting minutes, etc. of relevant municipal and district-level departments. 195 documents were collected. A total of 51 people from 10 municipal departments (units), Huangpu District Government and leaders of relevant district departments were interviewed and inquired, and the relevant management situation was investigated in detail.

After investigation, the joint investigation team restored the whole process of the incident, identified the relevant response situation, analysed the cause of the incident, identified the nature of the incident, put forward handling suggestions for the relevant responsible persons, and put forward rectification suggestions on strengthening the work related to urban public safety.

1. The basic situation of the place where the incident occurred

(1) The Bund Scenic Area

The Bund Scenic Area is a public area within the jurisdiction of Huangpu District. It starts from the Huangpu River Flood Control Wall in the east, extends to the west sidewalk of Zhongshan East No. 1 Road and Zhongshan East No. 2 Road in the west, starts from the north sidewalk of Dongmen Road in the south, and reaches the south bank of Suzhou River in the north, with an area of 3.1 square kilometers.

(2) Chen Yi Square

Chenyi Plaza is located in the middle of the Bund Scenic Area (across the road from No. 335 to 309 of Zhongshan East Road), adjacent to the east end of Nanjing East Road, and connected to Zhongshan East Road, with a public activity area of about 2,877 square meters. The Huangpu River viewing platform connected by large stairs and large ramps in Chen Yi Square is the best viewing location in the Bund Scenic Area. In addition, the traffic near Chenyi Square is convenient. It is about 580 meters away from the Nanjing East Road Station of Metro Line 2 and Line 10. It is the area with the largest flow of people and the highest density of people in the Bund Scenic Area.

(3) Crowded and stampede scene

The scene of the incident is located at the upper and lower pedestrian passage steps leading to the Huangpu River viewing platform at the southeast corner of Chen Yi Square. The stairs are divided into two groups of 17 steps from top to bottom. The distance between the two groups of steps is 2.3 meters. There are stainless steel bar handrails on both sides of the steps. The width of the steps is 6.2 meters, the highest point is 3.5 meters above the ground, and the depth is 8.4 meters.

(4) Surroundings of the Bund Scenic Area

On the east side of the Bund Scenic Area, across the Huangpu River is the Lujiazui area of Pudong where landmark buildings such as Shanghai Oriental Pearl and the newly completed Shanghai Center are located. Lane, Fuzhou Road, Hankou Road, Jiujiang Road, Nanjing East Road, Dianchi Road, Beijing East Road, South Suzhou Road and other roads are connected. Citizens and tourists can walk up the stairs to the viewing platform to watch the landscape lights and buildings on both sides of the Huangpu River.

The Bund Source is located at No. 33 Zhongshan East Road, adjacent to the Bund Scenic Area, about 550 meters away from Chen Yi Square.

2. Changes and preparations for New Year's countdown activities

(1) Relevant information on the change of the New Year countdown event

Since 2011, Huangpu District Government, Shanghai Municipal Tourism Administration and Shanghai Radio and Television have held New Year countdown activities at the Bund Scenic Area for three consecutive years. In view of certain uncontrollable factors in safety and other aspects, the Huangpu District Government, after negotiating with the Shanghai Municipal Tourism Administration and Shanghai Radio and Television Station, asked the municipal government for instructions on November 13, 2014 that the New Year's countdown event was temporarily suspended at the Bund Scenic Area. Held in another location, the audience of the event will be limited to about 3,000 people. The organisers are the Huangpu District Government and Shanghai Radio and Television Station. In this regard, the municipal government agreed to suspend the New Year's countdown event held in the Bund Scenic Area, and clearly required that "whoever hosts the event will be responsible" for the event held in another location, resolutely implement territorial jurisdiction, and effectively implement the responsibility.

On December 9, 2014, the 76th executive meeting of the Huangpu District Government decided that the 2015 New Year countdown event will be held in Bundyuan, which will be specifically organized by the Huangpu District Tourism Bureau. At the same time, the relevant departments of the district are required to implement various safeguard measures for the activities. On December 26, the Huangpu Public Security Bureau issued a decision on the safety permit for large-scale mass activities, and approved the application of the District Tourism Bureau to hold a countdown to the New Year.

(2) Preparations of Huangpu District

1. Huangpu District Government. On December 9, 2014, the 76th executive meeting of the Huangpu District Government made it clear: "The District Public Security Bureau should work with the District Municipal Committee (ie the Huangpu District Municipal Management Committee, hereinafter referred to as the 'Huangpu District Municipal Committee') and other departments to make a plan for activities, as soon as possible. Sort out all kinds of welcome activities in the whole area on the day of the event, carefully study the response plan, achieve overall coordination, orderly arrangement, and rationally deploy various support forces to ensure the safety and order of key areas such as the Bund, People's Square, and Xintiandi."

December On the evening of the 31st, the Huangpu District Government did not strictly implement the 24-hour special personnel on duty and leadership-led shift system.

2. Huangpu Public Security Bureau. On December 25, the Huangpu Public Security Bureau formulated a work plan for the safety and security of the New Year's countdown event. The main content is to establish a security work command for the New Year's countdown event. There are two sub-commands for on-site control and order maintenance along the Bund and Nanjing Road. A total of 771 security police and 180 host security guards were arranged for the New Year countdown event. Among them, there are 350 police officers for order maintenance along the Bund and Nanjing Road (60 in Chen Yi Square and 7 in the stairs), and the rest are used for on-site management and control of activities in the Bund, anti-terrorism, comprehensive security, public safety management, mobile armed police, etc. .

