



**HAL**  
open science

# The One-Man-Crowd: towards single-user capture of collective motions using virtual reality

Tairan Yin

► **To cite this version:**

Tairan Yin. The One-Man-Crowd: towards single-user capture of collective motions using virtual reality. Graphics [cs.GR]. Université de Rennes, 2024. English. NNT : 2024URENS011 . tel-04692549

**HAL Id: tel-04692549**

**<https://theses.hal.science/tel-04692549v1>**

Submitted on 17 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES

ÉCOLE DOCTORALE N° 601

*Mathématiques, Télécommunications, Informatique, Signal, Systèmes,  
Électronique*

Spécialité : *Informatique*

Par

« **Tairan YIN** »

« **The One-Man-Crowd : Towards Single-User Capture of Collective  
Motions using Virtual Reality** »

Thèse présentée et soutenue à « Rennes », le « 16/04/2024 »

Unité de recherche : « : Inria/IRISA »

## Rapporteurs avant soutenance :

Anthony Steed    Professor, University College London, UK  
William Warren    Professor, Brown University, USA

## Composition du Jury :

Président :	Anatole Lécuyer	Directeur de Recherche, Inria
Examineurs :	Anthony Steed	Professor, University College London, UK
	William Warren	Professor, Brown University, USA
	Nuria Pelechano	Associate Professor, Universitat Politècnica de Catalunya, Spain
	Ioannis Karamouzas	Associate Professor, University of California, Riverside, USA
	Anatole Lécuyer	Directeur de Recherche, Inria
Dir. de thèse :	Julien Pettré	Directeur de Recherche, Inria
Co-dir. de thèse :	Marie-Paule Cani	Professeure, Ecole Polytechnique
Co-Encadrant de thèse :	Ludovic Hoyet	Chargé de Recherche, Inria

## Invité(s) :

Marc Christie (Co-encadrant de thèse)    Maître de Conférence, Université de Rennes



# ACKNOWLEDGEMENT

---

Five years ago, when I expressed my desire to do a Ph.D. in France to my master's thesis supervisor, he told me, "If you want to do a Ph.D. in France, you have to go to INRIA because it is the best computer science research center in France". Three years ago, after much effort, I was lucky enough to get a PhD contract from INRIA. Now, as I finished writing this thesis, I couldn't help but relive the meaningful and colorful moments I spent with different people during my PhD.

First and foremost, I would like to thank my parents for their understanding and support of my decision to pursue a Ph.D. in France. Although they couldn't be very involved due to distance and language barriers, I know that they were always thinking of me from afar in China.

Next, I would like to thank the team secretaries, H el ene, Nathalie and Gwenaelle, who helped me with various administrative tasks. I would also like to thank Fran ois, who coordinated the meeting rooms when I was preparing for experiments, and Florence, who lent me the keys every morning during the weeks I was doing experiments.

I would also like to thank all the colleagues I have met in the laboratory: Thomas, Vincenzo, Maxime, Lev, Julien, Fabien, Wouter, Ad ele, Katja, Pierre, Alberto, Yuliya, Diane, Benjamin, Alexis, C eline, Tony, Arthur, and others I may have forgotten to mention. I will cherish the beer nights during the first year of my PhD. I am also grateful to my Chinese colleagues in the lab, who gave me a lot of help and advice: Lisheng, Hanwei, Xi, Qian, Kai, Shenghao, Yiqun, Xiaoyuan, Yahao, and Chencheng. Special thanks to Pratik, the postdoc in our group and my neighbor in the same dorm. Thanks to our proximity, we often engaged in academic discussions in our spare time. Most of all, I appreciate his advice not to hesitate to tell your supervisor if you don't understand something. I would also like to thank R emi and Anne H el ene for their proactive help during my thesis defense.

This Ph.D. journey was funded by the European Union's Horizon 2020 Research and Innovation Program under the Marie Sk lodowska-Curie Grant Agreement No. 860768 (CLIFE project). I feel very lucky to have been part of this project. Through the various workshops and secondments organized by the project, I had the opportunity to meet renowned researchers and industry professionals. More importantly, I met fellow PhD stu-

---

dents from excellent research institutions in other EU countries: Nefeli, Marilena, Rafael, Haoran, Vincenzo, Klara, Lisa, Bharat, Mirela, Yuliang, Radek, Kiran, Ariel, Eduardo. I am grateful for the time we spent together and the academic discussions we had. I would also like to thank Yiorgos and Nuria for their efforts in organizing the workshops.

I would like to thank the examiners who participated in my defense, especially Anthony and William, for their time and effort in reviewing my thesis.

Finally, I would like to thank each of my advisors for their guidance and support throughout my Ph.D. journey. My four advisors, who are also my co-authors, were deeply involved in the writing of each paper, greatly improving the quality of the writing and teaching me valuable writing skills. I feel very fortunate.

- I am deeply grateful to Marc for his guidance with my academic writing in English. His insights and feedback have been invaluable in helping me improve the clarity and precision of my writing.
- I am profoundly grateful to Ludovic for his ubiquitous presence in my research. He has guided me the most in my experimental design and code implementation, tirelessly correcting my grammatical mistakes in writing. His dedication and patience have been instrumental in my progress, and his support has been unwavering.
- I am grateful to Marie-Paule for her active participation in my thesis. Despite the distance between Paris and Rennes, she always tried to follow my experimental progress and often provided valuable suggestions. Her efforts to stay engaged despite the logistical challenges have greatly enriched my research experience.
- I am truly grateful to Julien for his support throughout all stages of my PhD. As my supervisor and the visionary behind One-Man-Crowd, as well as the leader of our team, Julien became increasingly busy over the years, making our weekly meetings difficult to maintain. Nevertheless, Julien's support was always timely and effective at various stages of my PhD. His optimistic attitude and deep understanding of the research topic boosted my confidence in my work. His excellent writing skills will always be something I aspire to. During the final stages of my PhD, I was under considerable mental pressure, but I did not share this with my advisors. Eventually, I realized that Julien had long sensed my stress and had been silently supporting me.

People often say that getting a Ph.D. is a challenging journey. In reality, it's more like being a frog in slowly boiling water. Most graduate students I know start their first year excited about the freedom to manage their own time and research topics, then grow

---

anxious as graduation approaches, and finally, for various reasons, experience immense mental pressure in the final stages. From another perspective, or at least for me personally, this journey has witnessed my desire to succeed and the efforts I have made to chase my dreams. At certain points, it taught me, or even forced me, to re-examine my attitude towards the world and my approach to self-care. This experience has shaped me not only as a researcher, but also as an individual, instilling in me a resilience and adaptability that I will carry forward. As I close this chapter and look to the future, I am filled with a deep sense of gratitude and anticipation for the next challenges and adventures that lie ahead.

# RÉSUMÉ EN FRANÇAIS

---

## Contexte

Depuis les années 1990, les études sur le comportement des foules se sont multipliées. Représentées par les premiers travaux de Helbing and Molnar [1995], ces études sur les foules tentent de comprendre les comportements des foules et de modéliser leurs mouvements. Ces efforts ont permis de simuler les mouvements de foule sur ordinateur, et donc de prédire le comportement des foules. Aujourd'hui, les résultats de ces études sur les foules sont étroitement liés à notre vie : dans les domaines de la conception architecturale et de la sécurité publique, les concepteurs utilisent des données empiriques pour évaluer la capacité d'évacuation des bâtiments où les flux humains sont importants, et utilisent la simulation de foule pour simuler le flux de personnes lors de grands événements (tels que les événements sportifs, les concerts et les rassemblements) ainsi que pour prédire les risques ; dans le domaine de la vie numérique et des loisirs, les personnages virtuels et les foules simulées sont couramment vus sur les écrans de télévision et de cinéma, et jouent également un rôle important dans les jeux vidéo en interagissant avec les joueurs ; d'autre part, dans le domaine universitaire, la recherche et la simulation du comportement des foules sont en constante évolution pour mieux comprendre le comportement des foules et les causes de ces comportements, et pour essayer continuellement de construire des simulations de foules plus réalistes, au service des deux domaines précédents. Il n'est pas difficile de voir que la simulation de foule joue un rôle crucial dans ces scénarios, et rendre les foules virtuelles générées aussi réalistes que possible est l'objectif le plus important de la simulation de foule. Dans ce contexte, les données relatives au comportement des foules ont joué un rôle de plus en plus important : des années 1990 à 2015 environ, les méthodes courantes ont recueilli des données sur le comportement des foules et ont procédé à une modélisation mathématique, formant des modèles qui couvrent les caractéristiques dynamiques des mouvements de foule, les comportements de choix des foules dans les évacuations d'urgence, etc. et les ont utilisés pour simuler le comportement des foules virtuelles. D'autre part, ces méthodes fondées sur la connaissance évaluent souvent le réalisme des simulations en les comparant à des données de foules réelles lors de la

---

simulation du comportement des foules dans différents scénarios. Depuis le milieu des années 2010, des méthodes de simulation visant à reproduire les caractéristiques des données réelles sont apparues, notamment des méthodes qui recherchent et comparent des ensembles de données réelles pour générer le comportement de foules virtuelles, ainsi que des méthodes plus populaires de réseaux neuronaux profonds basées sur l'apprentissage statistique. Bien que ces méthodes ne s'efforcent plus d'acquérir des connaissances à partir des données de la foule, elles dépendent tout autant des données de la foule, en particulier des données de haute qualité.

Jusqu'à présent, la plupart des données sur les foules proviennent d'expériences contrôlées ou d'observations sur le terrain. Toutefois, ces méthodes se heurtent à des limites importantes, notamment les défis techniques liés à la collecte de données de haute qualité, mais également les contraintes éthiques telles que l'obligation de ne pas blesser les sujets. Ces difficultés se traduisent par une couverture limitée des scénarios par les données et un manque de compréhension des détails du comportement individuel. D'un point de vue microscopique, les mouvements de foule sont une combinaison de multiples mouvements individuels, où les membres de la foule interagissent localement avec les autres. Ces interactions locales sont riches en comportements de bas niveau tels que l'observation (regard ou dévisagement), la communication (verbale ou non verbale) et l'évitement (mouvement de tout le corps ou d'une partie du corps). Ces comportements de bas niveau sont précieux pour comprendre les mouvements de foule et peuvent être des éléments importants pour la perception du réalisme des foules. Cependant, la plupart des ensembles de données existants ne couvrent que les informations relatives à la trajectoire des piétons et ne sont pas en mesure de refléter les détails des comportements complexes au niveau individuel. Dans le parcours des chercheurs qui cherchent continuellement à simuler des foules de haute fidélité, l'absence de telles données signifie qu'il est impossible de simuler des comportements similaires.

Récemment, avec la mise en œuvre de la technologie de la réalité virtuelle, de plus en plus d'études sur les foules ont choisi de mener des expériences dans des environnements virtuels pour étudier le comportement des piétons. De plus, avec la maturation continue des dispositifs de capture du comportement accompagnant la Réalité Virtuelle (RV) – y compris le suivi du regard, du corps, des mains et des expressions faciales – il est devenu possible de collecter les comportements d'interaction des utilisateurs réels et de la foule dans des environnements virtuels. Les plateformes d'expérience sociale de la RV, telles que VR chat et Ubiq [Friston et al. 2021], permettent idéalement à plusieurs utilisateurs



---

réels de participer simultanément à une scène de RV pour des expériences, ce qui permet d'obtenir des données comportementales de haute qualité pour chaque utilisateur. Cependant, la participation de plusieurs sujets implique la nécessité de disposer de plusieurs ensembles d'équipements de réalité virtuelle, ainsi que de relever des défis techniques tels que la synchronisation des données multi-terminales à faible latence. Par conséquent, les recherches sur les foules basées sur la RV conçoivent toujours des expériences impliquant un seul utilisateur en même temps. Par conséquent, pour créer un environnement de foule dans le monde virtuel, la plupart des études s'appuient généralement sur la simulation de foule pour construire des foules virtuelles. Cependant, comme nous l'avons mentionné précédemment, la technologie actuelle de simulation de foule n'est toujours pas en mesure de simuler un comportement complexe et réaliste de piétons semblables à des humains dans des foules virtuelles, et ce qui limite le développement de méthodes de simulation dans cette direction est précisément le manque de données. Il n'est pas difficile d'imaginer le dilemme que pose ce problème de l'œuf ou de la poule : si le comportement de la foule virtuelle simulée est trop irréaliste, l'interaction entre l'utilisateur humain et la foule virtuelle le sera également. Ainsi, si les données collectées ne sont pas suffisamment réalistes, comment construire une simulation réaliste sur la base de ces données ?

## Objectifs & Défis

Compte tenu de ces circonstances, nous pensons qu'il est crucial de proposer une méthode rapide et efficace pour capturer des données sur les mouvements de foule. Cette méthode devrait enregistrer finement la dynamique et les comportements de chaque piéton pendant qu'il se déplace dans une foule et être facilement évolutive pour répondre à d'éventuelles exigences futures en matière de capture de données. D'autre part, nous espérons que les données collectées avec cette méthode proviennent toutes d'actions et de comportements réels d'utilisateurs humains. À cette fin, nous envisageons d'utiliser la réalité virtuelle pour intégrer les utilisateurs humains dans un environnement virtuel, puis de collecter les comportements des utilisateurs humains dans le monde réel à l'aide de dispositifs portables de capture des mouvements. Il est vrai qu'il existe des solutions similaires dans la réalité qui utilisent simplement des dispositifs portables de capture de mouvement pour suivre et enregistrer le comportement des utilisateurs. Cependant, la RV a progressivement montré trois avantages majeurs par rapport aux recherches antérieures : Premièrement, les expérimentateurs ont un contrôle élevé sur l'environnement virtuel, ce qui

---

permet de modifier facilement les conditions expérimentales dans l'environnement virtuel, et donc de compléter efficacement les préparations expérimentales. En même temps, avec l'aide de la technologie de simulation et d'infographie, il est facile de construire des scénarios dans l'environnement virtuel qui sont difficiles à créer dans les expériences de foule dans le monde réel (comme les flammes) sans causer de dommages physiques directs aux sujets et aux expérimentateurs. Deuxièmement, les scénarios expérimentaux de la RV ont un degré élevé de répétabilité, ce qui permet aux expérimentateurs de présenter des situations expérimentales totalement cohérentes à différents sujets, par exemple en faisant en sorte que les foules virtuelles se comportent de manière identique devant différents sujets, ou en simulant précisément les mêmes flammes au même moment et au même endroit. En revanche, en raison de la grande complexité des mouvements de foule, même avec les mêmes conditions initiales, il n'y a aucun moyen de garantir qu'une foule réelle effectuerait exactement les mêmes actions dans deux expériences. Troisièmement, les scénarios de RV permettent une collecte et un traitement plus efficaces des données. Les expérimentateurs disposent d'informations complètes sur la scène virtuelle et peuvent calculer les objets des comportements d'interaction de l'utilisateur en temps réel grâce à la technologie de capture de mouvement qui complète les dispositifs de RV. Par exemple, si le suivi des yeux est effectué dans un environnement virtuel, les expérimentateurs peuvent non seulement savoir quel type de regard l'utilisateur a effectué, mais aussi calculer quels objets ou personnages virtuels l'utilisateur a regardés.

Cependant, l'utilisation de la RV pour la capture des mouvements de foule se heurte à des difficultés. Tout d'abord, des recherches antérieures ont révélé que, sous l'influence des dispositifs de RV, le comportement et la perception de l'utilisateur dans les environnements virtuels ne sont pas les mêmes que dans le monde réel [Feng et al. 2021]. Deuxièmement, la capture de mouvement de foules dans la RV ne signifie pas simplement l'application de la solution de capture de mouvement de la RV pour un seul utilisateur à plusieurs utilisateurs : cela nécessite d'abord l'achat d'une grande quantité d'équipement matériel pour la capture de mouvement de plusieurs utilisateurs, et signifie également la reconstruction de l'état réel de chaque utilisateur dans l'environnement virtuel en temps réel, ce qui pose des défis pour la stabilité de la transmission des données. De plus, pour éviter les collisions dans le monde réel pour les utilisateurs portant des HMD (Head-Mounted Display, signifiant *affichage tête haute* en français) et pour éviter les interférences de signaux complexes causées par un équipement excessif, des sites expérimentaux plus grands ou même plusieurs dispersés sont nécessaires pour accueillir ces utilisateurs, ce qui

---

nécessite ensuite une équipe technique plus importante pour assurer la maintenance et le bon déroulement des expériences, ce qui conduit finalement à une escalade des coûts expérimentaux.

Compte tenu des possibilités et des défis qu'offre la RV, cette thèse tente d'explorer une méthode qui nous permet de collecter des données sur les mouvements de foule tout en veillant à ce que cette méthode soit rentable. Étant donné que les expériences de RV courantes n'impliquent qu'un seul utilisateur, nous espérons réaliser la capture de mouvement d'une foule avec un seul utilisateur. De cette manière, les recherches futures auront la possibilité d'utiliser l'équipement existant de l'expérimentateur pour constituer l'ensemble de données de foule requis à l'aide de ce paradigme de capture de mouvement de foule. En même temps, sur la base de la mise en œuvre d'une telle méthode, nous devons également vérifier que l'ensemble des données collectées par cette méthode est valable : évidemment, les données individuelles issues de cette méthode doivent être semblables à celles d'un être humain, car les données de chaque piéton proviennent de la capture du mouvement d'un véritable être humain. Cependant, au niveau collectif, nous devons nous assurer que les données obtenues par cette méthode présentent une similitude avec les données du monde réel. À notre connaissance, il n'existe pas à ce jour de données de foule similaires basées sur la capture de mouvements VR, ni de méthodes d'évaluation correspondantes. Cela pose le deuxième défi de notre travail, qui consiste à valider notre méthode par des moyens appropriés.

## Approche

Pour atteindre l'objectif de capturer des données sur les mouvements de foule avec un seul utilisateur, nous restructurons l'objectif de « capturer plusieurs piétons simultanément » en « capturer un seul piéton plusieurs fois », puis nous fusionnons les données de chaque piéton individuel pour former un ensemble complet de données sur les mouvements de foule. Pour mettre en œuvre ce processus, nous nous appuyons sur une approche d'enregistrement et de relecture. Cette approche est similaire à celle d'un acteur jouant plusieurs rôles au cinéma et à la télévision, où un acteur enregistre d'abord le comportement d'un personnage, puis recharge les données enregistrées dans la scène virtuelle, et joue un autre personnage dans le monde virtuel pour interagir avec le comportement précédemment enregistré, construisant ainsi l'interaction entre les deux personnages. Les caractéristiques techniques de la réalité virtuelle rendent ce processus d'enregistrement et

---

de relecture facile à mettre en œuvre, et il a été envisagé et appliqué dans des recherches antérieures [Slater et al. 1995; Osimo et al. 2015]. Dans cette thèse, nous incarnons un utilisateur dans un avatar virtuel dans une scène virtuelle à l'aide d'un HMD et d'un système de capture des mouvements du corps et nous laissons l'utilisateur générer le mouvement d'un piéton dans une foule virtuelle en marchant dans le monde réel. Au cours de ce processus, nous enregistrons le comportement de l'utilisateur. Une fois ce processus terminé, nous réinitialisons le temps dans la scène virtuelle et utilisons un personnage virtuel pour afficher le mouvement récemment enregistré de l'utilisateur. Pendant ce temps, l'utilisateur est incarné dans un autre avatar virtuel pour jouer un autre piéton virtuel. Ce processus est répété de manière à ce que l'utilisateur puisse agir de manière itérative et générer des comportements de marche pour différents personnages virtuels jusqu'à ce qu'une foule complète soit capturée. Nous appelons ce paradigme le paradigme de la foule-homme (One-Man-Crowd, OMC).

À ce stade, le paradigme de l'OMC se distingue par le fait qu'il évite toute technique de simulation pendant le processus de génération des données de la foule. Cette approche permet principalement d'éliminer les artefacts qui pourraient résulter du manque de réalisme des méthodes de simulation. Cependant, cette exclusion totale de la simulation limite notre capacité à utiliser pleinement les avantages généralement tirés des environnements de réalité virtuelle dans d'autres études de foules basées sur la RV.

Pour y remédier, nous avons exploré plusieurs variantes du paradigme de l'OMC. Dans ces variantes, nous utilisons des techniques de simulation de foule pour construire des foules virtuelles réactives, fournissant des scénarios contextuels pour la génération de données par l'utilisateur. Il est important de noter que ces mouvements de foule simulés sont utilisés uniquement pour la création de situations et ne sont pas conservés dans l'ensemble de données final. En incorporant ces méthodes, nous étudions les performances de l'approche OMC dans des scénarios de foule multidirectionnelle, en équilibrant le besoin de réalisme contextuel avec les principes fondamentaux du paradigme.

Pour évaluer le réalisme des données générées par notre méthode, nous adoptons des techniques d'évaluation couramment utilisées dans d'autres études de foules basées sur la RV. Une partie essentielle de notre évaluation consiste à comparer les données produites par notre approche avec les données de foule obtenues à partir d'expériences réelles. Étant donné la rareté des ensembles de données de foule existants qui incluent des mouvements détaillés de piétons, nous nous concentrons principalement sur l'évaluation des trajectoires de piétons.

---

## Contributions

Cette thèse propose deux contributions principales basées sur deux études expérimentales axées sur l’exploration de notre nouveau paradigme de capture de mouvements de foule en tirant parti de la réalité virtuelle pour fournir des données de foule de haute qualité.

Dans la première contribution, nous avons proposé un nouveau paradigme de capture de mouvements de foule basé sur la réalité virtuelle pour permettre de générer un ensemble de mouvements de foule avec un seul utilisateur. Nous appelons ce concept le paradigme de l’homme-foule, ou *the One-Man-Crowd (OMC)* en anglais. Sur la base d’une approche d’enregistrement et de relecture, nous avons subdivisé la capture de plusieurs piétons dans une foule en plusieurs itérations ultérieures de capture d’un seul piéton. En utilisant la capacité d’enregistrement de données et de visualisation d’animation de la réalité virtuelle, nous avons immergé un utilisateur dans le monde virtuel, puis enregistré le mouvement de l’utilisateur au cours d’une itération de la capture, et animé le mouvement capturé dans la même scène virtuelle lorsque l’utilisateur a effectué les captures suivantes. Ce processus a finalement abouti à un ensemble de données sur la foule. Nous avons testé le paradigme à l’aide de trois expériences qui reproduisaient trois expériences de contrôle de la foule menées dans des conditions de laboratoire réelles. En comparant nos données avec celles du monde réel, nous avons observé des similitudes qualitatives prometteuses en termes de modèles émergents. Malgré la présence de différences quantitatives, nos résultats ont validé l’utilisation du paradigme homme-foule pour capturer des données de foule qui présentent des comportements individuels et collectifs réalistes. Ce travail a été présenté à la conférence IEEE VR 2022 et a été accepté en tant qu’article de revue TVCG (nominé pour le meilleur article de revue), et est présenté dans le chapitre 3.

Puisque le paradigme homme-foule implique la génération de mouvements de différents piétons par un seul utilisateur, sa nature basée sur l’enregistrement et la relecture entraîne deux problèmes qui peuvent biaiser le comportement de l’utilisateur. Tout d’abord, l’utilisateur doit lancer le processus de capture dans une scène virtuelle vide et effectuera probablement des mouvements qui ne présentent aucune interaction avec les autres personnages générés. Deuxièmement, comme les personnages générés sont programmés pour n’afficher que les mouvements enregistrés au lieu d’être interactifs pendant leurs mouvements, le paradigme impose à l’utilisateur d’exécuter et d’augmenter progressivement la difficulté d’éviter les collisions, ce qui conduit à des interpénétrations inévitables entre

---

l'utilisateur (l'avatar virtuel de l'utilisateur) et les personnages. Pour remédier à ces deux limitations, nous avons proposé deux paradigmes améliorés qui introduisent respectivement une foule contextuelle au début du processus de capture pour déclencher un comportement d'interaction, et améliorent l'interactivité des personnages générés en utilisant des méthodes de simulation de foule. Nous avons mené deux expériences qui reproduisent des expériences de foule multidirectionnelle dans le monde réel en laboratoire. Nos résultats suggèrent que les comportements observés chez les utilisateurs de RV avec les agents environnants dès le début du processus d'enregistrement sont rendus beaucoup plus naturels, ce qui permet aux paradigmes 3R ou 4R d'améliorer la cohérence des ensembles de données de foule capturées. Ce travail a été présenté à IEEE VR 2024 et a été accepté comme article de revue TVCG (mention honorable pour les meilleurs articles). Ce travail est détaillé dans le chapitre 4.

## Aperçu

Cette thèse est structurée comme suit : Le chapitre 1 présente l'introduction générale de cette thèse. Le chapitre 2 fait le point sur l'état de l'art en matière d'étude des foules, en discutant de la collecte de données réelles et virtuelles et des méthodes de simulation. Le chapitre 3 présente notre première contribution, le paradigme de l'homme-foule. Le chapitre 4 présente notre deuxième contribution, où nous avons utilisé et examiné l'effet des foules contextuelles et réactives pour améliorer la capture des foules basée sur la RV. La thèse se termine par le chapitre 5, qui résume les résultats et suggère des orientations pour les recherches futures.