3. Huangpu District Municipal Committee. On December 31, the Huangpu District Municipal Committee and its affiliated Huangpu District Bund Scenic Area Management Office arranged a total of 108 city management law enforcement officers and social auxiliary forces to participate in the management of the Bund Scenic Area during the middle shift period (the daily middle shift period). The working hours are from 14:15 to 22:15, and the scheduled working hours of the day are from 14:15 to 1:00 am the next day).

4. Huangpu District Tourism Bureau. On December 9, 2014, the 76th executive meeting of the Huangpu District Government passed the plan for the New Year's countdown to be held at the Bund Source formulated by the Huangpu District Tourism Bureau. At 9:30 a.m. on December 30, the Huangpu District Information Office held a press conference, and the Huangpu District Tourism Bureau released information on the countdown to the New Year.

(3) Relevant work of Shanghai Public Security Bureau

On December 19 and 24, 2014, the Shanghai Public Security Bureau held two party committee meetings successively to study and deploy the security and stability maintenance work for New Year's Day and Spring Festival. On December 25 and 28, a special meeting of the leaders of the public security sub-bureaus was held to forward the Ministry of Public Security's "Notice on Effectively Doing a Good Job in Security and Stability Maintenance During the New Year's Day and Spring Festival in 2015", and put forward clear requirements

for doing a good job in the security and stability maintenance work during the New Year's Day and Spring Festival in 2015. On December 30, the main leaders of the Shanghai Public Security Bureau emphasised at a video conference on mobilizing and deploying security and stability work that the lighting and light show ceremonies in the Shanghai Center may cause a large number of tourists to gather in Lujiazui, the Bund and other related areas in a short period of time. According to the requirements of "One Action Plan" and "One Point One Plan", a thorough security work plan and emergency response plan have been formulated, and the police force configuration at the event site has been strengthened.

3. Changes in the flow of people in the Bund Scenic Area on the night of the incident and its response

(1) The overall flow of people in the Bund Scenic Area on the night of the incident

From 20:00 on the night of the incident, there were more people entering and leaving the Bund Scenic Area, and a large number of citizens and tourists flocked to the Bund viewing platform, showing a gradual gathering of people. Afterwards, according to the comprehensive analysis of the data provided by the Shanghai Communications Administration, the Shanghai Public Security Bureau, the subway operating company (ie Shentong Group) and other departments, the flow of people in the Bund Scenic Area that night was about 120,000 people from 20:00 to 21:00. , about 160,000 people from 21:00 to 22:00, about 240,000 people from 22:00 to 23:00, and about 310,000 people from 23:00 to the time of the incident.

(2) Personnel flow monitoring and reporting in the Bund Scenic Area

At 20:12, the command center of the Shanghai Public Security Bureau asked the command center of the Huangpu Public Security Bureau to report the flow of people in the Bund Scenic Area and Nanjing Road Pedestrian Street on the hour. At 20:20, the command center of the Huangpu Public Security Bureau reported that 50% of the people flow on the viewing platform of the Bund Scenic Area (the police judged the density of people based on their experience), and about 5-6 people flow on Nanjing Road (Henan Road to Zhongshan East Road) become.

At 20:27, the command center of the Shanghai Public Security Bureau requested the command center of the Huangpu Public Security Bureau to report the flow of people in the

Bund Scenic Area and Nanjing Road Pedestrian Street every half hour. At 21:14, the command center of the Huangpu Public Security Bureau reported that 50% of the personnel flow in Chen Yi Square was normal; at 21:39, Sun Cheng, the commander of the command center of the Huangpu Public Security Bureau, called Cai Lixin of the Bund sub-command to inquire about the Bund Scenic Area and Nanjing Road. Cai Lixin's response to the flow of people on the pedestrian street was 60-70%, but the radio and telephone records did not show that it was reported to the command center of the Shanghai Public Security Bureau; at 22:45, the Huangpu Public Security Bureau reported that the flow of people on the viewing platform of the Bund Scenic Area was 5- 60%.

(3) The deployment of police force in Huangpu Public Security Bureau

At 20:25, Chen Ronglin, assistant to the director of the Huangpu Public Security Bureau, asked the command center to collect the number of police forces that can be called by each police station in the area.

At 21:09, Chen Ronglin called Cai Lixin. The Bund Scenic Area is still under the command of Chen Qi, deputy director of the Huangpu Public Security Bureau and the on-site commander of the Bund branch headquarters, and the deputy director Yao Hua will be in charge along the Nanjing Road, and the police deployment will be adjusted. . At 21:13, the Huangpu Public Security Bureau mobilized 40 police officers from 5 police stations in the area to reinforce Nanjing Road and arrived at the post around 22:00; 20 police officers from the police training team arrived.

At 21:20, Yu Lei, the commander of the command center, led 40 police officers to the Bund sub-command for reinforcements.

At 20:27, Zhou Zheng, director of the Huangpu Public Security Bureau, called Chen Qi to inquire about the flow of people and the allocation of police forces on the Nanjing Road Pedestrian Street. Zhou Zheng asked to take the "iron horse" measures and arrange a police force every 10 meters; a one-way traffic warning belt was set up at the entrance of the viewing platform of the Bund Scenic Area.

Since then, the Huangpu Public Security Bureau has adjusted the existing police force deployment according to the flow of personnel. At 22:44, Zhou Zheng called Yao Hua and asked the police along Nanjing Road to set up an interception line to slow down the flow of people towards the Bund.

After investigation, as of 23:30, the Huangpu Public Security Bureau deployed a total of 510 police officers in the Bund Scenic Area and along Nanjing Road, including 80 in Chenyi Square (13 at the stairs) and 150 along Nanjing Road.

(4) Shanghai Public Security Bureau's orders and police reinforcements

At 20:31, the main leader of the Shanghai Public Security Bureau called Zhou Zheng, asking to pay attention to the flow of people in the Bund Scenic Area and implement security measures.

At 22:00, Chen Zhen, Yu Lie, and Lu Min, deputy directors of the Shanghai Municipal Public Security Bureau, were in the lobby of the Municipal Bureau Command Center to check the flow of people in the Bund Scenic Area and other areas through video surveillance. Chen Zhen reminded Zhou Zheng through a video call that there were already many people in the Bund Scenic Area, and asked him to conduct a safety assessment, and if there was a need, he would submit it in time. In this regard, the Huangpu Public Security Bureau did not ask for it.

At 22:12, the command center of the Shanghai Public Security Bureau instructed to train 200 personnel from the third department, 100 from the fourth department, 100 from the Jing'an Public Security Bureau, 50 from the Hongkou Public Security Bureau, and 50 from the Zhabei Public Security Bureau. A total of 500 police officers reinforced the Huangpu Public Security Bureau. At 23:10, reinforcements from the Hongkou Public Security Bureau arrived at the intersection of Zhongshan East 1st Road and Dianchi Lake. The task was to be responsible for guarding and dispersing the crowd along the Zhongshan East 1st Road (Nanjing East Road to Beijing East Road); at 23:30, the reinforcements of the Zhabei Public Security Branch arrived. The entrance of the sightseeing tunnel on the north side of Chen Yi Square, the task is to guide and evacuate the crowd on the viewing platform. At 23:35, the reinforcement police force of the third training department and the fourth training department were located at the intersection

of Fuzhou Road, East Zhongshan Road, and East Beijing Road, respectively, and the reinforcements of Jing'an Public Security Bureau had not yet arrived.

At 23:24, the command center of the Shanghai Public Security Bureau called the command center of the Huangpu Public Security Bureau, suggesting that motor vehicles should be diverted and diverted at both ends of Zhongshan East Road along the Bund Scenic Area, and temporary traffic control measures should be implemented. The command center of the Huangpu Public Security Bureau replied that Zhou Zheng's opinion is to keep the traffic flowing, otherwise people will think that there is a large-scale event in the Bund Scenic Area.

4. Incident occurrence and emergency response and rescue situation

(1) After the incident

At 22:37, after the one-way traffic warning belt at the pedestrian passage steps on the north side of the southeast corner of Chenyi Square on the Bund was broken, the police on duty tried their best to maintain order, and a large number of citizens and tourists still rushed to the viewing platform. From 23:23 to 33:00, the flow of people up and down continued to hedge and formed a stalemate in the middle of the stairs, and then formed a "surge". At 23:35, the pressure on the flow of deadlocked people increased sharply, causing someone to fall out of balance at the bottom of the stairs, which in turn caused many people to fall and overlap, resulting in a crowded stampede incident.

(2) On-site rescue situation

After the crowded and stampede incident at 23:35, the police who maintained order at the scene tried to pull out the nearby fallen person together with the citizens and tourists, but because the fallen person was still squeezed by the flow of people above, many attempts were unsuccessful. After that, many citizens and tourists at the stairs climbed over the handrails with the help of others. The flow of people above the stairs began to retreat under the command of the police and enthusiastic citizens and tourists. The density of people above gradually decreased. Rescue on the ground. Many citizens and tourists spontaneously surrounded the adult wall with their bodies, creating an ambulance channel with a width of about three meters.

The doctors and nurses among the citizens and tourists at the scene spontaneously joined the rescue work and carried out emergency rescue for the injured with vital signs.

From 23:41:22, Shanghai "120" medical emergency center received emergency calls one after another. From 23:49, 19 ambulances arrived at Chen Yi Square successively, carrying out on-site treatment and the transfer of the wounded as soon as possible. The Shanghai Municipal Public Security Bureau and Huangpu Public Security Bureau quickly opened emergency channels, mobilized police, buses and other social vehicles, and sent injured citizens and tourists to Ruijin Hospital, Changzheng Hospital, Shanghai First People's Hospital and Huangpu District Central Hospital for rescue. At the same time, quickly organize forces to do everything possible to collect information on casualties, and promptly contact the units where the casualties are located and their families.

(3) Emergency response and aftermath after the incident

After the incident, the main leaders of the municipal party committee and the municipal government rushed to the scene to direct the emergency response work, and rushed to the hospital to visit the injured and their families. At the same time, an emergency meeting was held overnight, and it was decided to set up a special working group and a joint investigation group for medical treatment and aftercare, and each group started work immediately.