# TABLE OF CONTENTS

---

<b>1</b>	<b>Introduction</b>	<b>19</b>
1.1	Context . . . . .	19
1.2	Objective & Challenges . . . . .	21
1.3	Approach . . . . .	22
1.4	Contributions . . . . .	24
1.5	Overview . . . . .	25
<b>2</b>	<b>State of the Arts</b>	<b>26</b>
2.1	Crowd Study & Data Collection In Real World . . . . .	26
2.1.1	Laboratory Experiment . . . . .	28
2.1.2	Controlled Field Experiments . . . . .	32
2.1.3	Uncontrolled Field Experiment . . . . .	35
2.1.4	Field Recording . . . . .	37
2.1.5	Summary . . . . .	39
2.2	Crowd Study & Data Collection In Virtual Environment . . . . .	39
2.2.1	Crowd Related Research Topics . . . . .	40
2.2.2	Creating Immersive Experiences . . . . .	42
2.2.3	Validating Data From VR Experiments . . . . .	45
2.2.4	Summary . . . . .	47
2.3	Crowd Simulation & Crowd Data . . . . .	48
2.3.1	Knowledge-Based Approach . . . . .	48
2.3.2	Data-Driven and Deep Learning approaches . . . . .	50
2.3.3	Evaluation Methods . . . . .	52
2.3.4	Summary . . . . .	53
2.4	Conclusion . . . . .	54
<b>3</b>	<b>The One-Man-Crowd: Single User Generation of Crowd Motions Using Virtual Reality</b>	<b>55</b>
3.1	Introduction . . . . .	55



TABLE OF CONTENTS

---

3.2	Key Concepts & Notations . . . . .	56
3.2.1	Notations . . . . .	56
3.2.2	Formulating the One-Man-Crowd Paradigm . . . . .	57
3.3	General Hypotheses . . . . .	58
3.4	Experiment . . . . .	59
3.4.1	Experimental Design . . . . .	59
3.4.2	Scenarios & Tasks . . . . .	61
3.4.3	Apparatus . . . . .	65
3.4.4	Participants . . . . .	67
3.5	Evaluation . . . . .	67
3.5.1	Scenario I - Unidirectional Flow . . . . .	67
3.5.2	Scenario II - Bottleneck . . . . .	75
3.5.3	Scenario III - Inflow . . . . .	80
3.6	General Discussion . . . . .	83
3.6.1	Methodology . . . . .	83
3.6.2	S2N vs. N2N . . . . .	84
3.6.3	Character Animation Details . . . . .	85
3.6.4	Generalizing Interactions . . . . .	85
3.6.5	Generalizing Scenarios. . . . .	86
3.7	Beyond OMC: A First Step Towards Multidirectional Interactions . . . . .	86
3.7.1	Experiment . . . . .	87
3.7.2	Test Scenario A: Circle (Circle-24 & Circle-12) . . . . .	88
3.7.3	Test Scenario B: Crossing Flow . . . . .	92
3.7.4	Exploratory Insights . . . . .	95
3.8	Conclusion . . . . .	96
<b>4</b>	<b>Effect of Contextual and Responsive Crowds on VR-based Crowd Motion Capture</b>	<b>99</b>
4.1	Introduction . . . . .	99
4.2	Key Concepts & Notations . . . . .	103
4.2.1	Agents . . . . .	104
4.2.2	Crowd-motion-capture Paradigm . . . . .	105
4.3	General Hypotheses . . . . .	109
4.4	Experiment . . . . .	109

4.4.1	Experimental Design . . . . .	109
4.4.2	Scenarios & Tasks . . . . .	110
4.4.3	Participants . . . . .	113
4.5	Evaluation . . . . .	113
4.5.1	Analysis Framework . . . . .	114
4.5.2	Scenario IV - Circle . . . . .	118
4.5.3	Scenario V - Crossing Flow . . . . .	125
4.6	General Discussion . . . . .	133
4.6.1	Effect of Contextual Crowds . . . . .	133
4.6.2	Effects of Local Corrections . . . . .	135
4.7	Limitations . . . . .	137
4.8	Future Work . . . . .	137
4.9	Conclusion . . . . .	138
<b>5</b>	<b>General Conclusion &amp; Future Works</b>	<b>139</b>
5.1	Technical Limitations . . . . .	140
5.2	Discussion . . . . .	142
5.3	Future Works . . . . .	144
5.3.1	Short-Term Perspectives . . . . .	144
5.3.2	Long-Term Perspectives . . . . .	145
	<b>Bibliography</b>	<b>147</b>



# INTRODUCTION

---

## 1.1 Context

Since the 1990s, there has been an increasing number of studies on crowd behavior. Represented by the early work of Helbing and Molnar [1995], these studies on crowds attempt to understand the behaviors of crowds and to model their movement. These efforts have made it possible to simulate crowd movements in computers, thereby predicting crowd behavior. Today, the results of these crowd studies are closely related to our lives: in the fields of architectural design and public safety, designers use empirical data to assess the evacuation capacity of buildings with large human flows, and use crowd simulation to simulate the flow of people in large events (such as sports events, concerts, and rallies) as well as conduct risk prediction; in the field of digital life and leisure entertainment, virtual characters and simulated crowds are commonly seen on TV and movie screens, and also play an important role in video games interacting with players; on the other hand, in the academic field, research and simulation of crowd behavior are constantly iterating to better understand the behavior of crowds and the causes of these behaviors, and to continuously try to construct more realistic crowd simulations, further serving the previous two areas. It is not difficult to see that crowd simulation plays a crucial role in these scenarios, and making the generated virtual crowds as realistic as possible is the most important goal of crowd simulation. Against this background, crowd behavior data has played an increasingly important role: from the 1990s to around 2015, mainstream methods collected crowd behavior and conducted mathematical modeling, forming models that cover dynamic characteristics of crowd movements, choice behaviors of crowds in emergency evacuations, etc., and used them to simulate the behavior of virtual crowds. On the other hand, these knowledge-based methods often evaluate the realism of simulations by comparing them with real crowd data when simulating crowd behavior in different scenarios. Since the mid-2010s, simulation methods aimed at replicating the characteristics of real data have emerged, including methods that search and match from real datasets

in generating virtual crowd behavior, as well as later more popular statistical learning-based deep neural network methods. Although these methods no longer strive to acquire knowledge from crowd data, they are equally dependent on crowd data, especially high-quality data.

So far, most of the crowd data either comes from controlled experiments or field observations. However, these methods face significant limitations, including not only the technical challenges of collecting high-quality data but also ethical constraints such as not harming the subjects. These difficulties result in limited coverage of scenarios by the data and a lack of understanding of individual behavior details. From a microscopic perspective, crowd movement is a combination of multiple individual movements, where the crowd members interact locally with the others. These local interactions are rich in low-level behaviors such as observing (eye gaze or stare), communicating (verbal or non-verbal), and avoiding (full-body motion or partial-body motion). Such low-level behaviors are valuable to understanding crowd movements, and can be important elements regarding the perceived realism of crowds. However, most existing datasets only cover pedestrian trajectory information, and are unable to reflect the details of the complex behaviors at the individual level. In the journey of researchers continuously pursuing high-fidelity crowd simulation, the lack of such data means that it is impossible to simulate similar behaviors.

Recently, with the implementation of virtual reality technology, more and more crowd studies have chosen to conduct experiments in virtual environments to study pedestrian behavior. Moreover, with the continuous maturation of behavior capture devices accompanying VR (including gaze, body, hand, facial expression tracking), it has become possible to collect real user and crowd interaction behaviors in virtual environments. VR social experience platforms, such as VR chat and Ubiq [Friston et al. 2021], ideally allow multiple real users to participate in a VR scene simultaneously for experiments, thus obtaining high-quality behavior data for each user. However, the participation of multiple subjects implies the need for multiple sets of virtual reality equipment, as well as overcoming technical challenges such as low-latency multi-terminal data synchronization. Therefore, mainstream VR-based crowd research still designs experiments to involve only one user at a time. As a result, to build a crowd environment in the virtual world, most studies usually rely on crowd simulation to construct virtual crowds. However, as we mentioned earlier, current crowd simulation technology is still unable to simulate complex and realistic human-like pedestrian behavior in virtual crowds, and what limits the development

of simulation methods in this direction is precisely the lack of data. It is not difficult to imagine the dilemma brought about by this chicken-or-egg problem: if the behavior of the simulated virtual crowd is too unrealistic, then the interaction between the human user and the virtual crowd will also become unrealistic. And if the data collected is not realistic enough, how can we build a realistic simulation based on these data?

## 1.2 Objective & Challenges

Considering these circumstances, we believe it is crucial to propose a quick and efficient method for capturing crowd motion data. This method should finely record the dynamics and behaviors of each pedestrian while moving in a crowd and be easily upgradeable to accommodate potential future data capture requirements. On the other hand, we hope that the data collected with this method all comes from real actions and behaviors of human users. For this purpose, we consider using virtual reality to embed human users in a virtual environment and then collect the behaviors of human users in the real world using wearable motion capture devices. Admittedly, similar solutions exist in reality that simply use wearable motion capture devices to track and record user behavior. However, VR has gradually shown three major advantages from past research: First, experimenters have high control over the virtual environment, allowing for easy modification of experimental conditions in the virtual environment, thus efficiently completing experimental preparations. At the same time, with the help of simulation and computer graphics technology, it is easy to construct scenarios in the virtual environment that are difficult to create in real-world crowd experiments (such as flames) without causing direct physical harm to subjects and experimenters. Second, VR experimental scenarios have a high degree of repeatability, allowing experimenters to present completely consistent experimental situations to different subjects, such as having virtual crowds behave identically in front of different subjects, or precisely simulating the same flames at the same time and location. In contrast, due to the high complexity of crowd movement, even given the same initial conditions, there is no way to ensure that a real crowd would precisely perform the same actions in two experiments. Third, VR scenarios allow for more efficient data collection and processing. Experimenters have complete information about the virtual scene and can calculate the objects of the user's interaction behaviors in real-time with motion capture technology complementing VR devices. For example, if eye tracking is done in a virtual environment, experimenters can not only know what kind of gaze the user has performed

but can also calculate what virtual objects or characters the user looked at.

However, using VR for crowd motion capture faces challenges. First, past research has revealed that, influenced by VR devices, user behavior and perception in virtual environments are not the same than in the real world [Feng et al. 2021]. Secondly, conducting motion capture of crowds in VR does not mean simply applying the VR motion capture solution for a single user to multiple users: doing so first requires the purchase of a large amount of hardware equipment for motion capturing multiple users, and also means reconstructing each user’s real state in the virtual environment in real-time, which poses challenges for data transmission stability. Not only that, but to avoid collisions in the real world for users wearing HMDs and to avoid complex signal interference caused by excessive equipment, larger or even multiple dispersed experimental venues are needed to accommodate these users, subsequently requiring a larger technical team to maintain and ensure the smooth conduct of experiments, ultimately leading to escalating experimental costs.

Considering the opportunities and challenges brought by VR, this thesis attempts to explore a method that allows us to collect crowd motion data while ensuring that this method is cost-efficient. Given that mainstream VR experiments only involve a single user, we hope to complete the motion capture of a crowd with just one user. In this way, future research will have the opportunity to use the experimenter’s existing equipment to build the required crowd dataset using this crowd motion capture paradigm. At the same time, based on the implementation of such a method, we also need to verify that the dataset collected through this method is valuable: obviously, the individual data from this method must be human-like because the data of each pedestrian comes from real human motion capture. However, at the collective level, we need to ensure that the data obtained from this method shows similarity to real-world data. To our knowledge, so far, there is no similar VR motion-capture-based crowd data, nor any corresponding evaluation methods. This brings the second challenge to our work, which is to validate our method through appropriate means.

### 1.3 Approach

To achieve the goal of capturing crowd motion data with a single user, we restructure the objective of “capturing multiple pedestrians simultaneously” into “capturing a single pedestrian multiple times” and then merge the data of each individual pedestrian together

to form a complete crowd motion dataset. To implement this process, we draw on a record and replay approach. This approach is similar to one actor playing multiple roles in film and television, where an actor first records the behavior of one character, then reloads the recorded data into the virtual scene, and acts out another character in the virtual world to interact with the previously recorded behavior, thus constructing the interaction between the two characters. The technical characteristics of virtual reality make this record and replay process easy to implement, and it has been envisioned and applied in past research [Slater et al. 1995; Osimo et al. 2015]. In this thesis, we embody a user in a virtual avatar in a virtual scene using an HMD and body motion capture system and let the user generate the movement of a pedestrian in a virtual crowd by walking in the real world. During this process, we record the user’s behavior. After this process is completed, we reset the time in the virtual scene and use a virtual character to display the user’s recently recorded motion. Meanwhile, the user is embodied in another virtual avatar to act out another virtual pedestrian. This process is repeated so that the user can iteratively act and generate walking behaviors for different virtual characters until a complete crowd is captured. We refer to this paradigm as the One-Man-Crowd (OMC) paradigm.

At this stage, a distinctive feature of the OMC paradigm is its avoidance of any simulation techniques during the crowd data generation process. This approach primarily benefits from eliminating artifacts that might arise from the lower realism of simulation methods. However, this complete exclusion of simulation limits our ability to fully utilize the advantages typically gleaned from virtual reality environments in other VR-based crowd studies.

To address this, we have explored several variants of the OMC paradigm. In these variants, we employ crowd simulation techniques to construct reactive virtual crowds, providing contextual scenarios for the user’s data generation. It is important to note that these simulated crowd motions are used solely for situational creation and are not preserved in the final dataset. By incorporating these methods, we investigate the performance of the OMC approach in multidirectional crowd scenarios, balancing the need for contextual realism with the paradigm’s core principles.

To evaluate the realism of the data generated by our method, we adopt evaluation techniques commonly used in other VR-based crowd studies. A key part of our assessment involves comparing the data produced by our approach with crowd data obtained from real-world experiments. Given the scarcity of existing crowd datasets that include detailed pedestrian motion, our focus primarily centers on evaluating pedestrian trajectories.



## 1.4 Contributions

This thesis proposed two main contributions based on two experimental studies focusing on exploring our novel crowd motion capture paradigm by leveraging virtual reality to provide high-quality crowd data.

In the first contribution, we proposed a novel VR-based crowd motion capture paradigm to allow generating a set of crowd motions with only one single user. We refer to this concept as the One-Man-Crowd paradigm. Based on a Record-and-Replay approach, we subdivided the capture of multiple pedestrians in a crowd into multiple subsequent iterations of capturing a single pedestrian. Using the ability of data recording and animation visualization of Virtual Reality, we immersed a user into the virtual world, and subsequently recorded the user’s motion during one iteration of the capture, and animated the captured motion in the same virtual scene when the user performed the subsequent captures. This process finally resulted in a set of crowd data. We tested the paradigm with three experiments that replicated three crowd controlled experiments conducted in real-world laboratory conditions. Through comparisons between our data and the real-world data, we observed promising qualitative similarities in terms of the emergent patterns. Despite the presence of quantitative differences, our results validated the use of the One-Man-Crowd paradigm to capture crowd data that presents realistic individual and collective behaviors. This work was presented at IEEE VR 2022 conference and has been accepted as a TVCG journal article (nominee of the best journal paper), and is presented in Chapter 3.

Since the One-Man-Crowd paradigm involves generating motion of different pedestrians by a single user, its Record-and-Replay based nature leads to two concerns that can bias the user’s behavior. First, the user has to initiate the capture process in an empty virtual scene, and will probably perform movements that present no interaction with the further generated characters. Second, because the generated characters are programmed to display only the recorded movements instead of being interactive during their movements, the paradigm imposes the user to perform and will also gradually increase the difficulty of collision avoidance, leading to unavoidable interpenetrations between the (virtual avatar of) user and the characters. To address these two limitations, we proposed two improved paradigms that respectively introduce a contextual crowd at the beginning of the capture process to trigger interaction behavior, and enhance the interactivity of generated characters by using crowd simulation methods. We conducted two experiments that replicated real-world multidirectional crowd experiments in a laboratory condition.

Our results suggest that the behaviors observed in VR users with surrounding agents from the beginning of the recording process are made much more natural, enabling 3R or 4R paradigms to improve the consistency of captured crowd datasets. This work was presented at IEEE VR 2024 and has been accepted as a TVCG journal paper (honorable mentioned for the best papers). This work is detailed in Chapter 4.

## **1.5 Overview**

This thesis is structured as follows: Chapter 2 surveys the state of the art in crowd study, discussing real-world and virtual data collection and simulation methods. Chapter 3 introduces our first contribution, the One-Man-Crowd paradigm. Chapter 4 introduces our second contribution, where we used and examined the effect of contextual and responsive crowds to improve VR-based crowd capture. The thesis concludes with Chapter 5, summarizing the insights and suggesting directions for future research.

# STATE OF THE ARTS

---

In this chapter, we first review the current state of crowd studies in the real world, followed by an exploration of crowd studies in virtual environments. The study of crowd behavior data primarily relies on data obtained from the real world. However, whether through controlled experiments or field observations, the collection and reconstruction of crowd data face various challenges. In recent years, an increasing number of studies have begun using virtual reality technology to investigate human behavior in different crowd settings. To construct crowds in virtual environments, these studies heavily depend on crowd simulation methods. This chapter will also discuss the interrelationship between crowd simulation and crowd data.

## 2.1 Crowd Study & Data Collection In Real World

The study of crowds is fundamentally dependent on observing and collecting data relative to crowd behavior. Recently, many studies have been conducted on different research topics, observing and collecting motion data of crowds in different contexts in order to understand and investigate crowd behavior. Broadly speaking, research on crowds can be divided into two main areas: crowd behavior in normal situations and crowd behavior in abnormal or emergency situations [Feng et al. 2021; Haghani and Sarvi 2018]. The latter includes primarily the study of crowd behavior during evacuations, and also includes some research focused on crowd disasters that have resulted in significant casualties.

The studies that focused on crowd behavior under normal situations cover numerous aspects of crowd motion concerning different scales of crowd movements. This includes: 1) individual pedestrian behavior such as collision avoidance [Paris et al. 2007], 2) behavior of small groups and the influence of group on pedestrian behavior [Moussaïd et al. 2010], 3) statistical characteristics featured by a moving crowd (also known as macroscopic and microscopic characteristics of a crowd) [Seyfried et al. 2005b; Seyfried et al. 2009; Seyfried et al. 2010a], 4) the collective patterns resulting from the combination of the local behavior

of pedestrians (as known as emergent patterns) [Shi et al. 2018]. Apart from these scale-dependent crowd studies, other studies have addressed how social experience influences pedestrian behavior [Nilsson and Johansson 2009; Sieben et al. 2017]. In general, studies focusing on crowds in normal situations are dedicated to obtaining knowledge about crowd movement and to better describing crowd movements through models.

Meanwhile, the studies that focused on crowd behavior under abnormal situations are usually directly linked to public safety and thus have different concerns. By abnormal situations, an important part of research has focused on different scenarios of crowd evacuation. While some review articles [Haghani and Sarvi 2018; Feng et al. 2021] provide fine-grained categorizations based on the measured variables, we, for the sake of simplicity, distinguish only 2 broad types of studies based on their high-level research focuses, that is, choice behavior during evacuation [Haghani and Sarvi 2016; Fang et al. 2010] and evacuation efficiency [Ding et al. 2020; Daamen and Hoogendoorn 2010]. Another important research topic is to revise and understand crowd behavior during crowd disasters [Helbing et al. 2007; Helbing and Mukerji 2012; Wang et al. 2019]. These research are highly related to fire safety, building design and safety control during massive events.

To acquire data and enable these studies, different approaches have been adopted, involving experiments with either human or animals, and other methods such as interview [Haghani and Sarvi 2018]. In the context of this thesis, we focus only on data directly reflecting human behavior. In this regard, numerous review articles [Haghani and Sarvi 2018; Haghani 2020; Feng et al. 2021] have proposed to categorize the nature of the studies and have distinguished 1) studies with a laboratory nature, also known as controlled experiments, and 2) studies with a field nature, also known as field observations. However, from a technical perspective, such a binary categorization is not sufficient to define the limits between experiments with a more complex nature, neither enough to reflect the paramount difference of the challenges the researchers had to face. Thus, we extend this categorization into a continuum from controlled experiments to field observations, notably involving **laboratory experiments** (see Section 2.1.1), **controlled field experiments** (see Section 2.1.2), **uncontrolled field experiments** (see Section 2.1.3), and **field recording** (see Section 2.1.4). In the following sections, we review the research outcomes and technical challenges of each category, and provide a summary of the advantages and disadvantages of each 2.1.5.

## 2.1.1 Laboratory Experiment

### 2.1.1.1 Research Topics

Laboratory experiments take place in temporary setups and under the influence of the experimenters' manipulation of certain design factors [Haghani and Sarvi 2018]. The laboratory environment, although not always strictly located inside a laboratory [Moussaïd et al. 2009], gives the experimenters the highest controllability not only on the design factors, but also on the data collection system (e.g., optimizing the installation of recording devices, equipping the subjects with necessary devices, etc.). This allows for high-precision data capture as well as the possibility of individual-level analyses. On the other hand, the experimental spaces are usually limited, which restricts the investigation of large-scale crowd movements. Therefore, research using laboratory-controlled experiments often aims to replicate basic and critical crowd scenarios that reflect real-world collective movements. These basic scenarios typically mimic real-world architectural elements (such as doorways, corridors, and bottlenecks) by employing simple artificial physical barriers [Seyfried et al. 2010b]. These barriers are arranged to create scenarios with specific geometric configurations. In Figure 2.1, we leverage Duives et al. [2013] and Shi et al. [2018]'s classification based on the geometric configuration and designed crowd movement features (such as the number of flows, the walking directions) to illustrate the elemental scenarios that are frequently studied. In laboratory controlled experiments, studies utilizing these scenarios have covered scales from the collective level to the individual level. In the field of collective-level crowd behavior research, much emphasis has been placed on emergent patterns – those collective behaviors that emerge from the aggregation of the local interactions of individuals. This phenomenon of emergent patterns has received considerable attention and has been a central focus of this area of study. Interestingly, emergent patterns are usually highly context-dependent. Some emergent patterns have been frequently investigated by researchers using similar scenarios. For instance, the studies investigating lane and stripe formation phenomena have shown a special attention to corridor [Kretz et al. 2006] and bottleneck [Helbing et al. 2005; Seyfried et al. 2009; Seyfried et al. 2010b] scenarios. Similarly, the stop-and-go phenomenon is usually studied in a single-file scenario [Lemercier et al. 2012], which allows reducing the design factors and provides a clearer measure of the variables of interests. The doorway scenario [Garcimartín et al. 2014], and other scenarios with similar geometric shapes, are often used to study the faster-is-slower phenomenon for the understanding of stressful crowd evacuation.

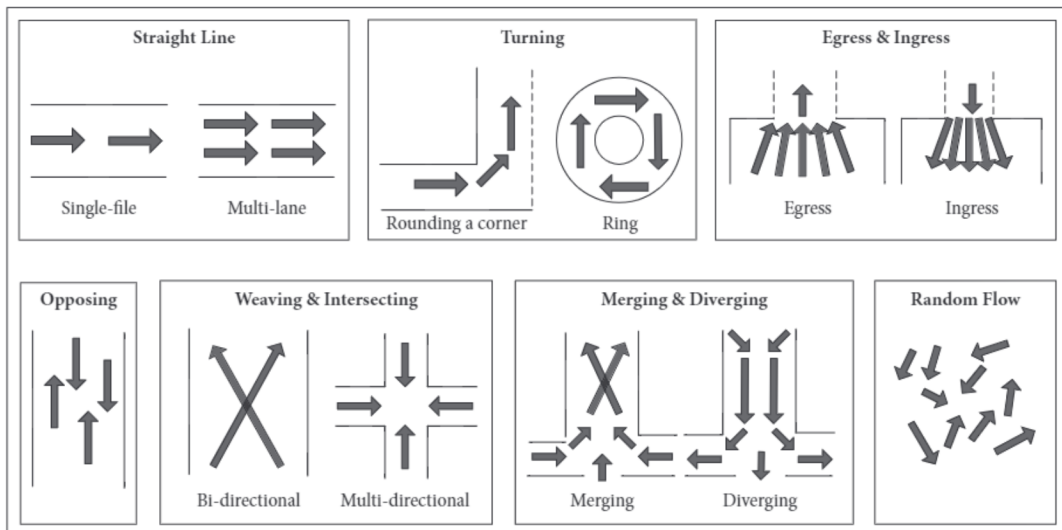
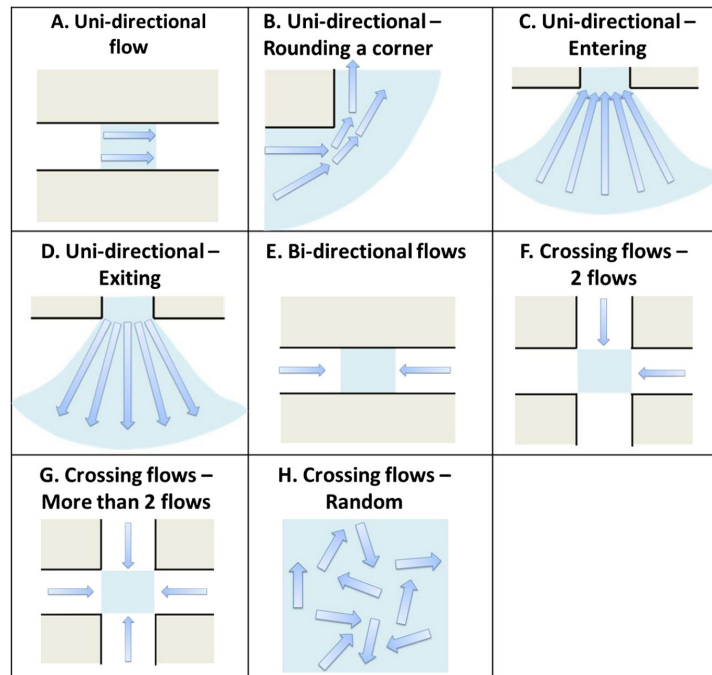


Figure 2.1 – The two different classifications of crowd movements provided by Duives et al. [2013](top) and Shi et al. [2018](bottom)

Another class of studies has focused on investigating the statistical characteristics of crowd movements. This involves both the macroscopic characteristics of a crowd, such as the density and flow speed, and the microscopic characteristics featuring the behavior of pedestrians, such as velocity and walking direction. Among these studies, we highlight two concepts that have been specifically emphasized: 1) the fundamental diagram, which reflects the relationship between the speed and density of a crowd, and 2) the pedestrian flow, which is often used to describe the evacuation efficiency of bottlenecks and narrow exits. Regarding different research context, the literature presented a 1D fundamental diagram used for the single-file and unidirectional crowd movement [Ziemer et al. 2016; Lemercier et al. 2012; Cao et al. 2016; Zhang and Seyfried 2012; Paetzke et al. 2023], and a 2D fundamental diagram used for crowd scenarios that involve both forward and lateral movements such as corridor, T-junction, merging flows with different merging angles [Seyfried et al. 2005a; Ren et al. 2019; Zhang et al. 2011; Shiwakoti et al. 2015; Shahhoseini et al. 2017]. Regarding the pedestrian flow, past studies have investigated the influence of different variables (e.g., the width of the corridor and door [Kretz et al. 2006; Tian et al. 2012], stress level [Daamen and Hoogendoorn 2010], selfish behavior [Nicolas et al. 2016]) on the flow speed.

TODO: add a figure to illustrate the 1D and 2D fundamental diagram

Some studies also focus on examining the connection between individual behavior and collective behavior, leading to the investigation of the way-finding behavior and navigational decision behavior [Haghani and Sarvi 2018]. These studies implicitly or explicitly consider psychological effects in crowd movement [Bode et al. 2015a; Konya and Sieben 2023]. Recent research pays more attention to the integration of these behaviors with crowd evacuation, exploring how a crowd’s exit choice during the evacuation process is influenced by various factors. These factors include herding effects [Haghani and Sarvi 2017b], queuing behaviors [Wagoum et al. 2017], and the trade-offs made by pedestrians during exit decisions [Liao et al. 2017].

On a smaller scale, some attempts to capture and model collision avoidance behaviors have been done. This class of studies focuses on the individual level behavior and seeks to explain how pedestrians adjust their velocity to avoid potential collisions with other participants. The experiments usually rely on multidirectional crowd scenarios, and are designed to create unavoidable interactions between pedestrians [Paris et al. 2007; Moussaïd et al. 2009; Huber et al. 2014; Parisi et al. 2016]. These studies addressed various aspects of the collision avoidance behavior, including side preference [Moussaïd et al.

2009], crossing angle [Asano et al. 2010; Daamen et al. 2014]. The experimental outcomes are frequently linked to the validation or calibration of relative crowd simulation (see developed discussion in Section 2.3).

### 2.1.1.2 Data Collection Methods

Most of the aforementioned studies go through data recording, data extraction, and data analysis.

- Data recording: This process requires experimenters to set up or install data capture equipment according to the site and purpose of the experiment. The vast majority of studies use video cameras to capture experimental procedures, which means that experimenters need to choose where to mount the cameras in advance (i.e., mounting the cameras on the ceiling to acquire an aerial view, or setting up cameras on the side of the crowds to acquire a horizontal view) Many experiments used multiple cameras to capture crowd movements from different viewpoints [Seyfried et al. 2010b; Liao et al. 2014], which meant that synchronization of the different cameras needed to be done prior to the experiment. Other experiments utilize motion capture devices to achieve a more accurate capture of pedestrian motion, and motion capture devices are generally able to record data at a higher frame rate than video recording-type methods [Paris et al. 2007; Huber et al. 2014].
- Data extraction: For all studies that use video camera to record experimental procedures, crowd motion data is stored as image frames in the video, so it is indispensable to recover high-quality crowd motion data from the video. Some early experiments were conducted by manually analyzing the video and extracting the relevant features of the crowd, including density, flow, velocity, etc., through image overlay and other methods with the help of image processing tools **empty citation** Due to the low efficiency of manual labor, these methods cannot provide lower level crowd motion data like trajectory. To automate the extraction of data, part of the research thus adopts automated image processing methods (which can automatically extract crowd features from video data, but still cannot provide individual trajectory information). The recent mainstream data extraction method is the pedestrian trajectory extraction based on color recognition. in the context of laboratory experiment, the experimenter has a high degree of controllability for both the experimental process and the subjects. In the context of laboratory experi-



ments, experimenters have higher controllability on both the experiment process and the subjects, who can be asked to wear a hat or marker with a special color before the experiment starts, used to detect each pedestrian by color detection in the recorded video, so as to achieve the recovery of the subject’s trajectory with a high precision. Early experiments using this method usually required the development of specialized software to realize tracking [Daamen and Hoogendoorn 2003b]. Later, Boltjes et al. [2010] proposed PeTrack, a software for automated pedestrian detection widely used in subsequent studies [Zhang and Seyfried 2012; Liao et al. 2014; Ma et al. 2020; Konya and Sieben 2023]. In some studies focusing on individual behavior [Lemercier et al. 2012], experimenters also used additional markers (usually located on the shoulder) to further improve the accuracy of tracking and data extraction. For data obtained from motion capture systems, the data extraction solution is usually pre-implemented by the manufacturer of the equipment, thus eliminating the need for the researcher to perform data extraction on their own, even though the process might require intensive manual pre-processing of the data (e.g., cleaning, labeling).

## 2.1.2 Controlled Field Experiments

Despite the large number of studies and data captured in pure laboratory conditions, studied walking scenarios are usually limited to the standardized and simplified situations, and sometimes are not sufficient to simulate or get close to real-world situations. One important reason for this is that the real-world crowd movements usually happen within specific scenes involving unique parameters, that are difficult, sometimes infeasible, to replicate within the laboratory conditions. To overcome this limitation, some research conducted experiments as well as crowd motion capture in special scenes such as classrooms [Guo et al. 2012; Zhu and Shi 2016], stadiums [Burghardt et al. 2013], supermarkets [Cao et al. 2018], and theaters [D’Orazio et al. 2016] to collect more authentic data of crowd movements, while still keeping these experiments under the control of experimenters.

### 2.1.2.1 Research Topics

Controlled field experiments usually involve scenes on a larger scale than laboratory experiments. Both the factor design and data recording are thus restricted by the physical

condition of the experimental scenes. As a result, the studies involving controlled field experiments focus less on the inter-individual interaction. On the other hand, since the experiments are conducted in real-life buildings, these studies inherently possess a nature of crowd evacuation.

Some experiments continued the idea of the laboratory experiment, focusing on the collective level of crowd behavior, especially the macro- and microscopic characteristics of crowds. For example, Huo et al. [2016] investigated how pedestrian flow is influenced in the case of staircases, and Burghardt et al. [2013] investigated the effect of bottleneck on crowd flow by using the stadium grandstand structure, and compared the data collected in experimental conditions with those of crowd in real life-situation by using the results of field observation in the same venue.

The majority of the experiments utilized real-world sites to study and test the performance of the pedestrians in the evacuation, and the effects of different factors on the pedestrians' decision. An important class of studies also focused on simulating different visibility conditions during evacuation to address crowd evacuation performance during emergencies like fire. These studies so far investigated macroscopic characteristics related to pedestrian flow (e.g., flow speed, evacuation time) [Cao et al. 2018; Guo et al. 2012; Zhu and Shi 2016], pedestrian way-finding and exit choice [Guo et al. 2012; Jeon et al. 2011; Porzycki et al. 2018; Ronchi et al. 2017]. Other studies examined how external factors influence the evacuation efficiency of the crowds. By external factors, we refer to the information that the pedestrians can perceive from either the environment or the others. For the former case, a main research direction focuses on how pedestrians perform way-finding based on the evacuation installation that provides instructive information (e.g., signs, sounds, lights) to assist the evacuation [Fridolf et al. 2013; Galea et al. 2017; D'Orazio et al. 2016; Jeon et al. 2011]. For the latter case, studies have focused on how social behavior can influence pedestrian decisions [Heliövaara et al. 2012; Kruchten and Schadschneider 2017; Haghani and Sarvi 2017a]

### 2.1.2.2 Data Collection Methods

In controlled field experiments, experimenters still have high controllability of the experimental scenario and subjects, although less than in the pure laboratory condition. In some cases, it is still possible to use the commonly used tracking methods described in the previous section, i.e., based on automated color recognition [Burghardt et al. 2013]. Again, this means that the experimenter needs to take the subjects through a lot of

preparation. However, in the bulk of scenarios, such experiments usually have to face three main challenges.

First, the design factor of the experiment and the means of recording the experiment are both physically limited by the experimental site, and most of the experiments are recorded by borrowing video surveillance equipment from the experimental site. Due to the location and viewing angle of these devices, it is often not possible to cover the full range of action of the subjects in larger scenarios. Some of the evacuation experiments were conducted to simulate the restricted visibility by turning off the site lighting [Jeon et al. 2011] or using artificial smoke [Fridolf et al. 2013], which limited the visibility of the pedestrians as well as the visibility of the normal video camera, and thus the experimental process and the path of the pedestrians could not be recorded using conventional methods. With the site lighting turned off, Jeon et al. [2011] equipped each subject with a GPS tracker to record their walking paths during the experiment. In contrast, in scenarios where artificial smoke is used, a common practice is to use a thermal imager to videotape subjects in the smoke. For example, Ronchi et al. [2017] arranged for a professional firefighter to follow the subjects with a thermal imager outside their field of view in an experiment using artificial smoke, which meant that the experimental recording process had a higher degree of complexity

Second, more constraints on the time of day to conduct experiments, especially those conducted in commercial buildings, can lead to experiments being scheduled at extreme times. For example, in the work of Jeon et al. [2011] a three-day experiment was scheduled between 00:00 and 05:30, which increased costs for both the experimenter and the subjects, and posed a higher degree of difficulty for the organization of the experiment.

Finally, compared to the experiments conducted in a pure laboratory condition, the controlled field experiments have a higher risk of personal injury compared to experiments conducted under purely laboratory conditions. This risk further limits the scope of experiment selection. In particular, experiments that alter visibility conditions (i.e., turning off lighting and artificial smoke) are typically more focused on the behavior of a single pedestrian. It is not difficult to envision that increasing the number of subjects simultaneously present in the scenario would further raise the risk factor of the experiment, and become the ethically more complex. While this limits the scope of research topics, it also limits the ability of such experiments to simulate more realistically real-life conditions in emergency evacuation scenarios, and the results obtained may systematically deviate from real situations.

For the above reasons, controlled filed experiments usually do not focus on high-precision pedestrian behavior such as individual trajectories, let alone lower-level individual behavior. The scope of studying and describing crowd behavior is usually limited to some high-level metrics such as flow rate, flow speed and evacuation time (in the context of studying evacuation efficiency), or tabular statistics of path and exit choices. (in the context of choice behavior study). Due to potential security risks, it is difficult for emergency and evacuation studies to simulate real-world situations with a high degree of realism.

### 2.1.3 Uncontrolled Field Experiment

The uncontrolled field experiment corresponds to the class of crowd related experiments that are not in the control of experimenters. In this thesis, we follow one criterion to distinguish the uncontrolled experiment from the controlled experiment: whether the subjects were aware that they were going through an experiment and their behavior would be recorded for one or multiple analyses in a specific context of crowd study. If not, then the experiment is categorized as an uncontrolled experiment. Since the participants are not aware they of the experiment, their behaviors are more natural, leading to crowd data of higher realism. As a trade-off, researchers cannot accurately obtain the subjects' demographic information, nor can they precisely identify or isolate all the factors that influence pedestrian behavior.

#### 2.1.3.1 Research Topics

Under the subcategory of uncontrolled field experiments, the vast majority of studies are based on analyses related to the recordings of evacuation drills, particularly unannounced evacuation drills in crowded buildings or facilities. In this regard, the uncontrolled field experiments have addressed similar research topics as the controlled field experiments, particularly, the pre-evacuation time, evacuation time, evacuation speed, route choice, and exit choice. The investigated scenarios have covered retail store [Shields and Boyce 2000], cinema theater [Nilsson and Johansson 2009; Lovreglio et al. 2015], teaching building [Fang et al. 2010], hotel [Kobes et al. 2010b], underground metro train [Fridolf et al. 2016].

As the evacuation drills were unannounced and were conducted in crowded places, the evacuations inherently involved high density crowds and the analyses are usually

performed on the collective level. Although rare, we remark a specific type of uncontrolled experiment, where the experimenter recruited participants for a decoy research topic while hiding the real experimental motivation. In an example study, Kobes et al. [2010b] settled participants in a hotel by convincing them that staying overnight was necessary to finish the decoy experiment. During the night, the experimenter informed the participants that a fire was going on, and they were required to evacuate from their room. One participant was informed at a time to observe the single participant’s way-finding behavior during the evacuation.

### 2.1.3.2 Data Collection Methods

Similar to our discussion in Section 2.1.2.2, the methods for observing and recording crowd behavior are limited by site conditions. Unlike controlled experiments, in the context of uncontrolled field experiments, experimenters cannot use any auxiliary tools (such as markers or trackers) to assist data recording. The data recording methods available to experimenters are largely limited to using video cameras at the experimental site to record parts of the experiment. Therefore, as with most experiments in Section 2.1.2, the data analysis of these uncontrolled field experiments also mostly remains at the level of statistical data analysis. At the same time, the observed pedestrians are in a non-experimental behavior mode, so the recorded data shows richer patterns and behaviors, such as groups of friends, groups of couples, etc. Although this enhances the realism level of the recorded data, it also poses more challenges for data analysis and processing. In addition, since the subjects are not in an experimental state, they may exhibit behaviors beyond the expectations of the experimenters, thereby affecting data collection. For example, in the work of Kobes et al. [2010b], the authors had to exclude some participants’ results due to unexpected circumstances. For instance, when one participant received a fake fire alert, three other participants were in the same room. Also, several other participants informed nearby people after receiving the fake fire alert. It is not difficult to imagine that this method of disguising the true purpose of the research may face more uncontrollable factors when organizing larger-scale experiments.

To supplement the analysis of experiments based on video data, some studies conducted interviews with participants after evacuation experiments to gather their psychological experiences. This approach aligns more closely with the research methods based on surveys or interviews as categorized by Haghani and Sarvi [2018], as well as Fang et al. [Feng et al. 2021] in their reviews. Considering our focus is elsewhere, we will not delve

deeply into these methodologies.

## 2.1.4 Field Recording

### 2.1.4.1 Research Topics

In the field nature of crowd data collection, an essential data resource is the crowd behavior recorded directly from everyday life. This data collection process does not involve any form of experimentation, and thus, we categorize this method of directly observing and gathering data from daily life as field recording, distinguishing it from other experimental methods that also possess a field nature. Numerous studies have utilized these captured data to explore crowd behavior in the real world's daily life. These studies encompass various scales of crowds, ranging from small groups of two to three people, everyday scenes with medium to low-density crowds, high-density crowd scenes, and massive events, up to the crowd movement on the scale of streets or even entire cities.

One class of studies has focused on social group behavior containing 2-5 people. These studies have investigated the walking behavior (involving speed, step frequency, group component) [Feng and Li 2016; « Improved Social Force Models Considering Heterogenous Characteristics among Social Groups » 2018] and shape patterns [Moussaïd et al. 2010], group splitting [Do et al. 2016], and the influence of group size [Duives et al. 2014]

Another class of studies has focused on estimating statistical characteristics of crowd, especially the pedestrian flow behavior in the real-world situation, involving variables such as density-speed and density-flow volumes [Fruin 1971b]. Numerous studies have covered different pedestrian walking dynamics on walkways [Virkler and Elayadath 1994; Lam et al. 1995; Corbetta et al. 2016], sidewalks [Tanaboriboon et al. 1986; Al-Azzawi and Raeside 2007], and stairways [Shah et al. 2013; Tanaboriboon and Guyano 1991].

Another main research interest in the context of field study is observing and analyzing dense crowd behavior during mass events such as festivals [Zhang et al. 2013; Ma et al. 2013; Larsson et al. 2021]. The observation of mass events can lead to various research topics, involving both statistical characteristics for crowds with a high density level [Johansson et al. 2008], and the collective emergent pattern such as stop-and-go waves [Zhang et al. 2012], lane formation [Zeng et al. 2014].

A set of research has also addressed crowd disasters, a research topic that cannot be simulated through human experiment, using video recordings from the surveillance cameras during the disasters. Such studies, although not exhaustive, revised disasters

related to high crowd density, have recovered crowd the collective patterns during the disasters [Helbing et al. 2007; Krausz and Bauckhage 2012] as well as the causality of crowd disasters [Helbing and Mukerji 2012]. Some other studies focused on the emergent crowd evacuation performance during natural disasters, such as earthquake [Yang et al. 2011b; D’Orazio et al. 2014; Bernardini et al. 2016; Shiwakoti 2016]. Similarly, crowd evacuation behavior under terrorist attack has also been revised [Wang et al. 2019].

#### 2.1.4.2 Data Collection Methods

Field recordings have collected a wealth of valuable crowd behavior data from daily life, most of which is obtained in the form of video recordings. Early studies analyzed crowd behavior in video recordings through manual analysis [Virkler and Elayadath 1994; Lam et al. 1995]. Subsequently, some research began to use image processing technology to directly calculate the macroscopic characteristics of crowds from video data, such as density and flow speed [Zhang et al. 2013]. Numerous studies have been dedicated to exploring computer vision techniques to achieve automated detection and tracking of pedestrians, a task also known as Multiple Pedestrian Tracking (MPT) [Sun et al. 2021]. This has evolved into two different technological paths: Model-free Tracking (MFT) and Tracking-by-Detection (TBD). MFT requires manually marking the tracking targets in the initial frame of the video [Feng and Li 2016], i.e., manually performing pedestrian detection, while TBD further achieves the automation of pedestrian detection. Recently, TBD has made several milestone developments in research [Singh et al. 2008; Zhang et al. 2008; Shafique et al. 2008; Yang et al. 2011a; Andriyenko and Schindler 2011]. With the rise of deep learning methods, the latest research attempts to use deep neural networks to achieve more efficient detection and tracking [Kim et al. 2015; Tang et al. 2017; Xu et al. 2017].

However, due to limitations in video camera imaging quality and occlusions in high-density crowd scenarios, achieving high-quality pedestrian detection and stable tracking remains a significant challenge in the field of crowd data reconstruction. To overcome these challenges, some researchers have turned to other types of image acquisition devices, such as infrared counters [Zhang et al. 2013] and depth cameras [Corbetta et al. 2016], to simplify subsequent analysis work. Moreover, other studies have utilized multimodal signal data (such as WiFi and Bluetooth signal transmission devices, GPS trackers, and even GSM network base stations) to track and reconstruct crowd movement, thereby achieving crowd data collection on a larger scale (up to the city level).

### 2.1.5 Summary

In this section, we review how past studies have captured real-world crowd movement data. These research efforts have obtained valuable data for understanding and modeling crowd behavior through two distinct approaches: controlled experiments and field observations. However, constructing these datasets often entails considerable costs. In controlled experiments, researchers typically invest significant time and effort in organizing participants and conducting experiments. This includes submitting applications to ethics committees for experimental permission, recruiting numerous participants (some studies even involved up to 1,000 participants [Ziemer et al. 2016]), equipping these participants with tracking garments or devices, and involving professionals and specialized equipment in constructing certain scenarios. Some experiments even pose risks of physical harm to participants. Consequently, the duration of data capture and post-processing in controlled experiments usually spans years.

In field observations, past research has explored various methods for extracting pedestrian dynamics data from video footage. However, the quality of data retrievable from field video is limited due to challenges like lighting conditions, camera angles, and obstruction among pedestrians. As a result, high-quality crowd motion data remains scarce. With the recent surge in deep learning and generative models, an increasing number of studies have started to focus on human body motion [Feldmann and Adrian 2023a; Boomers et al. 2023]. Among existing crowd data, few datasets provide data beyond trajectories, such as eye gaze data and body motion data of pedestrians. Overall, the current crowd datasets are still very limited and struggle to meet the growing demand for richer data in related research fields.

## 2.2 Crowd Study & Data Collection In Virtual Environment

In the field of crowd studies, with the proliferation of Virtual Reality (VR) technology, virtual experimental environments constructed using VR devices are increasingly favored by researchers due to their low-cost and high flexibility. These environments not only allow for the free setup of virtual objects and characters to interact with human users, but also provide immersive experiences while reducing the risk of physical harm associated with real-world experiments. However, the physical limitations of VR technology may lead to



differences in the behavior of human users in virtual environments compared to the real world. This section will focus on VR-based crowd experiments and their impact on human user behavior.

### 2.2.1 Crowd Related Research Topics

Similar to crowd studies in the real world, the VR-based studies mainly focus on the human user's dynamic and choice behavior. However, a remarkable difference is that most of the VR-based studies only capture a single user's behavior due to equipment limitations. Despite the limitation on the number of simultaneous participation, VR provides rich data capture solutions covering the user's trajectory, body motion, gaze, facial expression, voice, and sometimes interaction with the virtual environment. As a result, the VR-based crowd research tends and is able to pay more attention to the user's low-level behavior. Thus, in addition to the three-level categorization (operational level, tactical level, and strategical level) of crowd studies proposed by [Hoogendoorn and Bovy 2004], the VR-based research has also addressed a sub-operational level concerning those low-level behaviors of human behavior under various situations. Another important difference is that, because of the high controllability in the experimental design, numerous VR-based experiences have created rich plots to both enhance the immersion and create unexpected situations for the participants during the experiences.

The tactical and strategical level of pedestrian behavior, has been subjected to many research interests. Multiple experiments have been conducted to study behavior during evacuation, covering how the pedestrian behavior can be influenced by different factors. One important research focus is the social influence [Kinaterder et al. 2014a; Kinaterder et al. 2014b; Kinaterder and Warren 2016; Ríos et al. 2018] on the pedestrian choice behavior. These studies usually created virtual confederates by using crowd simulation. The confederates repeated the same movement or strategy in front of different participants during the experiments. These studies have led to observations that the subjects' choices can be influenced by the virtual agents, and that the emergence of herd behavior is linked to both environmental and personal factors [Lovreglio et al. 2016], as well as the uncertainty in the escape environment [Haghani and Sarvi 2017a]. Some other studies took advantage of the Virtual Environment's high controllability to investigate the influences of environmental factors. Many studies leveraged fire simulation to create emergency stimuli [Kinaterder and Warren 2016; Cao et al. 2019a] and to control the visibility during emergency [Kinaterder et al. 2014a], while some other studies chose to set up

different experimental conditions. For instance, Tang et al. [2009] investigated the impact of mistaken signage on the way-finding tasks, and Kinateder et al. [2019] revised the color design of emergency exit. Another class of studies investigated how the level of familiarity with the virtual environment can influence the choice behavior [Lin et al. 2020]. To let the participants get familiar with the virtual scene, a common method is to assign them some plot-based tasks (which also hide the real purposes of experiments), and trigger an emergency in the virtual world once certain conditions are fulfilled.

Research at the operational level of pedestrian behavior focuses on the interaction behaviors among pedestrians. This primarily includes collision avoidance behavior with obstacles and other virtual pedestrians, as well as following behavior. This covers topics similar to those found in real-world studies, such as inter-individual level interaction, where the captured user interacts with only one obstacle or one virtual character [Argelaguet et al. 2015; Nelson et al. 2023]. Due to the limitation of having only a single user in experiments, scenarios involving multidirectional flows are rare. However, by reducing the focused flow to just one pedestrian, some studies have utilized crowd simulation to conduct research on individual-crowd level interactions, which are less common in real-world experiments. For example, Bruneau et al. [2015] investigated whether the subject would choose to go through or go around a virtual crowd under different conditions, Nelson et al.; Koilias et al. [2019; 2020b] examined how the subject’s walking speed can be influenced by the virtual crowd’s speed, density, and direction.

With the maturation of capture technology accompanying virtual reality (VR), some recent studies have started using VR to delve deeper into the behavioral details of individuals within virtual crowds. Berton et al. [2020b] studied pedestrian gaze behavior during walking in the Virtual Environment, and compared with the real-world results. Raimbaud et al.; Raimbaud et al. [2022; 2023] investigated the stare interaction between a human subject and virtual crowds. Jovane et al. [2022] addressed the question of how can virtual characters attract the subjects’ attention. Some other studies involved the capture of the subjects’ full-body motion [Berton et al. 2020a; Koilias et al. 2020a]. These studies can be viewed as an extension of research on operational level behavior, forming a new category, namely the sub-operational level, which was absent in previous real-world experiments, constrained by the challenges of capture technology and unable to cover such detailed levels. Although such studies are still relatively few, they reflect a trend in future research developments.

Apart from the above-mentioned studies involving only a single subject at a time, we

highlight that, although rare, some studies have conducted multi-user VR experiments to further investigate the herd behavior [Berg 2016; Moussaïd et al. 2016]. The multi-user approach allows all pedestrians in a virtual scene to be controlled by real people, thus exhibiting more realistic behavior than a simulated crowd. However, this method, requiring the simultaneous participation of multiple users, complicates the design and organization of experiments. Technologically, this approach necessitates the interconnection of multiple virtual reality input devices over a network, facing additional challenges in engineering implementation and stable data transmission.

## 2.2.2 Creating Immersive Experiences

To accurately gather user behavior data, VR-based studies require participants to immerse themselves in the experimental scenarios, making decisions and choices. For instance, in researching stressful evacuation situations, researchers expect participants' behaviors to closely mirror real-life responses under escape pressure, rather than being indifferent or inappropriate due to the virtual setting. To achieve this, past VR experiments focused on enhancing immersion through measures in rendering, stimuli, and social presence. However, some experiments classified as VR-based studies in review articles such as Haghani and Sarvi; Haghani; Feng et al. [2018; 2020; 2021] do not involve Virtual Reality in the sense commonly understood today. This section will discuss previous efforts to build immersive experiences and the various forms of data collected through different technical means.

### 2.2.2.1 Immersive Rendering

Rendering of the Virtual Environment is crucial in VR-based crowd studies. Initially, this involved visualizing the virtual environment and crowd, with subjects viewing the virtual crowd behavior under different conditions on a screen and answering questions, like what choices they would make assuming they were in the depicted scenario [Lovreglio et al. 2014; Li et al. 2019; Bode et al. 2015b]. Some methods also present virtual environments on screens, allowing users to interact with the environment through input devices [Tang et al. 2009; Kobes et al. 2010a]. Recent mainstream approaches typically immerse subjects directly into a 3D virtual environment, employing two primary display technologies: 1) Cave Automatic Virtual Environment (CAVE), an experimental space defined by three walls, floor, and ceiling, offering immersive environments through mul-

tiple projectors. Users wear special goggles, with the system tracking these goggles and adjusting the jointly projected images to create a stereoscopic view. CAVE allows physical walking in the virtual environment but can host only one user at a time due to perspective dependency. 2) Head-Mounted Display (HMD), which creates a three-dimensional perception of the virtual scene in the user's brain by rendering images for each eye on separate screens. Thanks to built-in tracking and input solutions, HMD enables physical walking or controller-based movement. To enhance the similarity between virtual and real scenes, audio rendering, like simulating fire alarm sounds, is introduced [Kinaterder and Warren 2016; Kinaterder et al. 2018]. In VR environments, objects and virtual characters usually lack a physical entity, preventing tactile perception by human subjects. To address this, some studies have introduced haptic rendering, embedding subjects in a virtual avatar and creating artificial tactile feedback on the subject using wearable haptic devices, based on the avatar's collisions with other objects [Berton et al. 2020a; Koiliias et al. 2020a].

### 2.2.2.2 Navigation

In VR-based experiments, navigation is a crucial way for participants to explore and understand the virtual environment and characters, with trajectory data being one of the most important and common types of data collected. Research literature mainly presents two methods: virtual and physical navigation. In virtual navigation, participants use a joystick or VR controller to move in virtual space while remaining stationary in the real world. This method typically sets a uniform speed for all participants, maintaining a constant pace. Its advantage lies in allowing movement in large-scale [Natapov and Fisher-Gewirtzman 2016; Lin et al. 2020; Zhu et al. 2020] or non-flat virtual scenes [Feng et al. 2020; Lin et al. 2020; Zhu et al. 2020], but it cannot intuitively reflect the unique dynamic characteristics of participants and may lead to cyber sickness. In contrast, physical navigation requires participants to move in the real world, using tracking technology to ensure the virtual camera follows their position changes in real-time. This method records real path data, reflecting different speed preferences of participants, and reduces cyber sickness since the observed motion in the virtual world matches their physical movement. However, unlike virtual navigation, the range of movement in virtual space is identical to that in the real world, limited by the coverage of the tracking technology. Early experiments used a combination of inertial measurement units and ultrasonic ranging for positioning and tracking [Fink et al. 2007b; Kinaterder et al. 2018], while recent experiments typically utilize the tracking solutions built into commercial HMDs [Nelson et al.