Mobilize the city's high-quality medical resources to go all out to treat the wounded. On the basis of expert consultation and evaluation, in accordance with the requirements of "one person, one plan and one person, one expert", the medical plan is defined one by one, and every possible effort is made to save lives. As of January 20, 49 Among the injured, 46 people have been discharged from hospital after diagnosis and treatment (including 11 of the 13 seriously injured), and 3 wounded (2 serious and 1 minor) are still in hospital for treatment. Confirm the identity of the casualties as soon as possible through a variety of channels, publish the list of the victims to the public in a timely manner, and follow up the injured discharged from the hospital. Assign special personnel to do their best to receive and comfort the families of the casualties, and organize professionals to provide psychological counseling to the injured and the families of the casualties. By organizing collective interviews, written publications, "Shanghai Publishing" government affairs Weibo and WeChat, etc., timely release relevant information to the media and society.

On the morning of January 1, the municipal party committee and the municipal government held an emergency meeting of responsible cadres of the party and government in the city to fully deploy various aftermath work and safety precautions on the whole market, and express deep condolences to the victims before the meeting. On January 4th, the city leaders participated in the pre-meeting activities of the delegations of the third session of the 14th Municipal People's Congress and the 38th chairman meeting of the 12th CPPCC. Before the meeting, all the participants stood in silence to express their deep condolences to the victims. . On January 7, the municipal party committee and the municipal government held a city-wide safety work meeting, requiring a comprehensive investigation of various safety hazards, carefully sorting out weak links and shortcomings one by one, and solving them carefully one by one, and effectively do a good job in crowded places safety management work.

5. Reason Analysis

Regarding the gathering of people in the Bund Scenic Area, especially Chen Yi Square on the night of the incident, the leaders of the Huangpu District Government and relevant departments were paralyzed, and they seriously lacked the awareness of public safety risk prevention. The serious lack of preparedness and response, poor early warning and improper response measures on the night of the incident were the main reasons for the crowded stampede.

(1) The risk of changing the countdown to the new year has not been assessed

A large number of citizens and tourists believe that the New Year's countdown will still be held in the Bund Scenic Area. The Nanjing Road Commercial Street, the Shanghai Center on the other side of the Huangpu River, and the Oriental Pearl Tower have attracted some citizens and tourists to watch them. In this regard, when the Huangpu District Government changed the New Year's countdown activities, it did not attach great importance to the possible safety risks of people gathering, and did not conduct assessments.

(2) The publicity of the change information of the New Year's countdown event is seriously not in place

After the New Year countdown event is changed, the organiser shall fully inform the society of the event information in advance. However, it was not until December 30 that the Huangpu District Tourism Bureau officially released the information on the New Year's countdown activities. There was no special reminder and extensive publicity for the difference between "The Bund" and "The Bund Source", and the information announcement was not timely, in place, and insufficient. .

(3) A serious lack of preventive preparation

The Huangpu Public Security Bureau did not comply with the requirements of the executive meeting of the Huangpu District Government, and in the safety and security work plan for the New Year's countdown activity compiled, it only conducted a safety assessment for the New Year's countdown activity at the Bund Source, and did not conduct a special assessment of the security risks of the Bund Scenic Area. The Huangpu Public Security Bureau, together with the Huangpu District Municipal Committee and other relevant departments, only deployed 350 police officers, 108 city management and auxiliary personnel, and 100 armed police officers in the Bund Scenic Area and along Nanjing Road. The deployment of security personnel is seriously insufficient.

(4) Failure to timely study, judge and give early warning of changes in the flow of monitoring personnel, and fail to issue reminder information

From 20:00 on December 31 to the time of the incident, the flow of people in the Bund Scenic Area showed an upward trend. The command center of the Huangpu Public Security Bureau did not strictly implement the work requirements of the command center of the Shanghai Public Security Bureau to report the monitoring of personnel flow every half an hour, nor did it report to the duty office of the Huangpu District Party Committee and District Government in time. The Huangpu Public Security Bureau did not take effective measures in time for the rapid increase in the flow of people at various times, and did not report to the Huangpu District Government to issue an early warning to control the development of the situation. No response was made to the situational research and judgment requests repeatedly reminded by the Shanghai Public Security Bureau.

(5) Responding to Improper Handling

In response to the continuous increase in the flow of people on the night of the incident, the Huangpu Public Security Bureau only made partial adjustments to the deployment of the police force, and did not take other effective measures. The bureau reported that no reinforcement needs were made to the Shanghai Public Security Bureau, nor were the relevant instructions of the Shanghai Public Security Bureau implemented, and the disposal measures were inappropriate. The Shanghai Public Security Bureau did not properly guide and supervise the handling measures of the Huangpu Public Security Branch. The Huangpu District Government did not report the incident information to the municipal government in time.

# Appendix II

# Code Book

1. Inclusion criteria

The following criteria must be fulfilled for an incident to be included in this database.

I. Incident must occur between the 1st January of 1989 and 31st December of 2021.

II. The incident has been reported either in Chinese or English by a well-known news publisher.

III. People must get injured or died in the incident due to crowd force originated from high density of surrounding people.