2019; Berton et al. 2019; Kinateder et al. 2019], covering an area of about  $10; m \times 10; m$ .

### 2.2.2.3 Stimuli & Rewards

In most studies researching evacuation behavior, experimenters need to create different experimental conditions within the virtual environment to compare subjects' behaviors in normal and emergency states. However, since disasters in the virtual environment don't cause physical harm to subjects, they might behave differently due to perceived lower risk. To help subjects emulate real-world sensations, a common approach is to simulate fire and smoke in virtual fire scenarios [Kinateder et al. 2014a; Lin et al. 2020; Kobes et al. 2010a; Cao et al. 2019a; Zhu et al. 2020]. Questionnaire results from these studies commonly indicate that participants felt anxiety or stress in the face of virtual fires, confirming that stimuli like virtual flames or fire alarms can effectively alter emotions. Another method involves using alternative forms of reward to induce stress in participants. For example, in the study by [Moussaïd et al. 2016], the researchers incentivized participants in virtual emergency evacuations by promising monetary rewards to those who reach the exit first, thereby simulating the competitive nature of real-life emergency evacuations.

### 2.2.2.4 Virtual Crowds

In VR-based studies, as they typically accommodate only one user at a time, virtual characters are created in the virtual environment to interact with human subjects. The behavior of these virtual characters varies according to the purpose of the study. Some research employs static virtual characters [Berton et al. 2020a; Raimbaud et al. 2022], while others use dynamic, yet non-responsive characters to construct crowd dynamics in the virtual scene [Raimbaud et al. 2023]. In studies examining the influence of social factors on crowd evacuation behavior, virtual crowds are often used to simulate herd behavior to impact the subjects' decisions and actions. For research on operational-level pedestrian behavior, simulation-driven autonomous agents are commonly used, allowing virtual characters to interact bidirectionally with human subjects. For instance, Bruneau et al. [2015] utilized a modified RVO2 method, and [Koiliias et al. 2020b] employed Unity's navigation mesh. Unfortunately, many studies provide insufficient details about the behavior patterns and driving methods of virtual characters. Moreover, few studies explore the quality of animation used for these simulated characters. In psychology-related studies [Patotskaya et al. 2023] and those focusing on attention [Jovane et al. 2022], researchers directly use

virtual characters to display motion capture animations to ensure the authenticity of character animations.

### 2.2.3 Validating Data From VR Experiments

The method of using VR to study crowd behavior always faces this question: Does the behavioral data recorded in VR experiments align with people’s behavior in the real world? This question stems from two main concerns. One major concern is that the visualization of VR restricts how participants perceive others (usually simulated characters) in experiments, which can, sometimes, significantly differ from real-world crowd behavior. Another major concern is whether subjects are sufficiently engaged and motivated to perform tasks in VR as realistically and earnestly as they would in real-world scenarios. To ensure comparability between crowd behavior data collected in VR and the real world, a common approach is to conduct the same experiments in the real world as a control. If real-world experiments are not feasible, comparing and analyzing the differences between existing real-world data and data obtained from VR experiments can be considered. By following this idea, a number of different studies have examined and provided positive answers to these two concerns.

#### 2.2.3.1 VR Effect on User Behavior

When using VR to generate crowd datasets, it is of paramount importance to evaluate the similarity between user behaviors in real and virtual situations. Unfortunately, VR-based experiments have been shown to introduce a number of perceptual and behavioral biases, such as affected depth perception [Armbrüster et al. 2008; Renner et al. 2013], reduced walking speed [Banton et al. 2005; Fink et al. 2007a], increased social distance [Bailenson et al. 2001; Gérin-Lajoie et al. 2008; Bruneau et al. 2015; Llobera et al. 2010; Sanz et al. 2015], and increased eye gaze and head movements [Berton et al. 2019; Berton et al. 2020b]. Due to this, considerable work has been done to assess how these biases affect behavior in VR. The general finding is that VR motion capture leads to measurable changes in a majority of spatial variables, although the qualitative nature of participant behavior may remain consistent.

### 2.2.3.2 Interpreting Results from VR Experiments

In VR-based crowd studies, the objectives of these research efforts exhibit subtle nuances. One category of research is dedicated to replicating and studying human behavior in the real world by leveraging the advantages provided by VR, aiming to obtain as realistic data as possible. These methods face the challenge of the impact of VR equipment on human behavior, which may reduce the authenticity of the data. Therefore, the first question these studies need to address is whether the data collected still accurately reflects pedestrian behavior in the real world. An intuitive way to answer this question is to replicate experiments from the real world and compare them with the results of experiments conducted in virtual environments. For example, Kobes et al. [2010a] conducted one of the earliest validation studies to compare pedestrian evacuation behavior in a real-life hotel and a virtual hotel. Cirio et al. [2013] examined the similarity between the trajectories collected from virtual navigation and real navigation. These explorations have resulted in promising evidence suggesting the comparability between data from real-world experiment and from VR-based experiments [Haghani et al. 2017; Moussaïd et al. 2016].

Depending on the experimental method, especially in cases where some experiments are based on 2D display methods and cannot directly collect pedestrian behavior, another category of research chooses to acknowledge that the results obtained from studying crowds using VR will inevitably differ from real-world situations. Consequently, these studies propose that the data collected in experiments should be interpreted in a relative manner [Kinateder et al. 2014a], believing that these results still reflect real human behavior to some extent. In other words, researchers believe that valuable knowledge about crowd behavior can still be gleaned from these experiments.

However, a different strand of research chooses an alternative path, distinguishing the behavior exhibited by humans in virtual environments from that in the real world. These studies focus directly on how the behavior of virtual crowds affects human users, without seeking direct comparisons with real-world scenarios. Instead, this type of research is more inclined to focus on how to present crowd behavior in virtual environments that users perceive as 'believable'. Examples of such studies include [Nelson et al. 2019; Koilias et al. 2020b; Nelson et al. 2022; Trivedi and Mousas 2023]. Nevertheless, the findings from these studies can still provide insights and inspirations for the first two research approaches.

## 2.2.4 Summary

In this section, we review a series of studies focused on crowds within virtual reality environments. Recent research has employed CAVE-like devices or Head-Mounted Displays to provide participants with immersive experiences. It has been observed that although there are quantitative differences in participants' behavior in virtual environments compared to real-world settings, the experimental results obtained in virtual reality are qualitatively similar to those in real-world experiments. Conducting experiments in virtual environments reduces the need for physical space and environmental setup, allowing researchers to freely construct virtual experimental sites and scenarios. For high-risk situations in reality, such as emergency evacuation in a fire, replicating these scenarios in virtual environments and recording participant behavior is almost risk-free. Moreover, virtual reality-based experimental systems are inherently more suited for data collection. For instance, while studying gaze behavior in virtual crowds, it is possible to compute in real time the virtual characters or objects that participants focus on, significantly easing the burden of data post-processing. Additionally, due to the high controllability of virtual environments, researchers can present identical scenes and crowd movements to different participants, facilitating further precise control of experimental variables. In contrast, replicating the exact crowd movements and behaviors in real-world controlled experiments, even with confederates, is as impossible as stepping into the same river twice.

Nevertheless, virtual reality-based crowd studies face several limitations. First, discrepancies between VR display technology and human visual perception can lead to different participant behaviors in virtual environments compared to the real world. Second, the lack of physical constraints and collision feedback in virtual worlds can result in unnatural interactions between users and the virtual scenes. Third, most studies are confined to single-user experiments. Theoretically, multi-user experiments based on HMDs are feasible, but they pose higher requirements for experimental equipment, venues, and software development, significantly increasing the cost of multi-user experiments. Single-user experiments force researchers to use pre-designed virtual crowds as the interactive counterparts for human users. These pre-designed virtual characters either employ pre-recorded motion capture animations, sacrificing the characters' interactive abilities for authenticity, or rely on crowd simulation for real-time responsive virtual crowds. However, as mentioned in the section on crowd simulation, current crowd simulation realism is still limited. Consequently, virtual reality-based crowd studies urgently require more advanced crowd simulation and animation technologies to expand their application scope.



## 2.3 Crowd Simulation & Crowd Data

From the perspective of crowd study, crowd simulation has two primary application scenarios. Firstly, it is used to create or reproduce the normal behavior of real-world crowds in virtual environments; secondly, in architectural design and safety management of high-traffic scenarios, crowd simulation is utilized to simulate the behavior of evacuation. Crowd simulation can be broadly divided into two sub-categories: macroscopic simulation and microscopic simulation. Since macroscopic simulation involves modeling the crowd as a singular entity, it is beyond the scope of this thesis. We only focus on microscopic simulation, which models each individual as a separate unit [Toll and Pettré 2021].

In recent research, crowd simulation and crowd data have always maintained a closely complementary relationship. For instance, many calibration and validation methods of crowd simulation rely on data obtained from the real world. Moreover, numerous crowd data extraction methods based on computer vision have borrowed the prior knowledge embedded in crowd simulation models to enhance the quality of data extracted from video recordings. Additionally, in VR-based crowd experiments, the creation of virtual agents that interact with users is inseparable from the use of crowd simulation. In this section, we will review the main methods of crowd simulation and how crowd data is used to improve the quality of crowd simulation.

### 2.3.1 Knowledge-Based Approach

This simulation process involves moving from a predetermined starting point to a destination, ensuring no collisions with other agents in the vicinity. According to the research of Toll and Pettré [2021], microscopic simulation typically involves discrete and iterative computations of each agent’s movement (speed and position). This process can be decomposed into four steps: 1) searching for neighbors, 2) computing preferred velocity, 3) local navigation, and 4) movement. Most microscopic simulation algorithms focus on achieving collision avoidance while assigning a reasonable speed to each agent.

Early methods employed a force-based model for collision avoidance in simulation processes. Helbing and Molnár [1995] proposed the first force-based model for collision avoidance. This approach analogized crowd collision avoidance behavior using a physical force model and introduced a virtual force, known as the social force, to control the actions of each agent. In the simulation, each agent is initially influenced by an attractive force directed towards their destination, followed by a repulsive force emanating from other

agents or obstacles, which intensifies as their mutual distance decreases. The resultant force of these influences ultimately changes the agent’s velocity in the form of acceleration, thereby altering their trajectory. Karamouzas et al. [2009] improved upon this model by predicting the relative positions of agents in potential future collisions, defining the direction of the repulsive force to achieve more realistic collision avoidance. Similarly, subsequent studies explored different methods to define the repulsive force, enhancing the realism of agent behaviors [Zanlungo et al. 2011; Karamouzas et al. 2014].

Besides the force-based model, the velocity-based model is also an important method. This approach is characterized by allowing each agent to actively choose its next velocity after considering various possible velocities [Toll and Pettré 2021]. The basic idea is that, at each step of the simulation, an appropriate velocity is selected for the agent based on the information it can observe, aiming to make its behavior as realistic as possible. These methods typically evaluate an agent’s velocity through a score function. For instance, this score function could be a cost function that describes the energy expenditure of the agent [Guy et al. 2010] or other types of functions [Guy et al. 2011; Narang et al. 2017]. To find a velocity that makes the agent’s score as satisfactory as possible, velocity-based methods can be further divided into sampling-based [Berg et al. 2008] and optimization-based strategies [Van Den Berg et al. 2011].

Another type of simulation is vision-based approach. The approaches of this kind “attempt to replicate how human navigate literally based on what they see” [Toll and Pettré 2021]. the essence of vision-based methods involves rendering virtual visual information for each agent to simulate their observation of the virtual environment. Contrasting with previously mentioned methods, a distinctive feature of this approach is that agents cannot access the global information of the virtual scene. Instead, they rely solely on the information acquired through their virtual vision (typically stored in pixels) for anticipating and avoiding collisions. Additionally, due to the necessity of rendering virtual vision for agents, these methods also pay special attention to the agents’ speed and direction. This application requires agents to base their movement calculations more on visual information than on global data.

Ondřej et al. [2010] proposed one of the first vision-based crowd simulation approach. Based on the information lied in each pixel, this algorithm adjusts the agent’s speed and direction to avoid future or imminent collisions, while trying to keep the agent’s speed as constant as possible. Dutra et al. [2017] defined a cost function that involves the agent’s speed and angle. The cost function is computed using the rendered synthetic vision. At

each simulation step, the algorithm uses the gradient of the cost function to update the agent’s state. Other studies such as López et al. [2019]’s work leveraged optical flow as an alternative approach to extract inter-frame information from each pixel of the synthetic vision. Again, the authors defined a cost function based on the agent’s synthetic vision and computed the agent’s future movement through optimization.

## **2.3.2 Data-Driven and Deep Learning approaches**

### **2.3.2.1 Searching From Database**

The knowledge-based methods model pedestrian behaviors through manually designed rules or functions. In contrast, a more intuitive approach is to use data from datasets to generate individual behavior in crowds. To ensure the realism of the generated data, the most straightforward method is to directly reuse behaviors from the crowd dataset. For an agent in a certain state at a given moment, if we can find a pedestrian in the dataset with the same state, we can naturally copy the pedestrian’s next action to the agent’s subsequent moment, thus ensuring the realism of the agent’s behavior. However, finding an identical state in the dataset is rare. Therefore, it is worth considering searching for sufficiently similar data in the dataset and then transferring the corresponding subsequent action to the agent in the simulation. This approach requires clustering the existing crowd datasets and an efficient querying method [Lerner et al. 2007; Charalambous and Chrysanthou 2014]. Although this method demands a high level of data richness, the primary interest in current research is in finding the most appropriate match between the agent state used for querying and the data obtained. Illustrating this point, Zhao et al. [2013] employed a neural network classifier to establish an implicit criterion for such matching. Similarly, Boatright et al. [2015] contributed to this line of research by introducing macroscopic characteristics, like local density and crowd flow, as additional parameters to refine the query process. Further, Ren et al. [2018] developed a set of objective functions based on the action state of the agent and utilized optimization methods to retrieve the most suitable real data, thereby enhancing the match’s accuracy between the agent state and the data.

### **2.3.2.2 Trajectory Prediction Using Deep Learning**

With the widespread application of deep learning across various research fields, new methods based on deep learning have emerged in the field of crowd simulation. Unlike

traditional data query methods, these approaches directly use crowd data as training data for neural networks, aiming to fit an abstract mathematical model through the nonlinear computational process of deep neural networks. The primary goal of these methods is to predict the future trajectory of an agent based on past trajectory, thus termed Human Trajectory Prediction (HTP). In this field, network structures like Recurrent Neural Networks and Long Short-Term Memory are particularly emphasized [Alahi et al. 2016]. Some studies view HTP as a problem of generating future trajectories based on past trajectories and states, exploring the application of GAN [Gupta et al. 2018; Amirian et al. 2019], VAE, and Conditional VAE [Salzmann et al. 2020], as well as diffusion models [Gu et al. 2022; Rempe et al. 2023] for generating and predicting pedestrian trajectories to achieve crowd simulation. Other research focuses on generating trajectories in a broader sense, not limited to pedestrian trajectories, showing good applicability on human crowd datasets [Zhao et al. 2019; Ivanovic and Pavone 2018].

### 2.3.2.3 Reinforcement Learning

Reinforcement learning (RL) enables systems to learn through trial and error, focusing on optimizing state descriptions and reward functions. While traditional RL has clear, well-defined goals, it's not inherently data-driven unless input data similarity is integrated into its objectives. Notably, RL models prioritize optimizing objectives over mimicking input data, contrasting with approaches like generative models.

Recent advancements apply RL to local navigation and agent interaction, acknowledging the complexity of designing optimal problem setups and the potential benefits of automating this process. For instance, deep reinforcement learning leverages deep neural networks to process raw data, simplifying state descriptions and enhancing action decision-making. Pioneering implementations in agent navigation by Casadiego and Pelechano [2015], and robot navigation by Chen et al. [2017b], highlight RL's ability to surpass traditional methods.

Furthermore, applications in crowd simulation and footstep-based navigation, explored by Lee et al. [2018] and Haworth et al. [2020], respectively, demonstrate RL's adaptability and superiority in various domains. These findings emphasize deep RL's potential, particularly when coupled with straightforward reward functions, to outshine conventional approaches.

### 2.3.3 Evaluation Methods

The evaluation of a crowd simulation algorithm can concern different aspects such as the real-time performance, the ability to compute energy-efficient motion of agents, or the level of realism [Toll and Pettré 2021]. In this thesis, however, our review only focuses on the evaluation methods that concern the level of realism (the authenticity of the simulated results.)

In evaluating the level of realism in simulation results, one approach involves focusing on the trajectory’s displayed features, either partially or entirely, and assessing these features to gauge simulation quality. Daniel et al. [2021], after discussions with experts in crowd-related fields, identified features as illustrated in their publication. These features encompass both microscopic characteristics, such as average walking speed and angular velocity, and macroscopic characteristics like local density and the fundamental diagram. Based on these features, Daniel et al. [2021] proposed the concept of a quality function, which is a weighted sum of multiple feature scores, providing a comprehensive assessment. This score’s correlation with realism was validated through perceptual experiments. Another intuitive approach is to compare simulation results with real-world crowd data. If the simulation closely resembles actual crowd data, it is naturally presumed to possess a degree of realism. Multiple studies have followed this method to evaluate their simulation algorithms. As traditional simulation approaches often involve numerous parameters, real-world data are also utilized for parameter calibration, enhancing the quality of the simulation. Furthering this exploration, the challenge of accurately mirroring real-world crowd dynamics in simulations is underscored by the inherent variability and complexity of human and collective behavior. Despite this, innovative methodologies have emerged, seeking to refine the authenticity of simulation results. Fundamental diagrams serve as an initial benchmark, yet they capture only a fraction of the intricate dynamics present in dense crowd scenarios. More nuanced approaches involve matching simulated trajectories with actual human movements, acknowledging the unpredictable nature of individual and collective behaviors.

Pioneers in this field have ventured beyond rudimentary comparisons, employing a variety of techniques to bridge the gap between simulated and real-world crowd behaviors. From Pettré et al. [2009] focusing on velocity-based steering algorithms to Lerner et al. [2009] leveraging data-driven methods for a more granular comparison, the evolution of evaluation methodologies is evident. Charalambous et al. [2014] introduced an outlier detection mechanism, providing a unique lens to identify discrepancies in agent

trajectories, while Guy et al. [2012] proposed a comprehensive metric based on entropy to capture the nuances of crowd interactions and movements.

Furthermore, the work of Wang et al. [2017] and He et al. [2020] sheds light on the importance of understanding crowd behavior at both the micro and macro levels. By clustering paths and analyzing patterns based on shape, speed, and time, their research underscores the multifaceted nature of crowd simulations, highlighting the need for comprehensive and adaptable evaluation metrics to truly capture the essence of realism in crowd simulations.

Complementing these methodologies, the optimization of input parameters emerges as a pivotal aspect in enhancing the quality and realism of crowd simulations. Recent studies [Wolinski et al. 2014b; Berseth et al. 2014] have transitioned from manual to automated methods for parameter tuning, framing it as an optimization challenge with the aim of refining simulation outcomes.

Berseth et al. [2014] conceptualized a framework tailored to SteerBench metrics [Singh et al. 2009], delineating calibration methodologies and embracing metrics that mirror real-world trajectory data, thus enriching the fidelity of simulations.

Simultaneously, Wolinski et al. [2014b] accentuated the importance of congruence between simulations and actual datasets, dissecting both the individual and collective behaviors through a myriad of behavioral metrics. Their exploration of diverse algorithms for parameter optimization underlines the profound impact of fine-tuning multiple metrics concurrently, paving the way for simulations that not only mimic but resonate with the complex dynamics of real-world crowds.

### 2.3.4 Summary

In this section, we review research related to crowd simulation and how crowd data collected from the real world facilitates crowd simulation. Several conclusions can be summarized as follows:

First, crowd data plays a pivotal role in crowd simulation. In traditional approaches, crowd data is utilized to validate and calibrate simulation algorithms. This validation encompasses various aspects of crowd data, including microscopic characteristics such as individual acceleration and speed, as well as macroscopic elements like local density and flow speed. It even involves specific local interaction behaviors, including collision avoidance, group behavior, and following behavior. In the recent surge of machine learning methods, crowd data is directly employed in training neural network models.

Second, the efficacy of traditional crowd simulation is often influenced by multiple parameters, and the choice of these parameters is typically dependent on the simulation scenario. This implies that simulation methods are usually scenario-specific and lack universality, with no single method performing optimally across all scenarios. While some efforts have been made to enhance simulation quality by combining various simulation algorithms, the need for scenario-specific parameter settings is unavoidable. Incorrect parameter settings can lead to unrealistic local or even collective behavior. Calibration of the simulation process using real-world crowd data is possible, but the scenarios covered by crowd data are limited, necessitating the availability of targeted real data for effective calibration.

Third, data-driven simulation methods require real-world datasets for supervising the training process. When existing crowd data is insufficient, some studies opt to use simulation methods to generate synthetic data.

Last, most current crowd simulation methods focus primarily on determining the acceleration, velocity, and position of agents. Although a few simulation methods also consider agent body motion, in most simulations, agents are represented as discs or some variant thereof in a three-dimensional world. Typically, when visualization is required, a common approach is to first complete the simulation and then match each agent with a body animation based on their velocity state at each moment using a certain strategy. However, when constructing virtual environments in a three-dimensional world, our aim is to simulate both trajectory and body motion as realistically as possible. If any step in the aforementioned processes yields unrealistic results, it is foreseeable that the overall simulation outcome will also be unrealistic.

## 2.4 Conclusion

In this chapter, we have seen the importance of crowd data for both understanding and modeling crowd behaviors. Given the complexity of recording crowds in real life (Section 2.1) as well as the current limitations of VR-based crowd data collection (e.g., typically limited to single user – Section 2.2), there is a need for novel approaches for capturing complex crowd data, which is the focus of the next chapters.

# THE ONE-MAN-CROWD: SINGLE USER GENERATION OF CROWD MOTIONS USING VIRTUAL REALITY

---

## 3.1 Introduction

In this chapter, we present the first contribution of this thesis: the One-Man-Crowd paradigm, a novel approach for capturing crowd motion dataset based on immersing a single user in Virtual Reality with a single user.



Figure 3.1 – Snapshot of crowd motions generated using the one-man-crowd approach. A single user successively embodies each displayed virtual agent in the order indicated by the highlighting color (from blue to yellow). We studied 3 scenarios that replicated existing experiments from left to right: circular unidirectional flow, bottleneck situation, inflow behavior (entering a lift).

Our review of the state-of-the-arts in Chapter 2 has revealed that VR has the potential to capture the user’s behavior in a very detailed level, including the user’s motion, and with what object or which character the user is interacting. VR’s potential for capturing single user’s behavior inspires us to explore crowd motion capture in the virtual world.

Our first contribution in this thesis is dedicated to proposing and validating a VR-based crowd motion capture paradigm, Moreover, this study offers the possibility of cap-



turing crowd motion datasets with a single human user. To present this new paradigm, we first define key concepts and notations in Section 3.2, followed by the general hypotheses in Section 3.3. Then, we design and realize the technical procedure that concerns some hardware and software in Section 3.4. Our system enables to capture not only the trajectories of the user acting virtual pedestrians, but also his or her body motion. Research questions emerged naturally along with the realization of our crowd motion capture paradigm: can we use it to effectively collect a *valuable* crowd dataset? How can we validate the paradigm for the use of crowd motion capture? We chose to validate our method by demonstrating the similarity between our results and the real-world data which has been acquired in previous research. To this end, we adopted 3 classical unidirectional crowd motion scenarios that are of great interest of study in the crowd-controlled experiment literature, and replicated them using our OMC paradigm, to capture crowd data and compare the results with our references (Section 3.4). Our results showed positive outcomes on the three chosen scenarios (Section 3.5). However, a number of limitations and biases were observed, which we further explored as future work in this chapter, leading to a small further step towards the understanding of the paradigm’s performance on multidirectional crowd movement (Section 3.7).

In summary, the contributions of this chapter are threefold: (i) We introduce the one-man-crowd, a new paradigm for using VR to record crowd datasets. (ii) We Propose and evaluate two versions of this paradigm to observe behavioral variety in data. (iii) We evaluate this paradigm in three well-known real-world experiments, demonstrating that OMC preserves some emerging large-scale crowd motion patterns, despite the lack of variety and unilateral interaction bias.

## 3.2 Key Concepts & Notations

### 3.2.1 Notations

To ensure a clear explanation of our methodology, we define the following concepts to avoid nuances that might make our technical description difficult to understand.

- **Crowd:** In this thesis, we precisely define a **crowd** as a group of pedestrians moving in a shared space during the same period of time.
- **Agent and Characters:** We define both **agents** and **characters** as the animated virtual humans used to populate a virtual scene. In the context of using **agents**

to describe virtual humans, our focus is more on the nature of virtual humans being driven by algorithms, whereas the term **character** better captures the user’s perception of the virtual person in a virtual environment. Despite this nuance, Despite this nuance, these two terms might be used interchangeably in some parts of this thesis.

### 3.2.2 Formulating the One-Man-Crowd Paradigm

The crowd motion capture paradigm that we proposed in this chapter, i.e., the One-Man-Crowd paradigm, can be viewed as a decomposition of regular crowd motion capture in the real world. We may conceptualize the process of recording a crowd’s movements as a series of individual captures, each corresponding to a single person within the crowd. To facilitate description, let’s define the capture of an individual as a **unit capture**. Assuming a crowd of  $N$  pedestrians is to be captured, then  $N$  unit captures await to be (ideally) simultaneously executed. As such captures in the real-world face multiple challenges, Virtual Reality provides the possibility to remotely and distributively capture  $N$  users in parallel, and to visualize them together in a common virtual scene. Indeed, while this VR solution allows focusing on only one pedestrian in each unit capture, it would require  $N$  independent VR capture systems to work simultaneously, which, again, incurs costs and logistical challenges. To simplify such processes, we consider separating the  $N$  unit captures in time, by transforming “executing  $N$  unit captures in one time” into “executing one unit capture in  $N$  times”. Thus, the capture process can be restructured into  $N$  **iterations**, with each iteration involving a single unit capture of a different VR user. In the following of this thesis, we refer to this idea as a procedure of **N-users-to-capture-N-characters (N2N)**:

**N2N.** For a given iteration  $T_i$ , a **different** user  $U_i$  is immersed in the virtual environment in one specific character  $c_i$ , displaying the user’s co-located motions in real-time. Every character  $c_{j \in [1..i-1]}$  is displaying global and body motions recorded during previous iterations  $T_j$  of **each corresponding user**  $U_{j \in [1..i-1]}$ . Again, characters  $c_{k \in [i+1..N]}$  are handled differently depending on the specific crowd scenario to be captured.