2. Database variable

1) Case NO.

Numeric variable

Incidents in the database starts from Case NO.1 and follow the sequence of Arabic numerals.

2) Scenario

Categorical variable

This field contains the scenario in which the incident occurred. By scenario, we define it to be the main purpose of the victims and surrounding people to be present at the location of incident. In case of conflict, scenario when the incident took place will be recorded.

Scenario 1: Sports

The victims and surrounding people are present to watch a sports match.

Scenario 2: Performance

The victims and surrounding people are present to watch a performance that is taken place at a stage or a fixed area.

Scenario 3: Religion

The victims and surrounding people are present to attend an Islamic ritual, Hindu ritual or Christian ritual.

Scenario 4: Jostle

The victims and surrounding people are present to collect free item or pursue desired items that are limited in amount.

Scenario 5: Organized event

The victims and surrounding people are present to attend an organized event for celebrating a festival or memorable stuff other than religion.

Scenario 6: Nightclub

The victims and surrounding people are present for entertainment at a nightclub.

Scenario 7: School

The victims and surrounding people are present as students to study.

Scenario 8: Other

The victims and surrounding people are present for purposes other than listed in scenario 1-10.

Scenario 9: Unknown

The purpose for the present of victims and surrounding people are not reported.

3) Venue

Categorical variable

This field contains the venue in which the incident occurred. By venue, we define it to be the physical environment where victims get injured or killed.

Venue 1: Stadium

The physical environment where victims and surrounding people get injured or killed is a stadium.

Venue 2: Open area

The physical environment where victims and surrounding people get injured or killed is a spatial zone with a relatively low density of obstacles, where environmental variables such as wind, solar radiation, and temperature can be more easily measured and analysed.

Venue 3: Building

The physical environment where victims and surrounding people get injured or killed is an enclosed building other than stadium.

Venue 4: Restricted area

The physical environment where victims and surrounding people get injured or killed is a restricted area in outdoor space where their direction of movement is limited in at least one direction. Typical example falls under this category includes tunnel, bridge, street and waterside.

Venue 5: Other

The physical environment where victims and surrounding people get injured or killed is not listed in Venue 1 – Venue 4.

Venue 6: Unknown

The physical environment where victims and surrounding people get injured or killed is not reported.

Cause

The next three fields contain the causing factors for each incident, Information, Physical space and Time, identified by FIST model except crowd force.

4) Cause 1 Information
Categorical variable

This fields contains the information crowd act upon that leads to the high desired velocity in the incident.

Information 1: Imminent danger 1 fire

The information crowd act upon that leads to the high desired velocity is fire.

Information 2: Imminent danger 2 Other

The information crowd act upon that leads to the high desired velocity is other imminent danger that crowd will consider it dangerous or harmful to their body health other than fire.

Information 3: Unexpected incident

The information crowd act upon that leads to the high desired velocity is an incident that is not expected by the crowds, not harmful to their body health, but still gives them strong will of escape.

Information 4: Desirable 1

The information crowd act upon that leads to the high desired velocity is things desirable for most participants that limited in amount or things has to be done in a timely manner.

Information 5: Desirable 2

The information crowd act upon that leads to the high desired velocity is things desirable for most participants that their quality varied by distance with them.

Information 6: Official announcement

The information crowd act upon that leads to the high desired velocity is an official announcement from the event organisers or authorities of the venue.

Information 7: No

No information crowd act upon that leads to the high desired velocity is reported in the incident.

5) Cause 2 Physical space

Categorical variable

This fields contains the physical space of assembly facility that leads to crowd turbulent or clogging at low desired velocity (referred as normal walking) or high desired velocity (referred as rushing) in the incident.
This physical space should be either at or downstream of the incident location. When both situation applies, the physical space at the incident location is taken.
Here we choose to separate the case of bidirectional with unidirectional. Only if a bidirectional clash is clearly stated in the report, otherwise unidirectional is recorded.

Low 1: Staircase and escalators 1

The physical space of assembly facility that lead to crowd turbulent or clogging during bidirectional normal walking is staircase, escalators and lifts.

Low 2: Staircase and escalators 2

The physical space of assembly facility that lead to crowd turbulent or clogging during unidirectional normal walking is staircase, escalators and lifts.

Low 3: Entrance and exit 1

The physical space of assembly facility that lead to crowd turbulent or clogging during bidirectional normal walking is entrance and exit.

Low 4: Entrance and exit 2

The physical space of assembly facility that lead to crowd turbulent or clogging during unidirectional normal walking is entrance and exit.

Low 5: Bridge and Tunnel 1

The physical space of assembly facility that lead to crowd turbulent or clogging during bidirectional normal walking is bridge or tunnel.

Low 6: Bridge and Tunnel 2

The physical space of assembly facility that lead to crowd turbulent or clogging during unidirectional normal walking is bridge or tunnel.

Low 7: Other 1

The physical space of assembly facility that lead to crowd turbulent or clogging during bidirectional normal walking is not listed in normal speed 1, 3 and 5.

Low 8: Other 2

The physical space of assembly facility that leads to crowd turbulent or clogging during bidirectional normal walking is not listed in normal speed 2, 4 and 6.