Assuming the practicality of this subdivision, we can extend this concept to a more

innovative approach: allowing a single user to sequentially perform these iterations and individually complete the  $N$  unit captures. Similarly, we refer to this idea as a procedure of **Single-user-to-capture-N-characters (S2N)**:

**S2N.** For a given iteration  $T_i$ , a **single** user is immersed in the virtual environment in one specific character  $c_i$ , displaying the user's co-located motions in real-time. Every character  $c_{j \in [1..i-1]}$  is displaying global and body motions recorded during previous iterations  $T_j$  of this specific user. Characters  $c_{k \in [i+1..N]}$  are handled differently depending on the specific crowd scenario to be captured.

Thus, the readers would easily understand that the name One-Man-Crowd stands for the S2N procedure in principle. However, for the simplicity of text organization, we consider both N2N and S2N procedures as different realizations of the One-Man-Crowd paradigm. Note that, from the procedural perspective, the OMC paradigm can be as a repetitive process of **recording** human motion and then **replaying** the recorded motion through a virtual character. For this reason, in the following chapter where several variations of the OMC paradigm are presented and explored, we will further refer to this original approach as the **One-Man-Crowd** paradigm.

### 3.3 General Hypotheses

We seek to validate the idea of creating human crowd datasets using the One-Man-Crowd capture paradigm. From a technical perspective, the capture process is indeed feasible. We need, however, to prove that the captured datasets are *valuable*, i.e., comparable with the real-world crowd datasets. To this end, we set up several general hypotheses that are related to the performance of the OMC paradigm, and further accept or reject them through an experiment presented in Section 3.4.

Our general hypotheses are related to the possible effects of all the perceptual biases (e.g., distance or self-motion perception), behavioral biases (e.g., slower walking speeds), variety biases (e.g., single user in the S2N case) and asynchrony biases (e.g., unilateral interactions) on the data generated by the OMC approach.

- **H1:** At the individual level, quantitative differences are expected between real and OMC virtual data. Aside from this, we expect a high similarity with real data, and

to preserve existing relations between analyzed variables.

- **H2:** On the collective scale, the OMC paradigm will lead to a lack of variety in data, meaning that S2N data could be considered as a subset of real or N2N data (e.g., S2N data distributions included in real data distributions). Moreover, N2N data will be more similar to real data than S2N data.
- **H3:** We expect that both N2N and S2N conditions will enable us to observe the emergence of crowd motion patterns, as the scenarios that we replicate involve mostly unilateral interactions.

As an important note, we explore those hypotheses in the light of qualitative analyses described in the following sections. As we replicate experiments designed by the crowd modelling community, we also replicate the analysis framework that was performed, and discuss the most striking effects we can observe. This choice is discussed in Section 3.6.

## 3.4 Experiment

To investigate whether the OMC paradigm can be used to create *valuable* crowd motion datasets, we replicated three real crowd experiments using the OMC paradigm: 1D-flow, bottleneck, and inflow experiments (elaborated in subsequent sections). These scenarios are of great interest in the field of crowd modeling. As we advance the concept of leveraging VR to collect crowd data beyond previous research, we inherit the knowledge and findings from these studies. With such information, we assume that while the use of VR may introduce behavioral and perceptual biases, it does not fundamentally affect the capture process. Thus, by replicating these experiments, we do not seek *quantitative accuracy*, but rather a qualitative similarity that aligns with real-world situations. We hypothesize that the OMC paradigm may introduce two additional main biases: (i) Asynchrony biases due to unilateral interactions, where only the immersed participant can respond to virtual agents, and (ii) Variety biases arising from limited behavioral diversity, as the paradigm involves a **single** participant repeatedly walking. We utilized both the S2N procedure and the N2N procedure of OMC to investigate these potential biases.

### 3.4.1 Experimental Design

Upon their arrival, participants read and signed the experiment consent form, during which they were presented the task to perform. They were then equipped with a Xsens

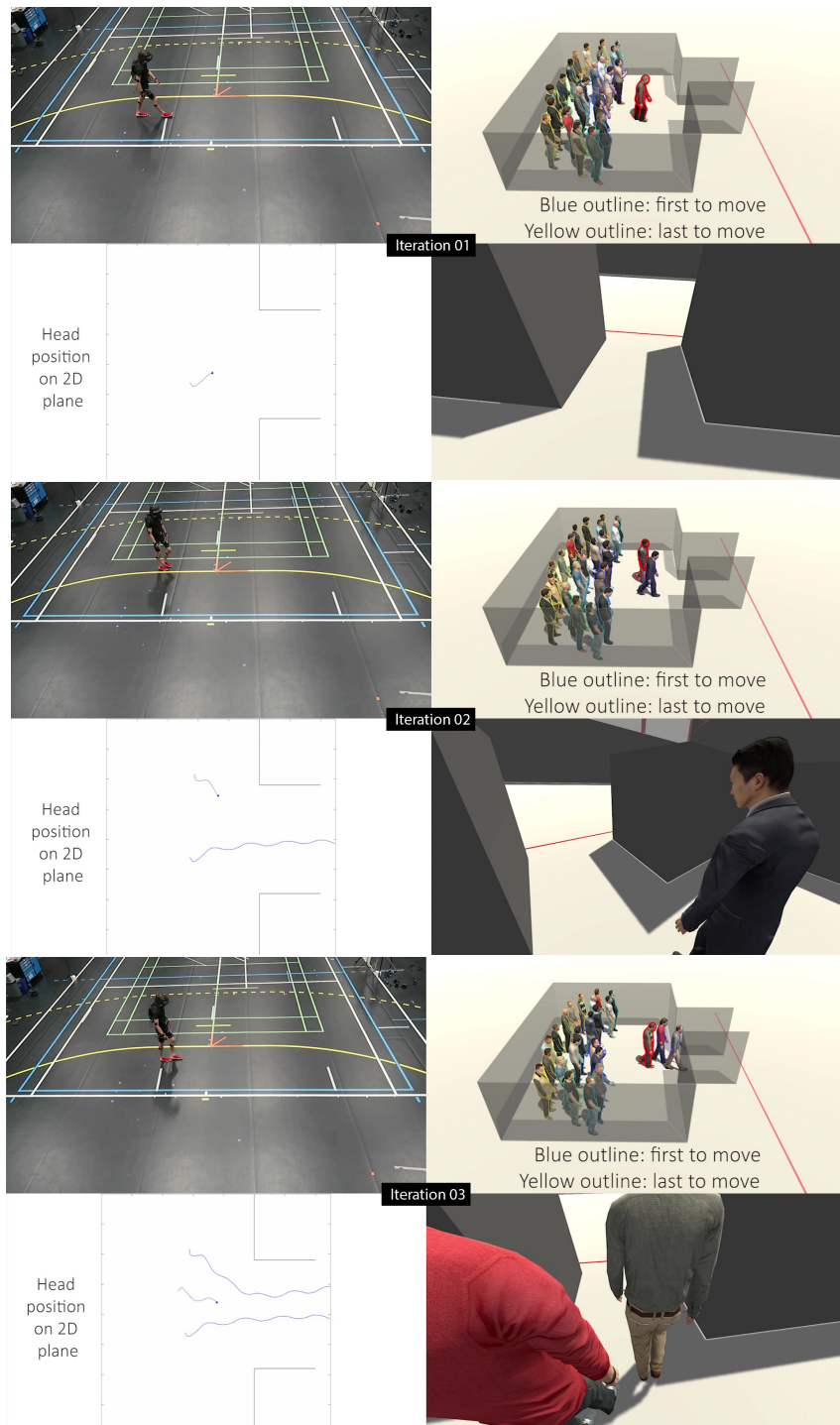


Figure 3.2 – Illustration of the first 3 iterations of the One-Man-Crowd motion capture paradigm. Each figure provides 4 different perspectives: 1) The user’s movement in the real world (top left). 2) The user’s avatar in the virtual scene, highlighted by red outline (top right). 3) The user’s trajectory from the beginning of each iteration to his current position, the trajectory is projected onto a 2D plan (bottom left). 4) The user’s first person view while walking during each iteration (bottom right).

motion capture system, a VR backpack, as well as a Pimax 5K HMD. Calibration of the motion capture system was then performed to ensure motion capture quality, as well as to resize the avatar to the participants' dimensions.

Participants were then immersed into the virtual world, and were instructed to accomplish numerous iterations of capture. During each iteration, participants started the iteration on a red spot (located at a fixed position during the entire experiment), then walked towards another spot located on the ground among a group of virtual characters (initial position of the character  $c_i$ ). Each scenario had a specific ending condition, which would terminate the iteration once activated. Participants would then be guided back to the initial red spot (not displayed during the actual iteration) so that they would always start new iterations from the same location. Figure 3.2 provides an illustration of the 3 initial iterations of such a procedure.

Participants first performed one iteration of each scenario presented in Section 3.4.2, contributing to the creation of the N2N baseline described below, after which they performed a full S2N process. The specialty of these two tasks are listed below.

**N2N baseline:** trajectories and motions of the  $N$  characters in the crowd were created successively by  $N$  different participants. For each scenario, a participant  $i$  only took the role of the  $i$ -th character  $c_i$ , while the  $c_{j \in [1..i-1]}$  previous characters were animated based on recorded animations and trajectories of the previous  $i - 1$  participants. Participants always performed an iteration of the 1D-flow scenario first, then the bottleneck scenario, and finally the inflow scenario.

**S2N procedure:** a single participant created successively the trajectories and motions of all the  $N$  characters in the crowd, by taking the role of one character at a time. Each participant took part in a single scenario, comprised of 24 or 25 iterations (depending on the number of characters to act). It is also important to point out that participants were never told during the experiment that they would be interacting with previous versions of themselves (see Section 3.6 for further discussion).

### 3.4.2 Scenarios & Tasks

The 3 scenarios that we chose to replicate are of great interest in the understanding of crowd behaviors and have been studied by multiple past research. Those research explored and established frameworks to evaluate human behaviors in the relevant crowd scenarios

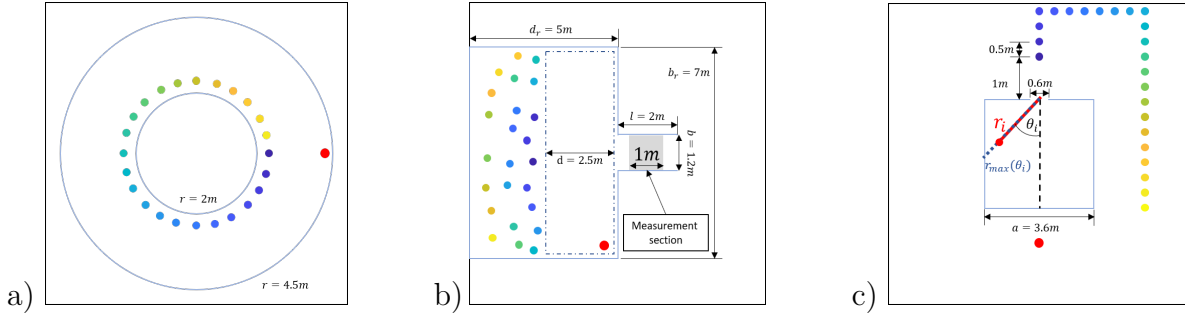


Figure 3.3 – Configuration of the three virtual experiments. Black squares delimit the  $10\text{ m} \times 10\text{ m}$  experimental ground. Blue lines and circles represent walls defining the virtual room of each scene. Red dots represent the initial position and ending position of each iteration. The other dots indicate the starting positions of each virtual character, colored from dark blue (first to move) to bright yellow (last to move). For each iteration, participants walked from the red dot to the position of one colored dot to trigger the experience. a) Characters are uniformly spaced out along the circular path. b) The initial positions of characters are generated by a Poisson distribution, and ordered based on their distance to the entrance of the bottleneck. c) The queue of characters waiting to enter the virtual lift was adapted to our experimental space.

from both individual and collective level. We hereby describe each scenario in detail and provide the references that we used to evaluate the performance of the OMC paradigm, then, we explain the exact task that participants needed to accomplish in each scenario

### 3.4.2.1 1D-flow

**Context.** Unidirectional (1D-)flows are composed of pedestrians moving in the same direction [Yamori 1998; Daamen and Hoogendoorn 2003a], and are interesting to study because they are frequent in real environments (e.g., queues or traffic in narrow corridors) with applications to safety (e.g., prediction of the evacuation time through a set of security exits). Experimentally, they are often reduced to a single lane of participants following one another [Seyfried et al. 2005b; Seyfried et al. 2010a]. 1D-flow experiments have often been replicated to understand the effect of various independent variables: cultural parameters [Chattaraj et al. 2009; Gulhare et al. 2018], age [Cao et al. 2016; Ren et al. 2019; Fang et al. 2019], lighting conditions [Cao et al. 2019b], environment complexity [Chen et al. 2017a; Huang et al. 2018; Sun et al. 2018], sound [Zeng et al. 2019; D. Yanagisawa et al. 2012], etc. 1D-flow data are analyzed in different ways. Individual variables explore relations between speed and headway. Their aggregation leads to rela-

tions between flow and density that are captured by fundamental diagrams. Finally, the emergence of stop-and-go waves (similar to the ones emerging in car traffic jams) was also investigated [Lemercier et al. 2012; Jelić et al. 2012a].

**Task.** In our experiment, the 1D-flow scenario was designed to replicate a real-world experimental condition used by Lemercier et al. [2012], where 24 real participants walked behind each other in a circular space. As in the real-world experiment, our participants were instructed to walk along a circular path (radius: 2.4  $m$ , i.e., trajectory length of 15.1  $m$ ) delimited by two circular walls (inner radius: 2  $m$ , outer radius: 4.5  $m$  – see Figure 3.3-a). The inner wall was 2  $m$  height, to prevent participants from anticipating their future movement based on characters too far down the line. During the experiment, participants were instructed to follow the character in front of them from the beginning of each iteration, while avoiding collisions. They were also not allowed to overtake the character in front of them. The following characters were animated using a pre-recorded motion capture animation (see below). They did not react to participants’ motions, but became transparent if they overtook or went through the participant’s avatar. For the first iteration, all the characters (except the participant’s avatar) were animated with a pre-recorded motion capture animation so that participants could continuously adjust their speed depending on a previous character as in the real-world experiment, and would not walk alone. This pre-recorded animation was motion captured on a male confederate prior to the experiments. The confederate was instructed to follow a virtual object in the same virtual scene, animated using the real trajectory of one experimental condition of Lemercier et al. [2012] so that participants would respond according to acceleration and deceleration observed in a previous real situation. Using the same pre-recorded trajectory and animation for each participant to initially follow also helped to ensure comparable situations in the following analyses. In each iteration, participants walked for a duration of 60s, after which they were instructed to return to the initial location indicated by the red spot.

#### 3.4.2.2 Bottleneck

**Context.** In crowd study, *Bottleneck* stands for a flow captured in an environment whose capacity locally decreases (typically a corridor that narrows). As for 1D-flow, the situation is of great interest because it corresponds to daily situations and is used to predict



environment flow capacities and jams as well as evacuation times where the environment imposes the presence of bottlenecks [Fruin 1971a; Helbing et al. 2006]. Such situations are analyzed first by calculating flows of people before and after the bottleneck, i.e., the number of pedestrians passing through the bottleneck per unit of time in relation with the corridor widths [Carstens and Ring 1970; Navin and Wheeler 1969; Older 1968; Liao et al. 2014]. It was proved that only the smaller width determines the output flow [Adrian et al. 2018], and that personal motivation has a large effect on the flow [Adrian et al. 2020]. Finally note that the number of lanes participants can form in the narrow part of the corridor is important to study because it determines the flow. Humans may create lanes with little separation by walking in a staggered pattern, called the *zipper effect* [Seyfried et al. 2010a; Seyfried et al. 2009].

**Task.** We designed our *Bottleneck* scenario to replicate the real-world experimental condition used by Liddle et al. [2009], where a group of 24 participants were asked to walk through a real corridor (bottleneck). As in the real-world experiment, our participants were instructed to walk through a virtual bottleneck (1.2 m wide, 2 m long – see Figure 3.3-b), while avoiding collisions with virtual characters and attempting to exit the bottleneck quickly. Before starting a iteration, participants joined the other virtual characters in a  $7\text{ m} \times 2\text{ m}$  waiting zone, located 2 m from the entrance of bottleneck. Unlike in the previous situation, only the first  $c_{j \in [1..i-1]}$  virtual characters were animated during iteration  $T_i$ . The other characters  $c_{k \in [i+1..N]}$  remained idle in the waiting zone, but were presented to participants at the beginning of the iteration so that they had a sense of the number of characters involved in the scenario. The iteration ended when the participant went beyond a red line on the ground, indicating that they had exited the bottleneck. Moreover, the order of acting the virtual characters was based on their initial distance to the entrance of the bottleneck (from minimum to maximum), so that participants would first act virtual characters closest to the entrance.

### 3.4.2.3 Inflow

**Context.** *Inflow* stands for successive entry of pedestrians into a closed area [Ezaki et al. 2015]. Understanding this process helps modelling traffic in public transport systems (e.g., entering buses, trains, elevators) [Ezaki et al. 2015; Ezaki et al. 2012; Liu et al. 2016a]. At the individual level, existing analyses focus on the decision-making process by

participants to pick a final position in the confined space. To this end, different calculations have been adopted such as Nearest Neighbour Distance [Liu et al. 2016a], Proxemic Floor Field [Ezaki et al. 2012] and more frequently, Voronoï Diagram [Ezaki et al. 2015; Liu et al. 2016a]. As shown by these studies, the search for a personal comfort zone plays a dominant role in the decision-making process [Evans and Howard 1973; Sommer 1959; Waş et al. 2006], and locations near the walls of the vehicle are preferred [Liu et al. 2016b].

**Task.** This scenario was designed to replicate the real-world experimental condition used by Ezaki et al. [2015], where a group of 25 participants were asked to enter one after another into a lift. As in the real-world experiment, our participants were instructed to stand in a line of virtual characters, then to enter a virtual lift of  $3.6\text{ m} \times 3.6\text{ m}$  (see Figure 3.3-c) and to choose a place to stay once they entered it. The entry of the lift was  $60\text{ cm}$  wide to limit participants' view so that they could not perceive the position of the virtual characters that had already entered the lift. As for the bottleneck scenario, only the first  $c_{j \in [1..i-1]}$  virtual characters in front of participants were animated during iteration  $T_i$ , using participants' recorded animations from previous iterations. The other characters  $c_{k \in [i+1..N]}$  remained in their original position, but were presented standing in line with an idle motion to participants at the beginning of the iteration so that they had a sense of the number of characters that should enter the lift. At the beginning of this scenario, we insisted that participants should take into account that all the 25 virtual characters should be able to enter the lift, and therefore that they should choose their position during each iteration to enable the remaining characters to enter the lift as well. The iteration ended when the participant considered that he had reached his wanted position and was facing the entrance, which was indicated by participants pressing the trigger on the HTC Vive controller that they were holding in this scenario.

### 3.4.3 Apparatus

We immersed the participants in a Virtual Environment (designed in Unity 2019.4) using a Pimax 5K Head-Mounted Display (specifications:  $90\text{ Hz}$ ,  $200^\circ$  fov,  $2560 \times 1440$  resolution), which provides a wide field of view beneficial for such situations involving close proximity with other characters. The HMD was used with 4 SteamVR 2.0 base stations, yielding a tracking area of approximately  $10\text{ m} \times 10\text{ m}$ . We also used one HTC Vive controller in the Inflow scenario (details below), which was held by the participants and used to trigger the end of each iteration. During the experiment, we recorded participants'

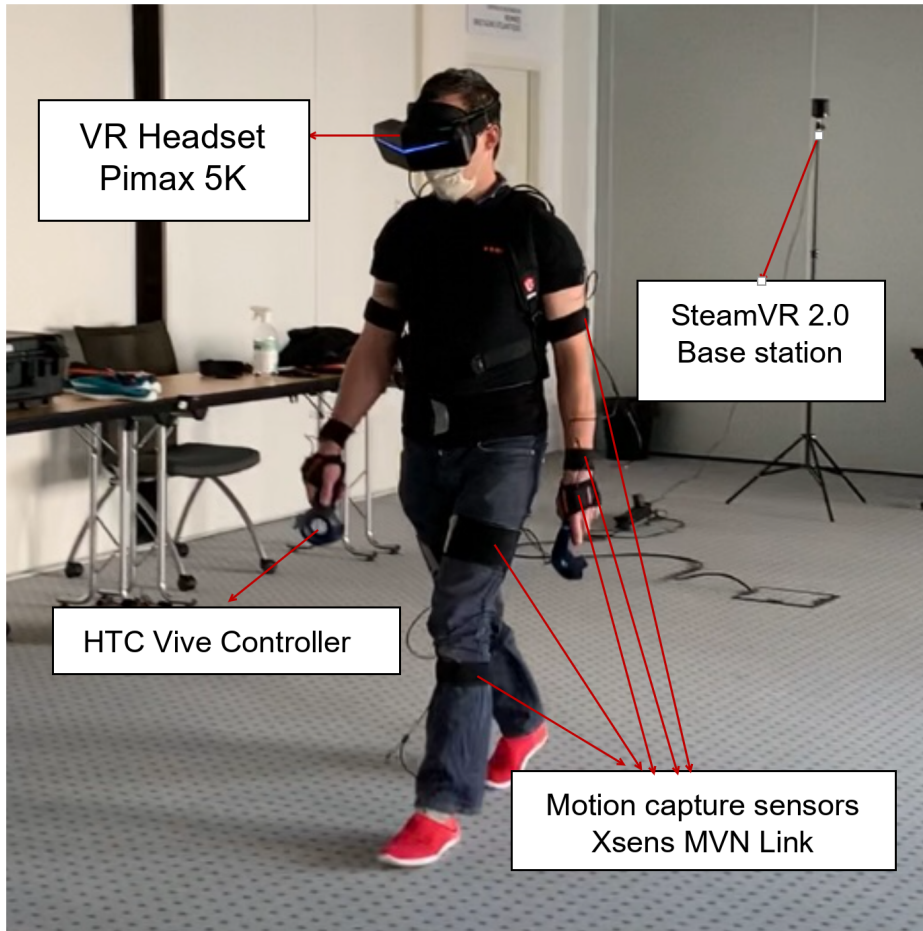


Figure 3.4 – An illustrative photo of the devices that used to conduct the experiment.

body motions using a Xsens motion capture system, which was used both to display the participants' motions in real-time on their avatar, as well as to animate the motions of the corresponding character in successive iterations. To enable participants to move freely in the environment, they were equipped with an HP Z VR G2 backpack (specifications: NVidia RTX 2080, Intel Core i7-8850H processor, 32GB RAM) running the experiment, on which were physically connected all the devices. This whole setup enabled participants to physically walk in the real space, while interacting with the other characters of the virtual crowd and seeing their own motions displayed on a co-localized avatar.

### 3.4.4 Participants

We recruited 25 participants for our experiment (age min=22, max=32, avg= $26.8 \pm 3.1$ ), which is the minimum number of participants required for the creation of our virtual baseline data. To avoid biases caused by motion characteristics coming from different genders, as well as to embody participants in gender-matched avatars, we recruited only male participants. Participants were recruited through internal emailing lists among students and staff, were all naive to the purpose of the experiment, had normal or corrected-to-normal vision, and gave written and informed consent before the experiment. The study conformed to the declaration of Helsinki.

For simplicity, as participants took part both in the N2N and S2N procedures, we attribute an ID from #1 to #25 to each participant, with respect to their participation order in our experiment. This means that participant # $i$  took the role of the  $i$ -th character in the N2N baseline scenarios. Participants #1 to #8 then performed the 1D-flow S2N scenario, participants #9 to #16 the bottleneck S2N scenario, and participants #17 to #25 the inflow S2N scenario.

## 3.5 Evaluation

We separately analyze and discuss the three different scenarios. As introduced in Section 3.4.2, for each scenario, previous research has proposed established evaluation frameworks, and each has different research focuses at both the individual and collective level. In each scenario evaluation, we explained in details our focuses and the corresponding results.

### 3.5.1 Scenario I - Unidirectional Flow

By virtually replicating Lemercier et al.'s experiment [2012], we expect to also replicate a number of real-world observations they made: the emergence of stop-and-go-waves, specific relations between individual speed and headway, as well as resulting fundamental diagrams. We also wonder whether the absence of physical constraints in VR will lead to unrealistic situations, such as overlaps between agent positions.

### 3.5.1.1 Analysis & Results

**Data processing.** For this scenario, participants' positions  $\mathbf{p}_i$  are computed according to a polar coordinate system centered on the circle path they follow:  $\mathbf{p}_i^{polar}(t) = [\theta_i(t); r_i(t)]$ . We are only interested in the evolution of  $\theta$ , as  $r_i(t)$  is almost constant since participants walk along a circle. We approximate  $r$  by its mean value  $r_{mean}$ . The position, speed, and acceleration can thus be reduced to:

$$\begin{cases} p_i(t) \approx \theta_i(t)r_{mean} \\ v_i(t) \approx \dot{\theta}_i(t)r_{mean} \\ a_i(t) \approx \ddot{\theta}_i(t)r_{mean} \end{cases} \quad (3.1)$$

**Stop-and-go waves.** For each S2N dataset, as well as the N2N dataset and our reference real-world result, we respectively plot the trajectories of all the virtual characters or real pedestrians in Figure 3.5. (Todo: the blue dashed line is missing in all of these results -> see the camera ready version article.) The upper blue dashed curve corresponds to the first agent moving in the circle, i.e., the pre-recorded confederate reference trajectory. As we said in Section 3.4.2, this trajectory reproduces a real case (center plot) and is made of periods where the confederate walks, stops, and walks again, etc. The plot makes evident that the stops of the confederate propagated backward in the 1D-flow, with a certain propagation speed. Jelić et al. [2012a] proposed a method to compute the speed of counter flow propagation, by computing the local extrema of  $v_i(t)$ , and by performing a linear regression on the obtained points, as illustrated in Figure 3.5. Figure 3.6 shows the distribution of the obtained speed values for all the iterations of each participant, as well as for the example from Lemerrier et al.'s real experiment. This figure shows that S2N experiments resulted in both faster and slower waves propagation speeds, depending on participants, and that N2N resulted in in-between values, close to real ones.

**Reaction time and following distance.** Lemerrier et al. [2012] found that participants in a 1D-flow were mainly matching the speed of their predecessor, trying to cancel  $\Delta v_i(t) = v_i(t) - v_{i-1}(t)$ . They also found that the propagation speed of waves was probably determined by participants' reaction time  $\tau$  when matching speed, and that the wave could be damped or resurge depending on the following distance between individuals

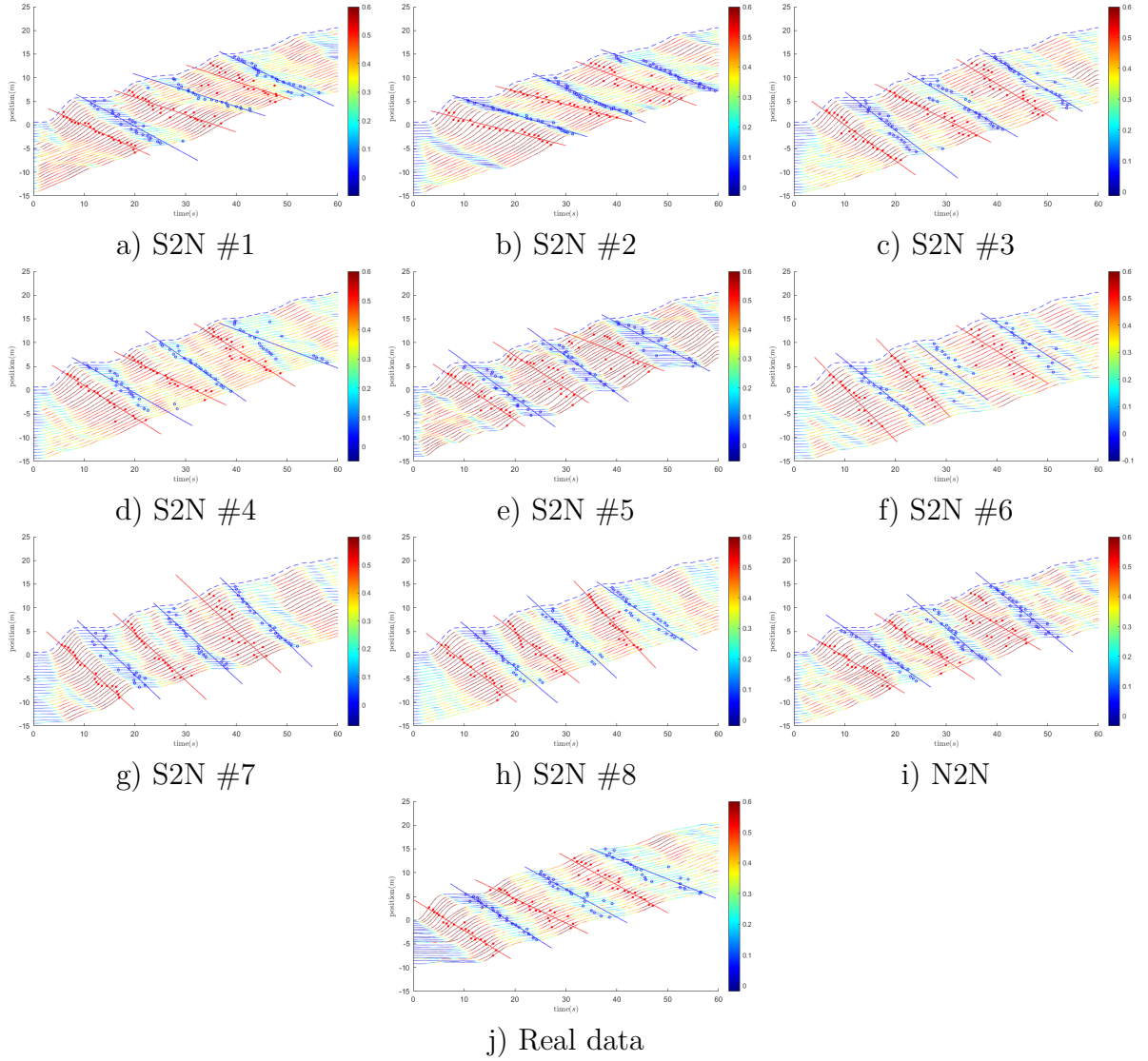


Figure 3.5 – Unidirectional flow scenario: evolution of pedestrians' position and the results of linear regression. Our real data refers to Lemerrier et al.'s work [2012].

$$p_i(t) - p_{i-1}(t).$$

$$\tau_i^*(t) = \arg \max_{\tau} (a_i(t) \star \Delta v_i(t + \tau)) \quad (3.2)$$

where  $a_i(t)$  is the participant's acceleration and  $\star$  the cross-correlation. Figure 3.7 plots the obtained distributions for reaction time as well as following distances for each dataset (S2N and N2N) in comparison to values obtained for Lemerrier et al.'s experiments [2012]. This figure shows that the range of observed reaction times is very close between OMC-S2N or N2N and real. On average, participants matched the speed of their predecessors

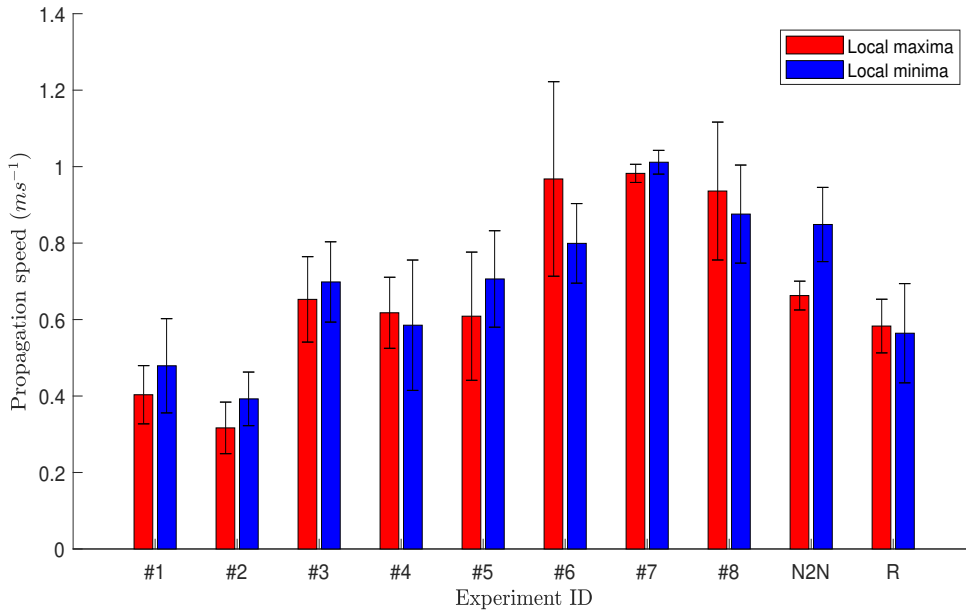


Figure 3.6 – Average and SDM results of all S2N (#1 to #8), N2N and real-world data (R).

with a delay of 0.5 s.

**Fundamental Diagram.** We computed the fundamental diagram that captures relations between speed and density. As in [Jelić et al. 2012a], we considered local density and speed, i.e., where  $\rho_i(t) = (p_i(t) - p_{i-1}(t))^{-1}$ . We smoothed values by adopting the following binning procedure. First, we calculated the density-speed pair  $(\rho_i(t), v_i(t))$  for each character  $i$  at each time step  $t$ . Then, we sorted all the density-speed pairs from the minimum density to the maximum, and separated them into 960 bins of  $K$  elements each ( $K = 180$  in our experimentation). Afterward, we computed the average density and speed per bin as:

$$\bar{\rho} = \frac{1}{K} \sum_{n=1}^K \rho_n \quad \bar{v} = \frac{1}{K} \sum_{n=1}^K v_n \quad (3.3)$$

Figure 3.8 plots the resulting pairs, for all S2N, N2N as well as the real example from Lemercier et al. [2012].

**Global density.** In real 1D-flow experiments, participants’ positions are restricted by the circle they follow, which gives a global density. In OMC virtual replications, since participants only see characters in front of them and cannot feel the physical constraints imposed by participants behind them, nothing prevents them from forming circles where

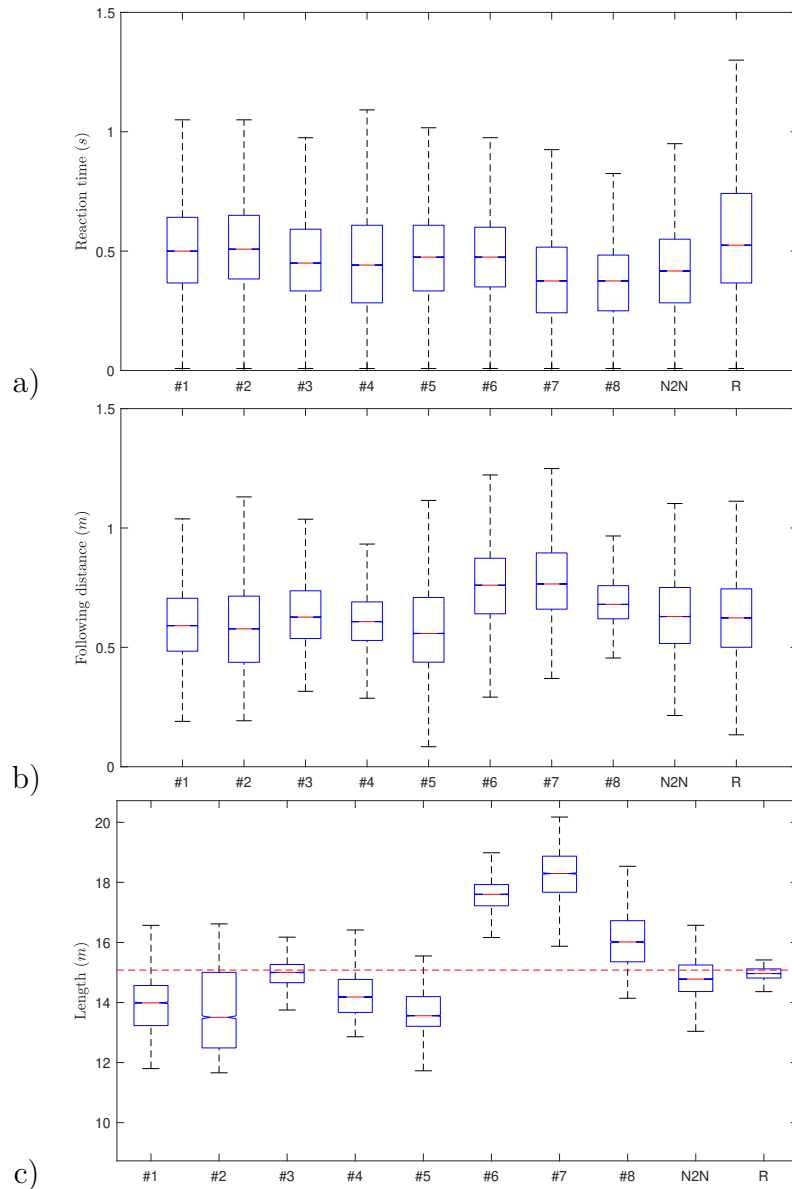


Figure 3.7 – a) Reaction time. b) Following distance. c) Total length of the circle formed by characters (the red dashed line indicates the physical length of the path). #1-#8: S2N results. R: Real-world data [Lemerancier et al. 2012].

the positions of the first characters overlap with the positions of the lasts. This explains why we gave an initial position to the participants out of the circle and asked them to get to their position in the queue: doing this allowed them to get an estimate of the global density, and we here check whether this global density was matched all along the iteration and whether character positions overlapped. To this end we compute the total



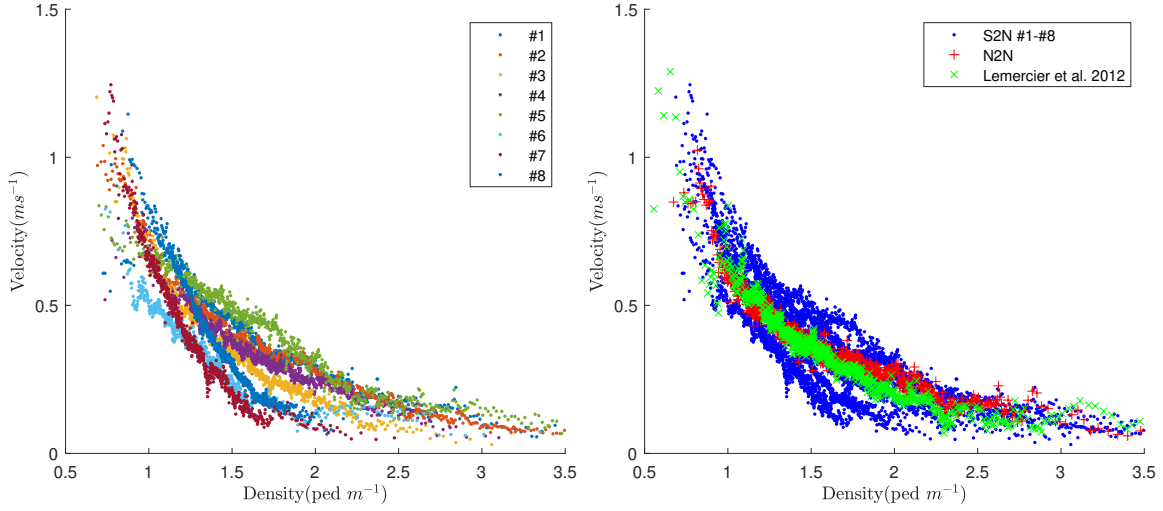


Figure 3.8 – Fundamental Diagrams showing the velocity-density relationship. Left: Results of the S2N procedure separately plotted. Right: Compared to real-world results, data generated by our method seems to be translated towards higher density level.

length  $L(t)$  of the circle formed by characters as the sum of the distance between each pair of predecessor-follower as follows:

$$L(t) = \sum_{i=2}^N p_i(t) - p_{i-1}(t) \quad (3.4)$$

As can be seen in Figure 3.7-c, the length of the 1-D queue exceeds the perimeter of the circle for a few participants, indicating the emergence of overlapping: the expected global density is not matched.

### 3.5.1.2 Discussion

We here discuss the comparison between the OMC 1D-flow and the real-world one, and justify our validation of the general hypotheses. Generally speaking, we observed the expected results in this scenario, i.e., similarities between the OMC results and the real-world reference.

First, Figure 3.5 illustrates the striking similarity in data we could obtain between Lemercier et al.’s real experiment and their OMC replication. Especially, we were able to observe as expected the propagation of stop-and-go waves for all participants, which validates **H3** (emergence of collective phenomena) for this scenario. The further analysis of wave characteristics, however, reveals quantitative differences, as expected by **H1**. Real

experiments resulted in wave propagation speeds limited to  $0.57 \pm 0.09 \text{ m.s}^{-1}$ , while a larger range could be observed in S2N datasets ( $0.69 \pm 0.24 \text{ m.s}^{-1}$ ). Moreover, due to a lack of variety (i.e., the same participant acting all the characters in the crowd), the S2N procedure resulted in participant-specific values. On the contrary, N2N results mixed and averaged these participant-specific behaviours. For instance, the propagation speed of waves in N2N results ( $0.66 \pm 0.03 \text{ m.s}^{-1}$ ) is close to the average of all our S2N datasets. As stated in **H2**, N2N lowers the effect of lack of variety, even though N2N wave propagation speeds remain different from real values. However, as real and OMC values were obtained over two different sets of participants, a more strict OMC-real comparison would need additional experiments, which is discussed in Section 3.6.

Then, according to their 1D-flow model, Lemercier et al. found that wave propagation speeds mainly depend on participants' reaction time to match the speed of their predecessor. This is consistent with our analysis of reaction time. For instance, participants #7 and #8 displayed both the lowest reaction times and the fastest wave propagation. As a result, when different participants mix in the N2N procedure, the resulting wave speed gets averaged as a result of individual differences in reaction times. We can however observe that participants with the lowest reaction times are the ones with the highest following distances. This suggests that they might simply be more careful and attentive, and that this individuality gets amplified by making them interacting with themselves. However, it is also possible that differences in reaction times in VR could be related to biases in the perception of distances, as objects typically look closer in VR than in real life [Loomis and Knapp 2003]. Exploring whether perceived proximity to predecessors could make participants more reactive would however need to be explored in future experiments.

Next, the analysis of fundamental diagrams leads us to similar conclusions. From Figure 3.8 we can see the speed-density pairs of each S2N result locate around the ones of real cases, with a distorted shape caused by the personal preference of each S2N participant and the lack of variety (**H1**). Thus, we could think that the real or N2N fundamental diagrams could result from the combination of the obtained S2N diagrams. Again, the N2N condition partly solves issues related to lack of variety, and the corresponding fundamental diagram has a more similar coverage compared to the real one (**H2**). However, note that fundamental diagrams of both the S2N and N2N procedure are also shifted compared to real data. This result indicates that participants tended to walk faster than in real life with the same local density. While this contradicts previous studies which found slower

speeds in VR [Fink et al. 2007a], one possible cause of this phenomenon is the absence of contact rendering combined with HMDs’ reduced field of vision: participants may not be able to see the entire body (of themselves and of their predecessor), especially the lower part. The physical space to perform walking motion was also not constrained by the presence of the physical legs of a real predecessor, resulting in larger steps and faster walking speed. Note that in real experiments, these physical constraints result in step locking behaviors as studied by Jelić et al. [2012b].

Moreover, we feared that the absence of physical constraints would lead to overlaps between agents’ position. For example, the beginning of the queue of agents may overlap its end. During the experiment, we did not render agents in this situation to avoid disturbing participants, as these agents would otherwise block their vision and traverse their avatar. We however compared the length of virtual queues against real ones (which are obviously bounded by the physical length of the circular path), as displayed in Figure 3.7-c. Our results provide mixed conclusions. Some participants clearly adopted following distances which, accumulated, violated the limit of the path length (e.g., #6 and #7), whereas most participants finally adopted distances that matched the expected global density. This suggests that our choice of making participants able to observe the global density by giving them a starting position out of the circle, then joining in the space allocated to them in the circle, somehow enabled most participants to be aware of the global density to be matched. This result also suggests that this choice might have had a persistent effect through the experiment, and that physical differences from real situations inherent to this virtual setup might be in some ways attenuated with careful organization.

As a final point, we discuss the choice to animate the first agent in the circle using motions pre-recorded by a confederate. It was motivated by two reasons. First, in real experiments, stop waves are initiated by the first participant in the circle being blocked by the presence of the last one. This would have raised a problem in the OMC replication, since the last agent motion is recorded last. Furthermore, as the confederate replicates the motion of the first participant in Lemercier et al.’s experiment, doing this enabled us to perform a strict comparison between real and virtual data as illustrated in Figure 3.5, as well as to ensure that all participants were presented with the same initial condition. While requiring to record specific animations prior to an experiment might be limiting in some situations, we believe that approaches replicating characteristics of an observed iteration could be also used, e.g., from recorded videos.

## 3.5.2 Scenario II - Bottleneck

In this replication of a bottleneck experiment [Liddle et al. 2009] we explore the biases introduced by the unilateral interactions, as participants need to negotiate with people coming from their side to enter the bottleneck. We evaluate the resulting number of lanes in the bottleneck as well as fundamental diagrams, since they are useful outcomes of such an experiment.

### 3.5.2.1 Analysis & Results

**Lane formation.** To count the number of lanes, we reproduced the method described by Seyfried et al. [2009] to estimate the probability to finding a pedestrian at a certain lateral position  $p_x$  inside a measurement zone (a corridor section along the  $y$  axis) with  $-0.6\text{ m} \leq x \leq 0.6\text{ m}$  and  $0.5\text{ m} \leq y \leq 1.5\text{ m}$  (see Fig. 3.3). In Figure 3.9, we present all the S2N results as well as the N2N result, and the real one. The distribution of lateral positions (binned) is presented in Figure 3.10. Real data corresponds to the results of the experiment performed by Liddle et al. [2009]. The number of formed lanes can be easily observed by counting the number of peaks in the distribution. For the real data, we can see that 2 lanes emerge. More difficult to appreciate from static plots only, the little distance separating the lanes results from the zipper effect (staggered walking pattern, see the companion video for illustration). Interestingly, OMC led to two extreme kinds of lane formation: a queue neatly lined up and a perfect formation of 2 lanes going through the bottleneck (as it can be seen in Figure 3.9-a and -h respectively). Meanwhile, most of our results present a mixed pattern between these two extreme results, resembling the N2N result presented in Figure 3.9-i which exhibits a normal distribution of lateral positions. Figure 3.9-a displays what can happen in the case of a single lane, where participants lined up by first walking directly towards the end of the queue, instead of gathering around the entrance (as observed in most other iterations, e.g., the N2N one).

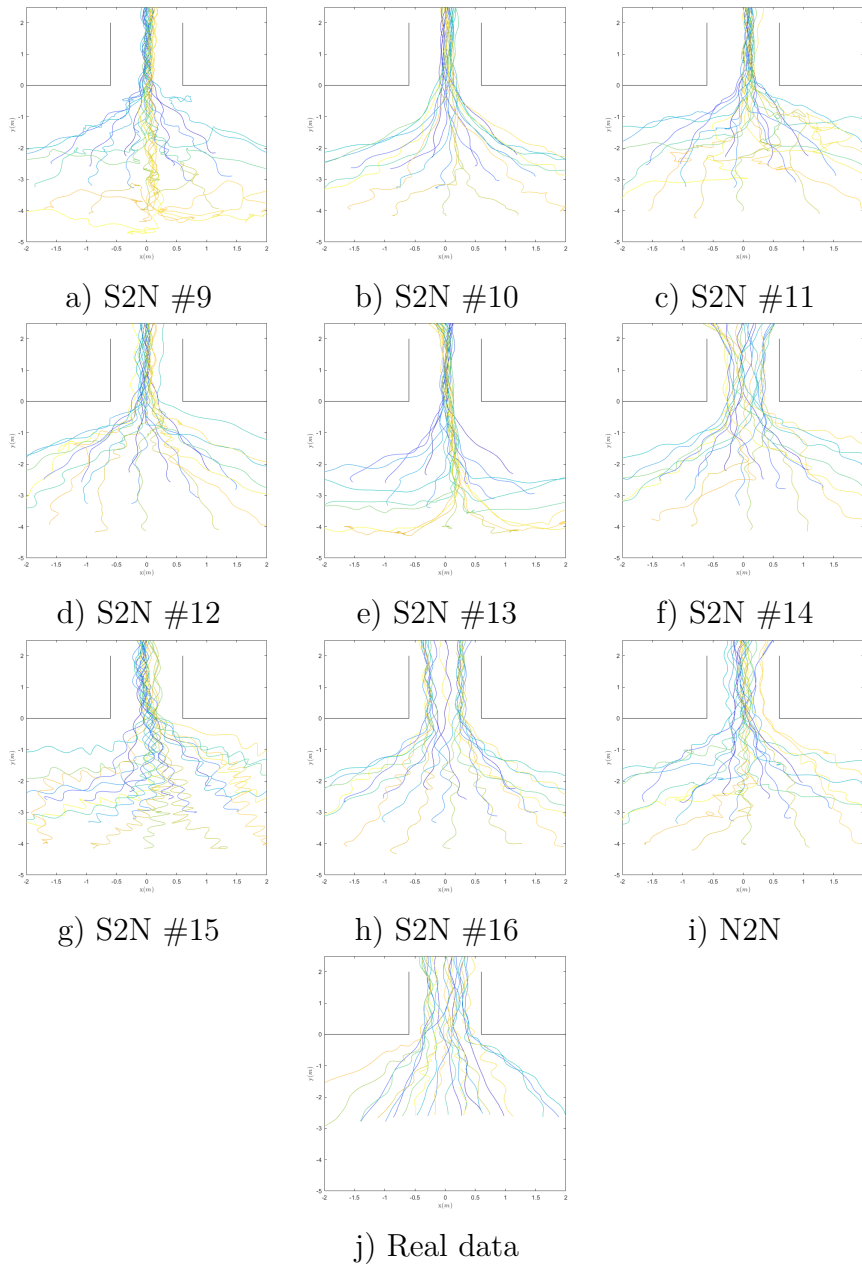


Figure 3.9 – Bottleneck scenario: Comparisons of lane formations in different conditions. Trajectories are colored by the order in which participants acted as virtual pedestrians (see Figure 3.3). We remark that the result #9 illustrates a typical situation where the lane formation can be clearly observed, whereas #16 illustrates another specific pattern where all the virtual characters converged on a single lane. The real data refers to the work of Liddle et al. [2009].

**Bottleneck fundamental diagram.** Fundamental diagrams capture relations between density and flow, but flow in a bottleneck linearly depends on its width [Seyfried et al. 2009]. To make the diagram independent to the bottleneck width, we study the *specific flow*  $J_s$  (as defined by Seyfried et al. [2009]) which represents the flow per unit of space, computed as the product of density  $\rho$  and the flow speed  $v$ :

$$J_s = \rho v \quad (3.5)$$

The estimation of  $\rho$  and  $v$  is performed on a section of the corridor ( $y \in [0.5 \text{ m}, 1.5 \text{ m}]$ ).  $\rho$  is computed according to the method proposed by Steffen and Seyfried [2010] (that guaranties continuity with respect to time):

$$\rho(t) = \frac{\int_A P(\mathbf{x}, t) d\mathbf{x}}{|A|} = \sum_i \frac{\int_A P_i(\mathbf{x}, t) d\mathbf{x}}{|A|} \quad (3.6)$$

where  $A$  represents the measurement area,  $P(\mathbf{x}, t)$  is the probability density of having a pedestrian at the position  $\mathbf{x}$  at the time  $t$ ,  $P_i(\mathbf{x}, t)$  is the corresponding probability density for a pedestrian  $i$  at the time  $t$ , and is supposed to be uniform on the Voronoï cell of the pedestrian. With this definition,  $N(t) = \rho(t)|A|$  denotes the number of pedestrians in the considered section of the corridor and is continuous even when an agent enters in or exits the section. We then compute the flow speed by a weighted averaging method as follows:

$$f(t) = \sum_i \frac{\int_A P_i(\mathbf{x}, t) d\mathbf{x}}{N(t)} v_i(t) \quad (3.7)$$

where the speed  $v_i$  of the pedestrian  $i$  is numerically derived from its position. We present the fundamental diagram resulting from specific flow against local density in Figure 3.11. This figure shows that the range of speeds participants walked at are similar between S2N and N2N. The specific flow values are also clearly lower for OMC than real data, confirming the effect of a 1-lane traffic where a 2-lane one was expected, resulting into a loss of flow. We also consider that walking speed, known to be slower in VR, may have accentuated this loss of flow.