High 1: Staircase and escalators 1

The physical space of assembly facility that lead to crowd turbulent or clogging during bidirectional rushing is staircase, escalators and lifts.

High 2: Staircase and escalators 2

The physical space of assembly facility that lead to crowd turbulent or clogging during unidirectional rushing is staircase, escalators and lifts.

High 3: Entrance and exit 1

The physical space of assembly facility that lead to crowd turbulent or clogging during bidirectional rushing is entrance and exit.

High 4: Entrance and exit 2

The physical space of assembly facility that lead to crowd turbulent or clogging during unidirectional rushing is entrance and exit.

High 5: Bridge and Tunnel 1

The physical space of assembly facility that lead to crowd turbulent or clogging during bidirectional rushing is bridge or tunnel.

High 6: Bridge and Tunnel 2

The physical space of assembly facility that lead to crowd turbulent or clogging during unidirectional rushing is bridge or tunnel.

High 7: Other 1

The physical space of assembly facility that lead to crowd turbulent or clogging during bidirectional rushing is not listed in rushing 1, 3 and 5.

High 8: Other 2

The physical space of assembly facility that lead to crowd turbulent or clogging during bidirectional rushing is not listed in rushing 2, 4 and 6.

No

The physical space of assembly facility that lead to crowd turbulent or clogging is not reported in the incident.

6) Cause 3 Time

Categorical variable

This field records the existence of long waiting time or slow crowd flow control in the incident.

Time 1: Ingress

Long waiting time or slow crowd flow control is identified in the incident during ingress flow.

Time 2: Egress

Long waiting time or slow crowd flow control is identified in the incident during egress flow.

Time 3: No

No indicator of time 1 or time 2 is reported in the incident.

7) Casualty

Categorical variable

This field contains the number of deaths of the incident. In case of conflict, the latest report from local news will be recorded.

0

No death in the incident.

1 -- 10

The number of deaths of the incidents are between 1 and 10.

11 – 50

The number of deaths of the incidents are between 11 and 50.

Over 50

The number of deaths of the incidents are over 50.

# Appendix III

# Questionnaire

**Information letter**

You are invited to take part in a research study about people's movement in crowd. Before you decide whether to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the participant information sheet carefully.

If you decide to participate you will only be able to withdraw at the end of each phase and you will be excluded from the rest of the experiment. Data that has already been recorded cannot be withdrawn because the results are heavily dependent on interactions of all participants.

**If you understand the above information and agree to complete the questionnaire, please tick the box below.**

☐ I confirm that I understand the information and agree to complete the questionnaire.

Principal Investigator: Professor Arnab Majumdar
Co-Investigator: Xiangmin Yang
Email: Xiangmin.yang18@imperial.ac.uk
Centre for Transport Studies,
Imperial College London, UK.

**Participant number: _____**

Please enter the number corresponding to your avatar above and leave no other personal information. All the data in this research will be processed in a fully anonymous manner.

1. How would you rate the level of immersion you experienced during the experiment, i.e., the degree to which you felt as if you were physically present, on a scale from 1 (not immersive at all) to 10 (extremely immersive)?

2. How closely do you believe the experiment reflects your behavior in real-life scenario, on a scale from 1 (not close at all) to 10 (extremely close)?

3. How do you rate the effectiveness of the dark circle with indicative arrows serves as an alternative of physical contact, on a scale from 1 (very bad) to 10 (very well)?

# Appendix IV

Below are the permissions obtained for reproducing figures from the publishers.

This Agreement between Xiangmin Yang ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

| | |
|---|---|
| License Number | 6224230517627 |
| License date | Mar 08, 2026 |
| Licensed Content Publisher | Springer Nature |
| Licensed Content Publication | Journal of Ambient Intelligence and Humanized Computing |
| Licensed Content Title | A review on technological advancements in crowd management |
| Licensed Content Author | Deepak Sharma et al |
| Licensed Content Date | Nov 18, 2016 |
| Type of Use | Thesis/Dissertation |
| Requestor type | academic/university or research institute |
| Format | print and electronic |
| Portion | figures/tables/illustrations |
| Number of figures/tables/illustrations | 1 |
| Will you be translating? | no |
| Circulation/distribution | 1 - 29 |
| Author of this Springer Nature content | no |
| Title of new work | A Socio-Technical Approach to Crowd Management: Understanding Decision-Making and Movement Patterns with Virtual Reality and Machine Learning |
| Institution name | Imperial College London |
| Expected presentation date | Apr 2026 |
| Portions | Figure 1 |
| The Requesting Person / Organization to Appear on the License | Xiangmin Yang |

| | |
|---|---|
| | Xiangmin Yang<br>Imperial College Road |
| Requestor Location | |
| | London, SW7 2BB<br>United Kingdom |
| Payment Type | Invoice |
| Email Address | xiangmin.yang18@imperial.ac.uk |
| | Xiangmin Yang<br>Imperial College Road |
| Billing Address | |
| | London, United Kingdom SW7 2BB |