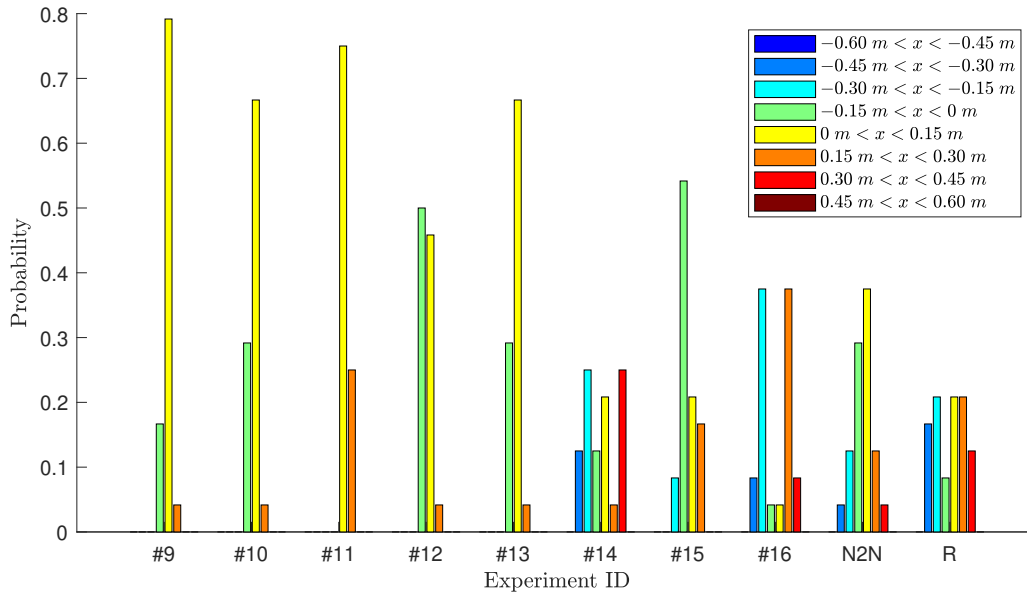


Figure 3.10 – Probability to finding a pedestrian at position  $x$  inside the exit measurement zone  $-0.6\text{ m} \leq x \leq 0.6\text{ m}$ ,  $0.5\text{ m} \leq y \leq 1.5\text{ m}$ .

### 3.5.2.2 Discussion

In this second experiment, we attempt to validate the idea of generating a crowd walking through a bottleneck with the OMC paradigm. We selected a bottleneck width where, in real experiments, a 2-lane traffic starts emerging (the full real experiment studied corridor width as a dependent variable). Our attempt was not always successful: few participants formed 2 lanes, most only 1 lane, and we could observe specific queuing up behaviors. We believe that the non-emergence of a 2-lane traffic can be explained by perceptual and variety biases: i) some participants may have perceived the corridor to be smaller and considered that only a 1-lane traffic would fit, and ii) the S2N procedure resulted in the same decision made by the participant iteration after iteration, and sometimes resulted in specific trajectories like the one showed in Figure 3.9-a. This validates hypothesis **H1**, showing quantitative differences between real and OMC virtual data at the individual level. However, we believe that this effect could be further studied based on previous work on passability judgement tasks in VR [Geuss et al. 2010]. Indeed, the question of the number of lanes fitting a given corridor width could be turned into one of passability into a gap. It might therefore be relevant to replicate the full experiment of Liddle et al. [2009] with corridor width as a dependent variable, so as to observe transition width (from 1 lane to 2 lanes for example) in a VR context. As expected from **H2**, N2N

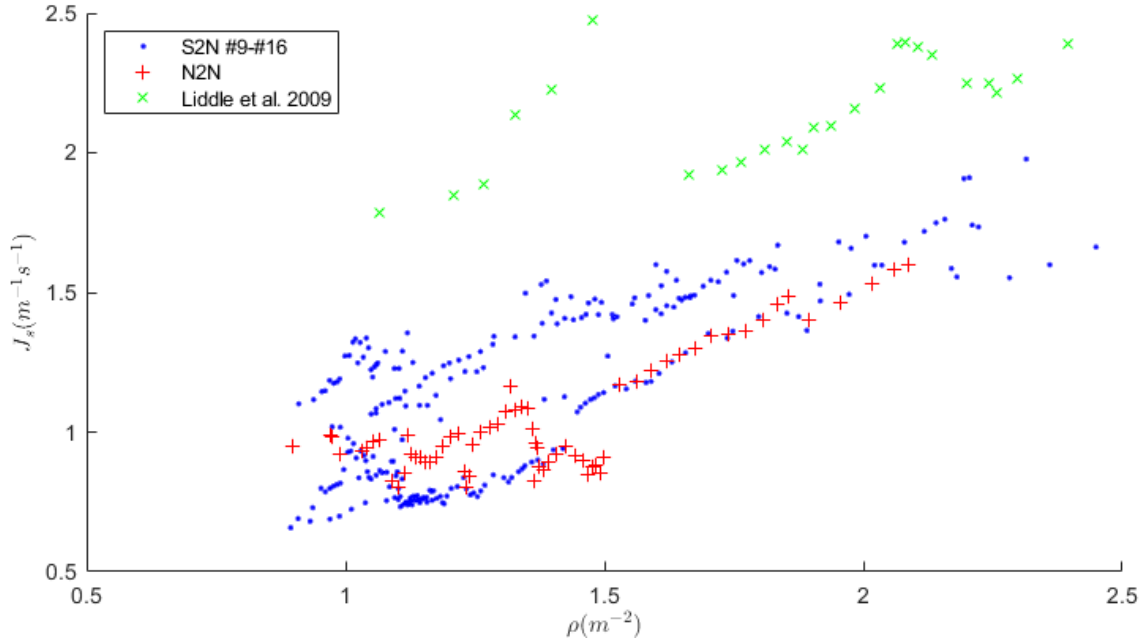


Figure 3.11 – Estimated specific flow with respect to instantaneous local density.

prevents this individual-specific decision on the number of lanes to form to be repeated iteration after iteration. Our interpretation is that, all together, participants varied the decision to form a 1-lane or a 2-lane traffic, which resulted in a uniform distribution of lateral positions. **H3** is therefore only partly validated, as the expected pattern could be observed in some iterations, but this was not always the case.

Additionally, it is quite clear that the fundamental diagram analysis enables observing the large quantitative differences in the measured flows: the absence of emergence of a second lane in most cases completely offsets flow values in similar densities. This highlights the need for awareness of quantitative differences of experiments performed in VR. Nevertheless, we also observed a consistency between the S2N and N2N results, as well as that the relation between specific flow and density is preserved compared to real world results (flow decreases when density increases, with similar slopes).



### 3.5.3 Scenario III - Inflow

Different from the two previous experiments where we study moving crowds, we here focus on entering into a virtual confined space by replicating Ezaki et al.'s experiment [2015] and investigate how participants decide which position to occupy. While previous experiments explored motion interactions with predecessors, here we explore more how people interact with their environment and the whole crowd: people pick the position they would prefer, also being conscious that a number of people have to fit. We here present the corresponding analysis in terms of participants' final positions and personal space.

#### 3.5.3.1 Analysis & Results

**Final positions.** Final positions picked by participants are illustrated in Figure 3.12, along the corresponding Voronoï diagram, for the 9 S2N examples (from #17-#25), the N2N and a real data example from [Ezaki et al. 2015]. The figures show both the order in which the positions were picked, and the personal space participants preserved around them. The personal zones are colored by their area.

Following the analysis performed by Ezaki et al. [2015], we can further analyze the position picking strategy by representing positions according to the polar coordinate system  $(r_i, \theta_i)$  centred at the entrance. For each direction  $\theta_i$ , we calculate the distance  $r_{\max}(\theta_i)$  between the pole and the wall in the corresponding direction. As the maximum distance is different in each direction, we compare the normalized distance  $r_{n_i} = r_i/r_{\max}(\theta_i)$ , to figure out how far, between the entrance and the wall, participants went. Figure 3.13 shows the normalized distance to the walls regarding the order of entrance. In comparison to OMC, the real results show a higher preference towards positions close to the wall. The more noisy N2N plot (but with no repetition) exhibits more varied strategies, and little influence of the walls.

**Personal space.** To further explore individual behaviors, we computed the final personal space for each agent, which is here defined as the area of each agent Voronoï cell (see Figure 3.12). The result is presented in Figure 3.13-b, with personal space presented with respect to the order of entering.

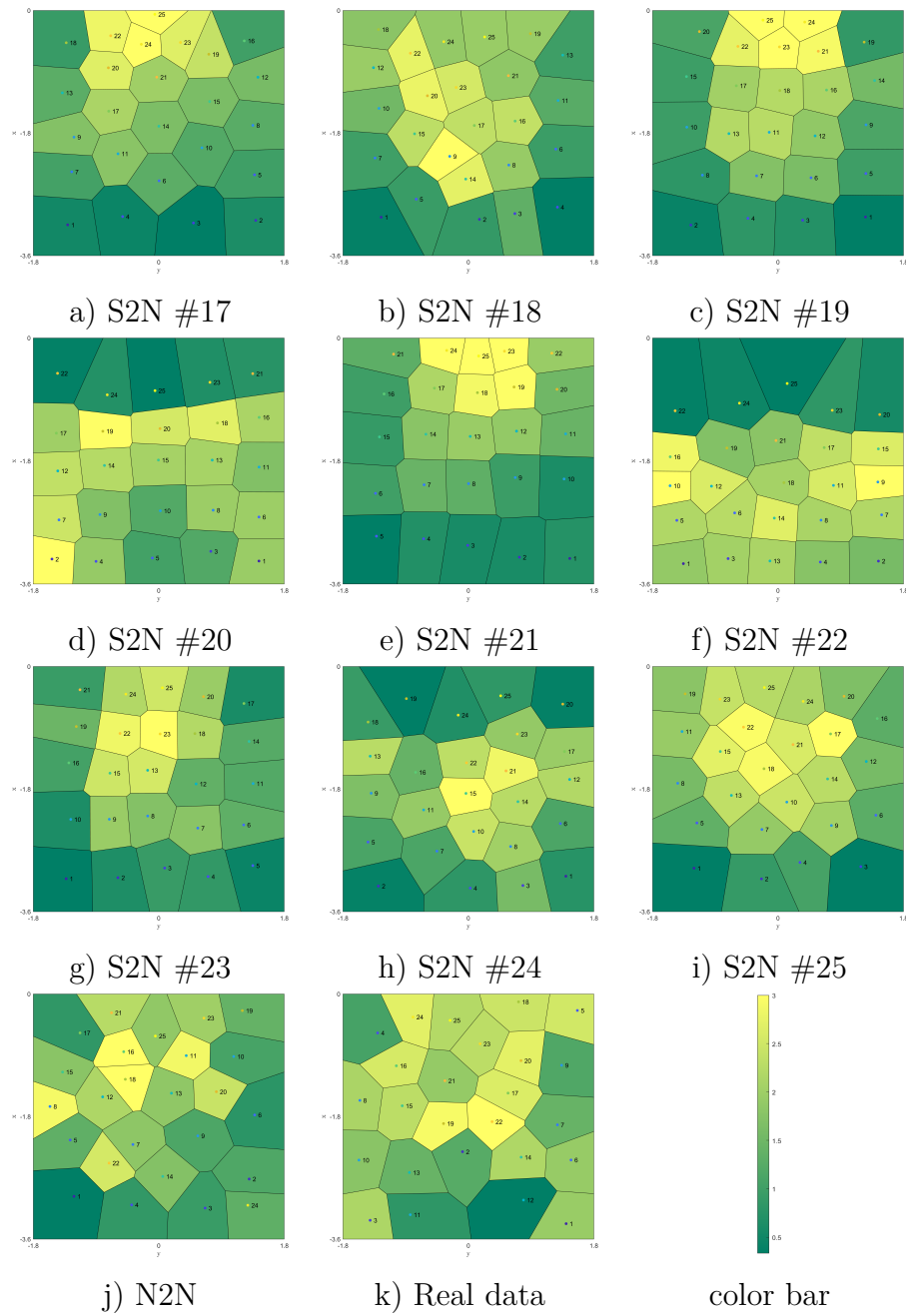


Figure 3.12 – Inflow behavior scenario: final positions of each pedestrian and personal space. Final position for different experiences. The entrance of the lift is located on top-middle. Each cell is colored by the local density of each position, truncated at  $3 \text{ ped.m}^{-2}$ . The real data refers to the work of Ezaki et al. [2015].

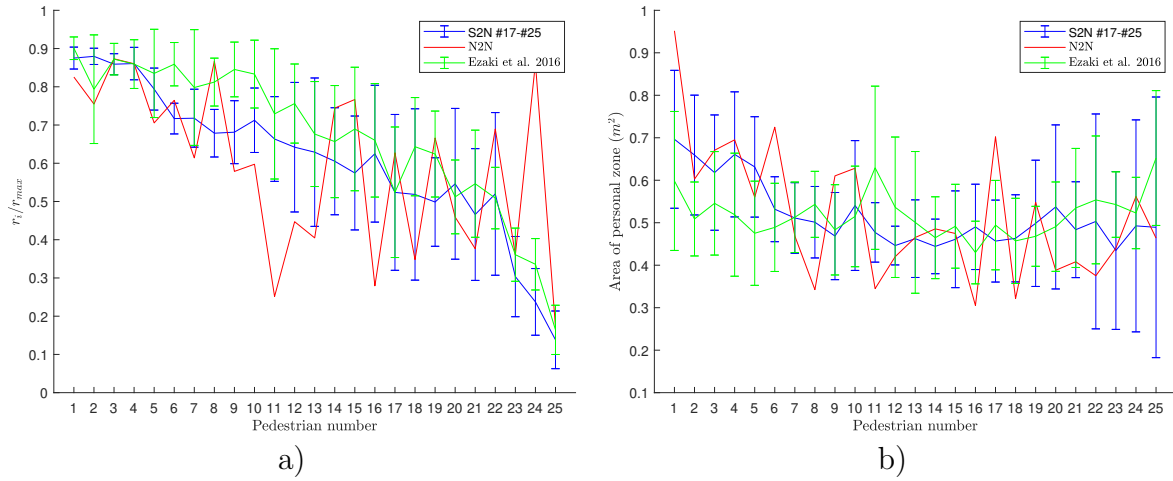


Figure 3.13 – Evolution of a) how far pedestrians walked along the direction of their final position and b) personal space area, with respect to the order of entering, for S2N, N2N and real-world results, respectively averaged over 8, 1 and 9 experiences.

### 3.5.3.2 Discussion

In this experiment, we explored the strategy of picking a position in a virtual room when being filled by a crowd. By analyzing the position picking order, it is clear that emergent patterns appear, but differ between real and virtual ones. Pedestrians in the real-world experiment chose their position in a sequence of U-shape formations with outer, middle and inner layers (Figure 3.12-k). In contrast, pedestrians in our OMC experiment would rather fill the room from the back to the front in a succession of flat rows, more or less organized depending on participants (e.g., Figure 3.12-e and -g). Looking at all OMC results, the corner area near the entrance seems to be rarely used, which differs from real-life scenarios. This difference can be explained by hypothesis **H1**. We have two possible reasons in mind. First, real and virtual walls play different roles: real walls can for example be used to lean on, which is obviously not the case of virtual ones. This can explain the preference to stand close to them first in real experiments (U-shape formations). Moreover, in the OMC procedure, participants do not have to exit the room at the end of the iteration: the VE disappears, and they move back to the initial position. This might decrease the importance of standing close to the entrance (that a U-shape formation will give). We also observed that OMC participants stood further from walls than in reality, suggesting that the perception of the distance to objects in VR might have had quite an important effect on final positions, at least those close to walls.

More generally, the evolution of the virtual pedestrians’ personal space (Figure 3.13-b)

indicates that during each S2N experience, participants usually left themselves a larger personal space than in real-life. Furthermore, once they found out that the first virtual agents had occupied a too large space for everyone to fit comfortably, they would overreact by enormously sacrificing their personal space. This is responsible for the fast dropping of personal area in the S2N results, and the relatively large personal space in the N2N result, which is not observed in real experiments.

The lack of variety also led to specific emergent patterns, validating **H2**. This can be seen in Figure 3.12-e and -g, where the participant clearly applied a strict strategy to fill the room, ending up in a very regular squared pattern. The N2N procedure successfully breaks this regularity (Figure 3.12-j), and we can see some irregular strategies, such as participant #24 picking the corner position in the back of the room whereas other last participants more preferably picked positions next to the door. Nevertheless, emergent patterns resulted from the OMC procedure, validating **H3**.

## 3.6 General Discussion

Combining the results and analyses from three experimental scenarios, we provide a general discussion of several important aspects. First, we evaluate the research methodology employed in this chapter. Then, we discuss the similarities and differences between S2N and N2N. Finally, we present our thoughts and deliberations on topics beyond the primary focus of this chapter, including animation, interaction, and scenario choice.

### 3.6.1 Methodology

In this chapter, our methodological approach consisted in replicating several existing experiments in VR with our OMC paradigm. We relied on existing analysis frameworks typically used in the crowd modelling community to evaluate our results according to common practice from this community. While this involved relying on qualitative evaluations, we believe this approach was more relevant to provide a first evaluation of the OMC paradigm. Apart from some limitations discussed below, our conclusion is that we *successfully* replicated three experiments that involve, mostly, *unilateral interactions*. Successfully, because we were able to replicate the observation of some existing, already known and observed, emergent phenomena. This is a major difference with previous studies in VR on collective behaviors that focused on individual analyses of local variables

(e.g., [Bruneau et al. 2015; Rio et al. 2018; Ríos and Pelechano 2020]). It is also clear that the biases we expected influenced to some extent the observed emergent phenomena. However, they mostly affected the characteristics of those phenomena without preventing their natural emergence. We were also able to observe which characteristics were affected and how, and could provide interpretations based on some formulated hypotheses.

While our analyses relied on *qualitative analyses*, we also believe that *quantitative analyses* will be important to explore in the future. For instance, a quantitative analysis would ideally require to perform again the real experiments with the same group of participants experimenting OMC. Given that our goal was to evaluate the potentialities of the OMC paradigm, the effort required to conduct such a real experiment would have prevented us from exploring the variety of scenarios we could cover here. At this stage, it would also be relevant to quantitatively evaluate the effect of several specific biases on our results. One interesting direction would be for instance to replicate studies on the effect of known dependent variables (e.g., age, density, corridor width). E.g., does the effect of age on 1D-flow can be reproduced using OMC? While characteristics of emergent phenomena are biased, would the effect of dependent variables still be observed and exploitable?

### 3.6.2 S2N vs. N2N

Our results suggest that the variety bias seemed to have influenced the most the characteristics of emergent collective phenomena. More specifically, the S2N procedure led in some cases to a replication across all agents of some specific individual behavior, that could be observed as a sort of pattern exaggeration. Such extreme observations, which were reported in the discussion of each experiment, included the formation of single or extremely distinct lanes (Figure 3.9-a and -h), or entering the virtual room in an overly-organized manner (Figure 3.12-e and -g). Since N2N solves the effect of this bias efficiently, we could claim that OMC-N2N is the only viable solution here. However, S2N is more convenient in practice because N2N still requires involving many participants to create new crowd motion datasets. Therefore, to further explore the potentialities of OMC-S2N, we still have two main questions opened to be addressed in the future. First, would the average of data resulting from multiple S2N datasets be equivalent to a N2N dataset? In other words, could we deduce the characteristics of a phenomenon emerging from interactions between various people given multiple ones emerging from interactions with oneself? The second is about possibilities for a M2N procedure, where  $M$  participants would control  $N$  agents, while keeping  $M$  as small as possible compared to  $N$  in order to

keep the advantage of recruiting few participants to generate large datasets. Future work would then investigate what would be the required  $M$  value to cancel the effect of the variety bias.

### 3.6.3 Character Animation Details

OMC agents replicate participants' behaviors in a limited way. E.g., we captured their full-body motions using a motion capture suit, but neglected their gaze activity, which is known to play a great role in human interactions, while difficult to study in practice. Other modalities, like sound, were not recorded. Therefore, we would like to explore the possible benefit of capturing more detailed behaviors and incorporating them in the VE, as well as to explore novel ways of making participants aware of the physical presence of characters (e.g., evaluating the benefits of rendering contacts using haptics [Berton et al. 2020a]). Another option would be to edit recorded behaviors, for example to change agents' motion style, which may provide another solution to lower the effect of variety bias. This would also raise novel questions, e.g., would that prevent participants to understand that they are interacting with themselves? Would a single participant behave in a more varied way given that he/she visualizes more varied motions?

### 3.6.4 Generalizing Interactions

Did participants realize they were interacting with themselves? The gross answer is that half of them reported yes, while this proportion depended on scenarios (all participants for 1D-flow, 2 for bottleneck, and 3 for inflow). For them, it was clear that interactions would be limited to unilateral ones, and that visible agents would never respond to their actions. While this is a limitation of OMC that will be difficult to overcome, we believe a number of solutions can be explored in the future to generalize the type of interactions that can be performed with OMC. The first one is to decompose interactions into multiple sections of time where participants control alternatively agents, similar to what Osimo et al. [2015] did for building conversations with one-self (which was one source of inspiration for OMC). The second solution would be to perform the S2N procedure multiple times. For example, in a dual interaction between agents A and B, let the participant controlling A, then B to unilaterally interact with A, and then A again to interact unilaterally with B, etc. This raises new important questions, such as whether multiple unilateral interactions can converge (or not) toward the result of a bidirectional interac-

tion. The third solution is to use a collision avoidance techniques to override previously recorded trajectories when a character is on a short-term collision course with the user. This consists in making a local and interactive adaptation of the trajectory, only when necessary, to avoid unrealistic situations. This solution has however a clear drawback: as mentioned in the beginning of this paper, modifying trajectories in such a procedural manner might reduce the realism of the generated dataset. With three possible choices, we believe that those solutions will enable us to generalize the use of OMC for other types of scenarios, e.g., studying collision avoidance strategies, large-crowd interactions, etc.

### 3.6.5 Generalizing Scenarios.

As previously mentioned, our study prioritized the replication of crowd scenarios mainly involving unilateral interactions. However, daily situations would include multidirectional interactions as well, which makes scenarios extending beyond unilateral interactions our next target. Intuitively, we anticipate new artifacts can result from those scenarios. Thus, we have conducted a preliminary experiment exploring them to gain insights about their real nature and importance. We present a detailed description of these preliminary attempts in the following section (3.7).

## 3.7 Beyond OMC: A First Step Towards Multidirectional Interactions

So far, our study in this chapter has validated the utility of the OMC method by replicating three real-world experiments and comparing the results, thereby demonstrating its capability to construct valuable datasets. Following the confirmation of OMC’s performance in the three unidirectional crowd scenarios discussed earlier, it was a natural progression to explore its efficacy in more complex scenarios, specifically within multidirectional crowd movement contexts. This endeavor aims to investigate the performance and limitations of the OMC method in multidirectional crowd movement contexts, thereby providing insights that will form the rationale for the more complete approach to be presented in the next chapter. We employ the same experimental approach as outlined in Section 3.2.2, attempting to construct three multidirectional crowd movements.

To this end, we set up and explored the 3 following preliminary research questions:

- **Q1:** In multidirectional crowd scenarios, a participant needs not only to react to

the virtual characters in the front, but also to pay attention that virtual characters may come from his lateral direction, and even from behind. In this case, can the participant generally complete the creation of a crowd scenario without colliding with virtual characters?

- **Q2:** During the experiment, as the participant successively embodies different characters, he/she will have to interact with more and more moving characters. Will this continuously increase the difficulty for the participant to perform collision avoidance? Will the participant have to gradually increase the complexity of his movements?
- **Q3:** The difficulty of generating multidirectional crowds can be influenced by both the number of interactions and their directions. Do these two factors have an equal level of impact on the generating procedure?

### 3.7.1 Experiment

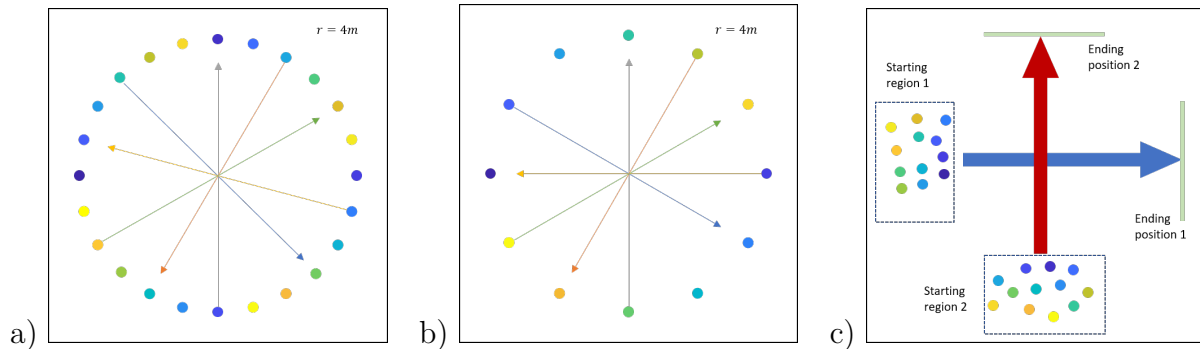


Figure 3.14 – Configuration of the preliminary experiments. Black squares delimit the  $10 m \times 10 m$  experimental ground. Each dot represents the initial position of a character, indicating the starting order. The first to start is colored by dark blue, while bright yellow colors the last. The arrows indicate the direction of each character. a) Configuration of Circle-24 and b) Circle-12, where participants are instructed to walk to the position diagonally opposite in the circle. c) Configuration of the crossing flow virtual experiment, the initial positions in each starting zone are generated by a Poisson distribution.