This Agreement between Xiangmin Yang ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

| | |
|---|---|
| License Number | 6224630846022 |
| License date | Mar 09, 2026 |
| Licensed Content Publisher | Elsevier |
| Licensed Content Publication | Information Fusion |
| Licensed Content Title | Classical dynamic consensus and opinion dynamics models: A survey of recent trends and methodologies |
| Licensed Content Author | Hossein Hassani,Roozbeh Razavi-Far,Mehrdad Saif,Francisco Chiclana,Ondrej Krejcar,Enrique Herrera-Viedma |
| Licensed Content Date | Dec 1, 2022 |
| Licensed Content Volume | 88 |
| Licensed Content Issue | n/a |
| Licensed Content Pages | 19 |
| Start Page | 22 |
| End Page | 40 |
| Type of Use | reuse in a thesis/dissertation |
| Portion | figures/tables/illustrations |
| Number of figures/tables/illustrations | 1 |
| Format | both print and electronic |
| Are you the author of this Elsevier article? | No |
| Will you be translating? | No |
| Title of new work | A Socio-Technical Approach to Crowd Management: Understanding Decision-Making and Movement Patterns with Virtual Reality and Machine Learning |
| Institution name | Imperial College London |
| Expected presentation date | Apr 2026 |
| Portions | Figure 6 |
| The Requesting Person / Organization to Appear on the License | Xiangmin Yang |

| | |
|---|---|
| | Xiangmin Yang<br>Imperial College Road |
| Requestor Location | |
| | London, SW7 2BB<br>United Kingdom |
| Publisher Tax ID | GB 494 6272 12 |

This Agreement between Xiangmin Yang ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

| | |
|---|---|
| License Number | 6224621211437 |
| License date | Mar 09, 2026 |
| Licensed Content Publisher | Springer Nature |
| Licensed Content Publication | Springer eBook |
| Licensed Content Title | The Working Principle of Cellular Automata |
| Licensed Content Author | Xuewei Li, Jinpei Wu, Xueyan Li |
| Licensed Content Date | Jan 1, 2018 |
| Type of Use | Thesis/Dissertation |
| Requestor type | academic/university or research institute |
| Format | print and electronic |
| Portion | figures/tables/illustrations |
| Number of figures/tables/illustrations | 1 |
| Will you be translating? | no |
| Circulation/distribution | 1 - 29 |
| Author of this Springer Nature content | no |
| Title of new work | A Socio-Technical Approach to Crowd Management: Understanding Decision-Making and Movement Patterns with Virtual Reality and Machine Learning |
| Institution name | Imperial College London |
| Expected presentation date | Apr 2026 |
| Portions | Diagram 2.1 |
| The Requesting Person / Organization to Appear on the License | Xiangmin Yang |
| Requestor Location | Xiangmin Yang<br>Imperial College Road<br><br>London, SW7 2BB<br>United Kingdom |

This Agreement between Xiangmin Yang ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

| | |
|---|---|
| License Number | 6224501090405 |
| License date | Mar 08, 2026 |
| Licensed Content Publisher | Elsevier |
| Licensed Content Publication | International Journal of Disaster Risk Reduction |
| Licensed Content Title | Propagation characteristics of the pedestrian shockwave in dense crowd: Experiment and simulation |
| Licensed Content Author | Jinghong Wang,Manman Chen,Bowei Jin,Jia Li,Zhirong Wang |
| Licensed Content Date | Nov 1, 2019 |
| Licensed Content Volume | 40 |
| Licensed Content Issue | n/a |
| Licensed Content Pages | 1 |
| Start Page | 101287 |
| End Page | 0 |
| Type of Use | reuse in a thesis/dissertation |
| Portion | figures/tables/illustrations |
| Number of figures/tables/illustrations | 1 |
| Format | both print and electronic |
| Are you the author of this Elsevier article? | No |
| Will you be translating? | No |
| Title of new work | A Socio-Technical Approach to Crowd Management: Understanding Decision-Making and Movement Patterns with Virtual Reality and Machine Learning |
| Institution name | Imperial College London |
| Expected presentation date | Apr 2026 |
| Portions | Figure 5 |
| The Requesting Person / Organization to Appear on the License | Xiangmin Yang |

Xiangmin Yang
Imperial College Road

Requestor Location

London, SW7 2BB
United Kingdom

Publisher Tax ID GB 494 6272 12

This Agreement between Xiangmin Yang ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

| | |
|---|---|
| License Number | 6224500022728 |
| License date | Mar 08, 2026 |
| Licensed Content Publisher | Springer Nature |
| Licensed Content Publication | Sustainability Science |
| Licensed Content Title | A transdisciplinary typology of change identifies new categories of adaptations and forms of co-adaptation in coupled human and natural systems |
| Licensed Content Author | Garry Sotnik et al |
| Licensed Content Date | Jun 3, 2021 |
| Type of Use | Thesis/Dissertation |
| Requestor type | academic/university or research institute |
| Format | print and electronic |
| Portion | figures/tables/illustrations |
| Number of figures/tables/illustrations | 1 |
| Will you be translating? | no |
| Circulation/distribution | 1 - 29 |
| Author of this Springer Nature content | no |
| Title of new work | A Socio-Technical Approach to Crowd Management: Understanding Decision-Making and Movement Patterns with Virtual Reality and Machine Learning |
| Institution name | Imperial College London |
| Expected presentation date | Apr 2026 |
| Portions | Figure 1 |
| The Requesting Person / Organization to Appear on the License | Xiangmin Yang |
| Requestor Location | Xiangmin Yang<br>Imperial College Road<br><br>London, SW7 2BB<br>United Kingdom |

**Order Date**
13-Mar-2026
**Order License ID**
1709202-1
**ISSN**
1742-5468
**Type of Use**
Republish in a thesis/dissertation
**Publisher**
IOP Publishing
**Portion**
Image/photo/illustration

## LICENSED CONTENT

**Publication Title**
Journal of Statistical Mechanics : Theory & Experiment
**Author / Editor**
Institute of Physics Publishing., International School for Advanced Studies (Trieste, Italy)
**Date**
01/01/2004
**Language**
English
**Country**
United Kingdom of Great Britain and Northern Ireland
**Rightsholder**
IOP Publishing Limited
**Publication Type**
e-Journal
**URL**
http://iopscience.iop.org/1742-5468

## REQUEST DETAILS

**Portion Type**
Image/photo/illustration
**Number of Images / Photos / Illustrations**
1
**Format (select all that apply)**
Print, Electronic
**Who Will Republish the Content?**
Academic institution
**Duration of Use**
Life of current edition
**Lifetime Unit Quantity**
Up to 499
**Rights Requested**
Main product
**Distribution**
Worldwide

**Translation**
Original language of publication
**Copies for the Disabled?**
No
**Minor Editing Privileges?**
No
**Incidental Promotional Use?**
No
**Currency**
GBP

## NEW WORK DETAILS

**Title**
A Socio-Technical Approach to Crowd Management: Understanding Decision-Making and Movement Patterns with Virtual Reality and Machine Learning
**Instructor Name**
Washington Ochieng
**Institution Name**
Imperial College London
**Expected Presentation Date**
2026-04-01

## ADDITIONAL DETAILS

**Order Reference Number**
N/A
**The Requesting Person / Organization to Appear on the License**
Xiangmin Yang

## REQUESTED CONTENT DETAILS

**Title, Description or Numeric Reference of the Portion(s)**
Figure 4
**Editor of Portion(s)**
N/A
**Volume / Edition**
2012
**Page or Page Range of Portion**
P02002
**Title of the Article / Chapter the Portion Is From**
Ordering in bidirectional pedestrian flows and its influence on the fundamental diagram
**Author of Portion(s)**
J Zhang, W Klingsch, A Schadschneider and A Seyfried
**Issue, if Republishing an Article From a Serial**
02
**Publication Date of Portion**
2012-02-03

This Agreement between Xiangmin Yang ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

| | |
|---|---|
| License Number | 6224260969745 |
| License date | Mar 08, 2026 |
| Licensed Content Publisher | Springer Nature |
| Licensed Content Publication | Nature |
| Licensed Content Title | Simulating dynamical features of escape panic |
| Licensed Content Author | Dirk Helbing et al |
| Licensed Content Date | Sep 28, 2000 |
| Type of Use | Thesis/Dissertation |
| Requestor type | academic/university or research institute |
| Format | print and electronic |
| Portion | figures/tables/illustrations |
| Number of figures/tables/illustrations | 1 |
| Will you be translating? | no |
| Circulation/distribution | 1 - 29 |
| Author of this Springer Nature content | no |
| Title of new work | A Socio-Technical Approach to Crowd Management: Understanding Decision-Making and Movement Patterns with Virtual Reality and Machine Learning |
| Institution name | Imperial College London |
| Expected presentation date | Apr 2026 |
| Portions | Figure 1 |
| The Requesting Person / Organization to Appear on the License | Xiangmin Yang |
| Requestor Location | Xiangmin Yang<br>Imperial College Road<br><br>London, SW7 2BB<br>United Kingdom |

This Agreement between Xiangmin Yang ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

| | |
|---|---|
| License Number | 6224260499478 |
| License date | Mar 08, 2026 |
| Licensed Content Publisher | Elsevier |
| Licensed Content Publication | Journal of Building Engineering |
| Licensed Content Title | A study on the arch mechanism of pedestrian evacuation and congestion alleviation strategies at building exits |
| Licensed Content Author | Ran Ye,Jiachen Li,Hongcheng Lu,Jinghong Wang,Yong Pan,Yan Wang |
| Licensed Content Date | Jul 1, 2024 |
| Licensed Content Volume | 88 |
| Licensed Content Issue | n/a |
| Licensed Content Pages | 1 |
| Start Page | 109159 |
| End Page | 0 |
| Type of Use | reuse in a thesis/dissertation |
| Portion | figures/tables/illustrations |
| Number of figures/tables/illustrations | 1 |
| Format | both print and electronic |
| Are you the author of this Elsevier article? | No |
| Will you be translating? | No |
| Title of new work | A Socio-Technical Approach to Crowd Management: Understanding Decision-Making and Movement Patterns with Virtual Reality and Machine Learning |
| Institution name | Imperial College London |
| Expected presentation date | Apr 2026 |
| Portions | Figure 24 |
| The Requesting Person / Organization to Appear on the License | Xiangmin Yang |
| Requestor Location | Xiangmin Yang<br>Imperial College Road |

London, SW7 2BB
United Kingdom
Publisher Tax ID GB 494 6272 12