Following the same experimental framework, we designed a preliminary experiment to replicate three multidirectional crowd motion scenarios listed below. As we were interested in preliminary insights about how OMC would behave with multidirectional interactions, and also due to the time constraint, only one naive participant participated in this preliminary study, and performed the three scenarios. Our main focus is to explore the nature and



importance of artifacts we may encounter in multidirectional interactions, in preparation for larger scale experiments. For the same reason, our analyses are not statistical.

The 3 virtual scenarios that we designed to explore multidirectional crowd movements are: one crossing flow situation and two circle crossing movements with different numbers of characters but 1 participant.

**Circle-24.** This scenario was designed to study multidirectional interactions. In this scenario, 24 characters are uniformly spaced out on a circle (radius: 4 m – see Figure 3.14-a). The goal of the participant is to reach the opposite side of the circle, without colliding with any other character. Participants are asked to act all the 24 characters present in the circle, starting from the one on the position  $(-4, 0)$ . Then, the participant always acts the 6th neighbor on their right-hand side, skipping it if it was already embodied.

**Circle-12.** This scenario was designed as a controlled experiment for Circle-24, to study the effect of density. The task is the same as in the Circle-24 scenario, except that the number of virtual characters is reduced to 12 so that each virtual character can have more space for its movements (see Figure 3.14-b). The starting order is decided in the same way as Circle-24.

**Crossing Flow.** This scenario was designed to study the performance of OMC during crossing flows. In this scenario, 24 virtual characters are equally divided into 2 groups. The two flows cross around the center of the virtual scene at a 90-degree angle (see Figure 3.14-c). The participant is asked to act alternatively one character of each group, until he has acted all 24 characters. The participant is instructed to traverse the virtual room and to reach the ending position, while avoiding any collision with the other virtual characters. The participant alternates iterations involving characters starting from a different starting region (i.e., from one of the two groups of characters). For each group, the starting positions were randomly generated by a Poisson distribution, and the virtual characters are acted by participants in the order of increasing initial distance towards the center of the scene.

### 3.7.2 Test Scenario A: Circle (Circle-24 & Circle-12)

To maximize the use of collected data, we adopted a cross-comparison method to extract as much information as possible from the experimental results. As Circle-12 is a

simplification of Circle-24, we present their results together.

### 3.7.2.1 Trajectories

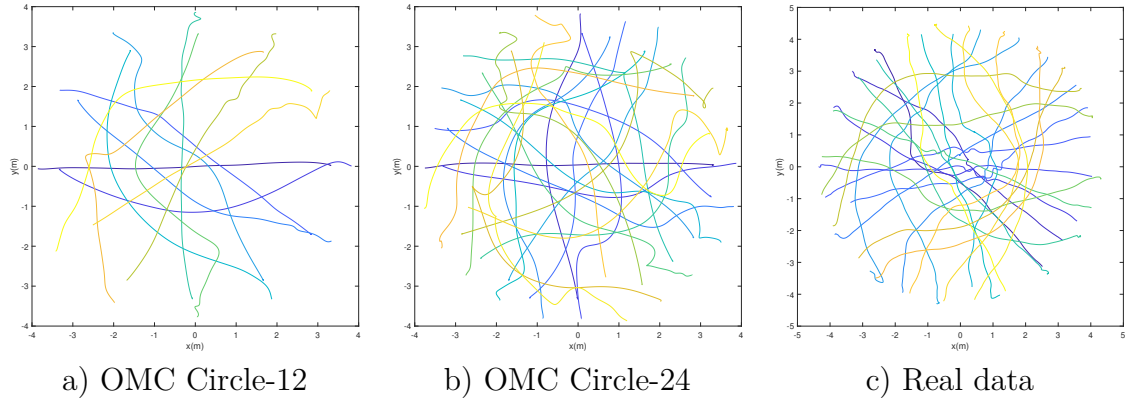


Figure 3.15 – The generated trajectories, colored regarding the starting order, from dark blue (first to move) to bright yellow (last to move).

In Figure 3.15 we compare two sets of trajectories generated with OMC (Circle-12 and Circle-24), with data from real experiments.

**Circle-24.** From the trajectories displayed in Figure 3.15-b, it appears that the participant adopted two entire different strategies. More specifically, we can see that in early iterations he used a strategy going through the center of the circle, while in later iterations he adopted longer trajectories, enabling him to circle around the center and therefore to limit the quantity of interactions in this area. We hypothesize that the difference in strategies is due to the increase in collisions happening with the increase in characters moving in the scene in such a complex situation. However, it is also important to mention that both types of strategies were also observed in real experiments (Figure 3.15-c), where we can see that some agents (possibly moving faster) went almost directly to their destination through the center of the circle, while others (possibly starting later) displayed more indirect trajectories.

**Circle-12 vs. Circle-24.** The Circle-24 scenario was supposed to be more difficult than the Circle-12 scenario with doubled density. We can see that the participant adopted similar trajectories for both scenarios, even though with fewer trajectories going around the center in the Circle-12 situation, as can be observed from Figures 3.15-a & -b.

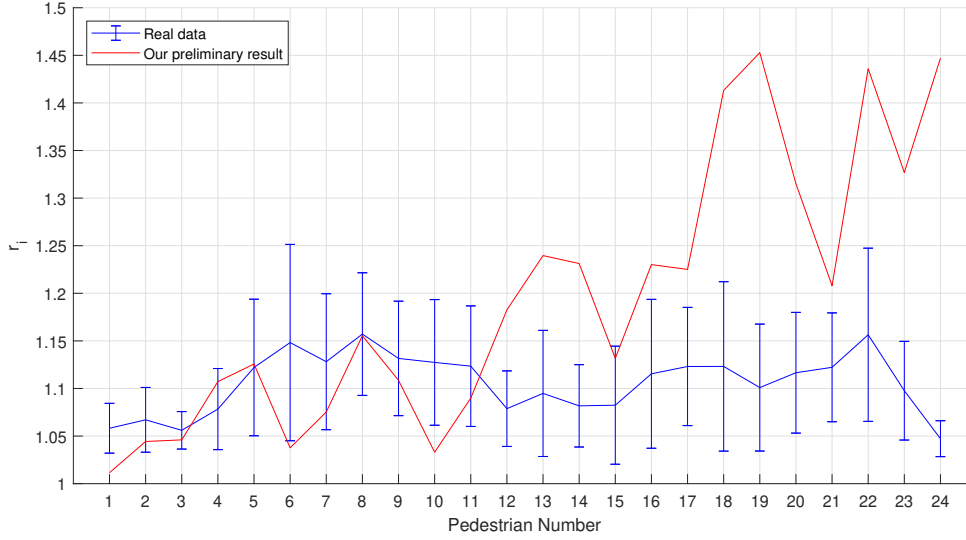


Figure 3.16 – The behavioral complexity of the user when acting as different characters, illustrated by the ratio between the distance that the participants walked and the distance between the starting and ending positions, depending on the iteration number, compared to real-world result. The real-world result is averaged over 21 data. Note that in the real-world case, the pedestrians walked simultaneously, and their number are not order-dependent. In our experiment, on the contrary, the pedestrian number the same as the iteration number.

### 3.7.2.2 Behavioral Complexity

As shown in Figure 3.16, the behavioral complexity continued to increase during the experiment. An important common phenomenon of the three experimental results is that the first pedestrian always walked directly towards his goal, leading to an almost straight first trajectory (dark blue horizontal trajectories). As the number of iterations increases, the global shape of trajectories becomes more and more complex. We can estimate behavioral complexity of each iteration  $i$ , through the distance ratio  $r_i$ :

$$r_i = d_i / \|\mathbf{p}_i\| \quad (3.8)$$

where  $d_i$  is the distance that the participant walked through during the iteration  $i$ , and  $\mathbf{p}_i$  is the vector from the starting position and the ending position. We present the evolution of behavioral complexity in Figure 3.16. As the number of moving virtual pedestrians increases, the behavioral complexity presents a clear tendency of increase, both for the Circle-12 and Circle-24 scenarios.

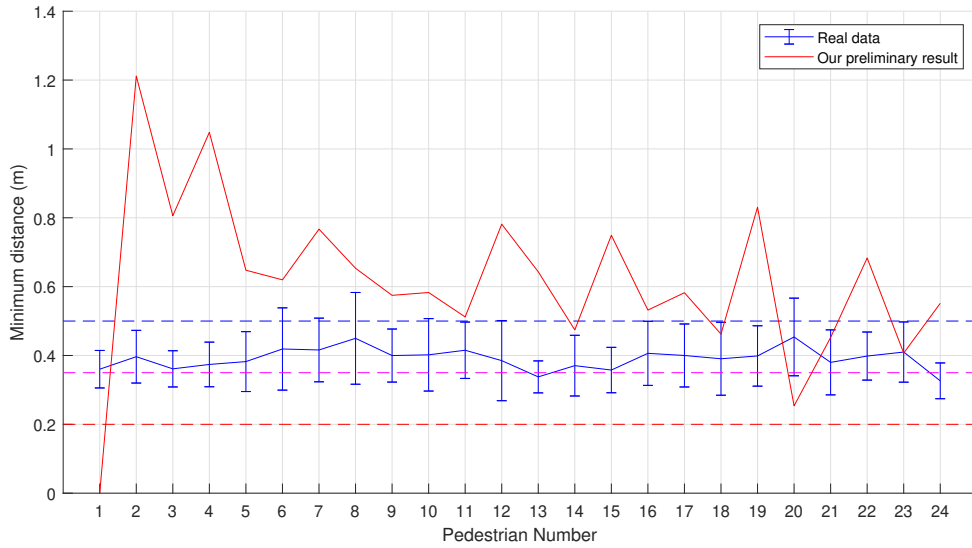


Figure 3.17 – The shortest distance between the user and any pedestrian for each iteration of the experiment. The real-world results are averaged over 21 data. The dashed lines highlight 3 important levels of distance. The blue dashed line indicates the distance where 2 pedestrians are very close. The magenta dashed line indicates the emergence of a slight touch. For the real-world condition, the red dashed line indicates collision, whereas in the virtual environment, this would indicate severe interpenetration between the participant’s avatar and a virtual character

### 3.7.2.3 Collisions

We detected collisions between the participant and the virtual characters through post-processing of the generated data. Specifically, we calculated the shortest distance between the participant and every moving characters during the iteration time. As shown in Figure 3.17, the participant was generally successful in keeping a relative large distance to all the moving characters and only had slight touch with them during the iteration 20, where he needed to react to 19 moving characters. Severe interpenetration was not observed (geometries did not fully interpenetrate).

### 3.7.2.4 Discussion

Through this exploratory experiment of potentially going beyond the limitations of OMC, our goal was to further explore 1) if participants can successfully complete the task without colliding with characters, 2) how increasing complexity of the scene might influence the participant’s strategies, and 3) how the complexity of the scenario might influence the participant’s behavior as well. Even though we cannot draw any decisive

conclusion due to the preliminary nature of the experiment, the trajectories captured with OMC highlight interesting features and tendencies which are consistent with real observations. We observed that the participant adopted both direct (i.e., going through the center) and indirect (larger curved trajectories, avoiding the center) for both Circle-12 and Circle-24. Of course, this choice of strategies might be relative to participants' individual preferences.

In terms of the behavioral complexity, we observe that the participant made more and more complex behaviors during the procedure. A naive explanation of the situation is that, as more and more virtual pedestrians moved in the scene, he needed to make more and more effort to avoid colliding with pedestrians from each direction. Meanwhile, the occlusion among the characters made it more difficult to predict which character was going to be where. Surprisingly, even with such difficulties, the participant accomplished the experiment without barely colliding with any virtual character. Although more participants would be required to conclude whether the answer to **Q1** can be positive, this preliminary result suggests that the OMC paradigm can handle scenarios of multidirectional movement, including complex ones (**Q3**). The observation of Figure 3.16 is likely to lead us to a confirmatory response of **Q2**.

### 3.7.3 Test Scenario B: Crossing Flow

Similarly to the Circle-12 and Circle-24 scenarios, we analyze the trajectories and behavioral complexity.

#### 3.7.3.1 Trajectories

Figure 3.18 shows the trajectories generated by the participant. Different from real-world situation, the participant has shown a strong preference for the strategy of detouring the center of the virtual scene (i.e., the friction zone) by adopting large curved trajectories going through the edges of the scene. We can observe that in the real-world experiment, the shapes of the two flows were basically consistent during the whole crossing. Although the large motion strategy in our result can also be observed in the real case, its emergence is less frequent and the motion is less amplified. We believe this was because the participant had prioritized the task of collision avoidance, and thus adopted strategies which granted him more time to predict the agents' trajectories and more space to act in avoidance movements.

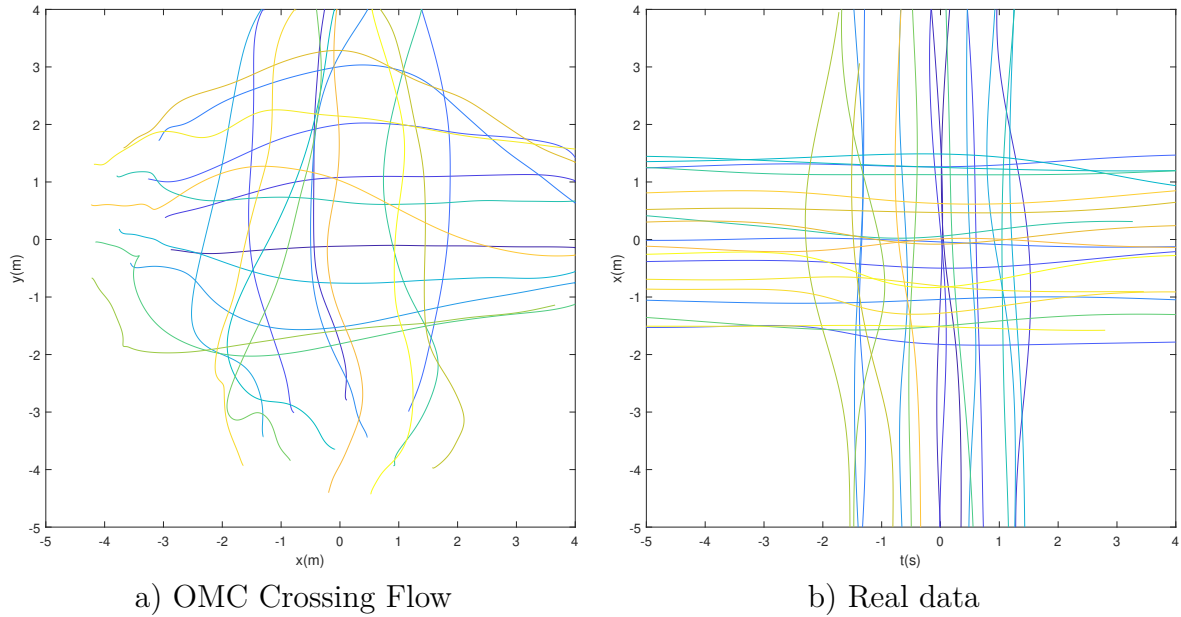


Figure 3.18 – Generated trajectories for the Crossing Flow scenario, colored with respect to the starting order, from dark blue (first to move) to bright yellow (last to move). a) Results generated by the participant using OMC. b) Real-world result.

### 3.7.3.2 Behavioral Complexity

As previously, we estimated the participant’s behavioral complexity during each iteration using Equation 3.8. The results are presented in Figure 3.19, and show that the behavioral complexity stayed at a low level for the first four iterations. Afterwards, it evolved in a peak-valley manner, while slowly increasing on average.

### 3.7.3.3 Collision

Again, we calculated the shortest distance between the participant and any moving character during each iteration. A close touch was detected during iteration 16 (Figure 3.20) where the participant needed to react to 15 characters (8 characters coming from the lateral side, as well as 7 characters moving in the front). This might be a chained reaction after what happened during iteration 15, where we observe a peak of behavioral complexity (Figure 3.19). Interestingly, no other collision happened during the following iterations.

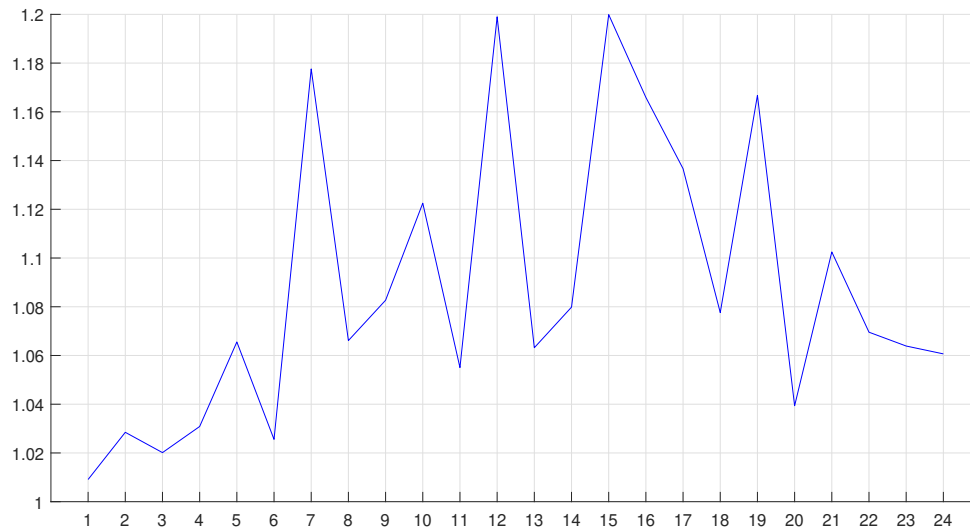


Figure 3.19 – The behavioral complexity of the user when acting as different characters, illustrated by the ratio between the distance that the participants walked and the distance between the starting and ending positions, depending on the iteration number, for the Crossing Flow scenario.

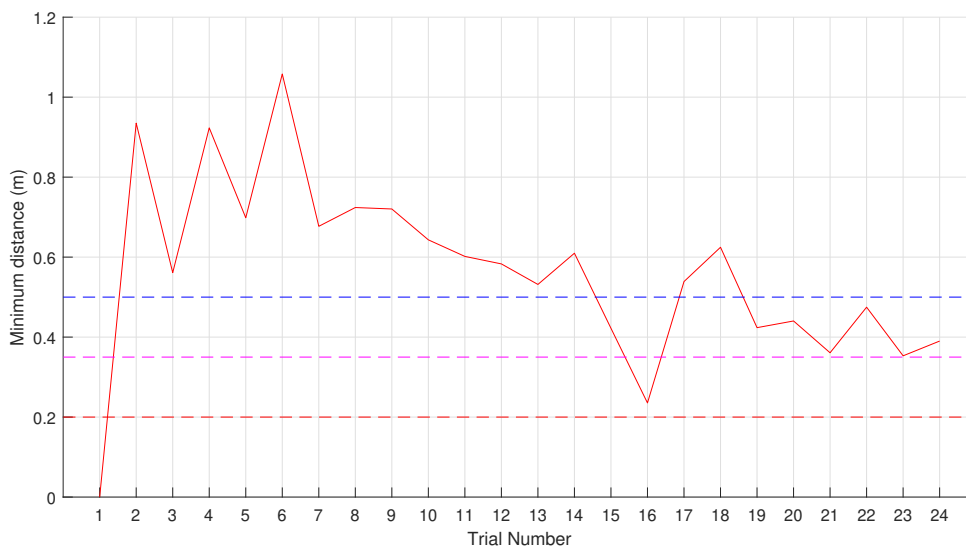


Figure 3.20 – The shortest distance between the user and any pedestrian for each iteration of the experiment. The dashed lines highlight 3 important levels of distance. The blue dashed line (50 cm) indicates the approximate shoulder distance. The magenta dashed line indicates the emergence of a slight touch. For the real-world condition, the red dashed line indicates collision, whereas in the virtual environment, this would indicate severe interpenetration between the participant’s avatar and a virtual character

#### 3.7.3.4 Discussion

In the crossing flow scenario, multiple agents walk through the center of a virtual scene from two fixed directions, at a  $90^\circ$  angle. The participant therefore needs to react to the movement of the agents in front of them, as well as to those coming from their lateral side. As the participant's moving speeds are limited by the agents in front of him, different agents will arrive at the crossing zone at different time and will there need to perform local collision avoidance. This might explain the shape of the participant's behavioral complexity plot shown in Figure 3.19. Meanwhile, the difficulty of collision avoidance increased during the generation of more agents' movements. However, the participant only experienced one light collision during the experiment, and its influence remained to be local. One possible explanation for this small number of collisions is that the participant chose the strategy of large movements and detouring the friction zone through the edge of the virtual room. The participant's performance gave out a positive response for **Q1**. For **Q2**, the difficulty of collision avoidance did increase, but the participant still managed to finish the experiment with barely any collision. The complexity of his behavior, however, did not continuously increase.

#### 3.7.4 Exploratory Insights

In this section, we designed 2 different scenarios which involved multi-directional interactions, using the same OMC-S2N setup. The first scenario consists of a *Circle*, where participants stand on the perimeter of a circle and are asked to reach the antipodal position on the circle. We performed this scenario under two different density levels (12 or 24 characters around the circle). The second scenario consists of a  $90^\circ$  *Crossing Flow*, where two groups of 12 characters each cross at a  $90^\circ$  angle. We compared both the OMC results to real-world recordings of similar situations. Our preliminary observations are the following. In multidirectional interaction OMC scenarios, we observed that the participant displayed strategies that are visually similar to real-world ones. For example, users chose to cross the circle either by going straight through the circle center, or by going around the center, avoiding this area of higher friction between people. In other words, they used strategy consisting in playing on speed or curvature of their trajectory to avoid collisions. We also observed some specific behaviors. For instance, in the *Crossing Flow* scenario, we observed that the participant often passed through the scene's edge to avoid interactions with other characters. This was not observed in real situations. Considering



that the results were generated by only one participant, we still need to conduct further formal experiments to investigate whether such observations are systematically related to the OMC paradigm and the scenarios.

Noticeably however, our experimental results suggest that even though introducing multidirectional interactions makes the participant’s task more difficult, it is still possible to create such crowd movements using the One-Man-Crowd paradigm. In both cases, the participant finished the scenario without interpenetrating (or in a very light way) with virtual agents. Knowing collision avoidance was a prioritized task, the participant put great effort in observing and predicting the agents’ movement. This leads us to a positive answer to question **Q1**, but with a condition, that the participant was instructed to prioritize the collision avoidance. In addition, in both of the scenarios, we observed that the difficulty of keeping large distance to all the moving characters continuously increased. In terms of the behavioral complexity, we have, however, observed that, in the scenario of *Crossing Flows*, the behavioral complexity of the participant seems to be limited in a specific region. This resulted in a positive answer to **Q2**, with a nuance about the progressive increasing of behavioral complexity. Meanwhile, this might imply that the number of directions and the number of interactions do not have the same impact on the generating procedure, with the number of directions seemingly having a greater impact (**Q3**).

In terms of behavioral variety, which has been a major focus of the principal study of this chapter, we made different observations between the *Crossing Circle* and the *Crossing Flow* scenarios. In the *Crossing Circle* scenario, the participant generated a set of trajectories with various movements, which can also be observed in real cases. In the *Crossing Flow* scenario, the lack of behavioral variety has led to an individual-specific result, which is unrealistic. It is thus interesting to evaluate the use of local adjustments for the movements that have already been recorded, to improve the interactivity of the generated crowds, and reduce the frequency of generating individual-specific trajectories. More explorations need to be conducted in this direction.

## 3.8 Conclusion

In this chapter, we introduced and rigorously tested a groundbreaking crowd motion capture paradigm. This approach requires a user to sequentially impersonate various virtual characters across different iterations. Subsequently, a virtual reality platform is

utilized to display these characters simultaneously, all successively enacted by the user, within the same scene.

In the first place, to assess the viability of this novel method, and as a first investigation in this area, we selected three typical unidirectional motion scenarios from crowd-controlled experiments as benchmarks for comparison. Our experimental findings and analytical review affirm the utility of this technique. Significantly, we established that the crowd motion data generated through this paradigm bears a qualitative resemblance to real-world environments, both at the individual and collective levels.

Nonetheless, our observations revealed that some participants produced unique S2N patterns, particularly in bottleneck and inflow scenarios. Unlike the simpler 1D flow, these scenarios entail a greater degree of interactions. However, in our setup, each character only replays pre-recorded movements, which hinders any real-time interaction with the participants. After the generation of certain characters in the environment, the participants are unable to engage with the existing virtual crowd and are merely integrated passively into it. Furthermore, at the beginning of this process, when participants navigate through an empty experimental space, they tend to move directly towards their target. This behavior contrasts with real-life situations, where an individual's path is often influenced by the movements of surrounding people. Given these considerations, we posit that it is crucial to significantly enhance the interactive aspects of the One-Man-Crowd paradigm for a more authentic and dynamic representation of crowd movements.

After the first study, we conducted an exploratory experiment to gain insights about the ability of OMC to handle complex situations that involve dense, multidirectional, interactions. This exploratory experiment involved one participant generating crowd movements in three scenarios (Circle-12, Circle-24, and Crossing Flow).

Although the exploratory analysis was not conclusive due to the minimal participation in the exploratory experiment, we observed that it was possible for one human user to generate datasets covering multidirectional crowd movements without generating too many collisions. However, similar to our observation in the inflow scenario, our strong instruction on collision avoidance had again driven the participant to adopt unnatural strategies (i.e., always detouring to the center of the crossing area) and behaviors (i.e., frequent head rotation to watch pedestrians from all directions) to over compensate the non-responsiveness of the agents when generating trajectories. This implies that to acquire natural behavior, it would be necessary to alleviate the firmness of our instruction as well as addressing the non-responsiveness of the agents. We also observed specific initial

behaviors (e.g., participants walking directly to the goal during the first trial), which, despite being relevant since they were alone in the scene, would need to be mitigated to increase the naturalities of the final crowd animations.

All these results therefore suggest that OMC is a novel, relevant paradigm for simplifying the process of capturing crowd animations in VR. The following chapter will then focus on exploring potential improving of the method, by exploring the value of providing contextual crowds (i.e., providing to the user a crowd context during the initial iterations of the capture) and reactive crowds (i.e., providing the virtual crowds with the ability to respond to the user) during the crowd motion capture process.

# EFFECT OF CONTEXTUAL AND RESPONSIVE CROWDS ON VR-BASED CROWD MOTION CAPTURE

---

## 4.1 Introduction

In the previous chapter, we proposed the One-Man-Crowd paradigm as a novel approach to capturing human crowd motion datasets. We noticed that the paradigm suffers from the lack of interactivity in two main aspects. (1) At the beginning of the generation process, the users walk alone in an empty virtual scene, and tend to walk straight towards their destinations. Such behaviors are unrealistic in the real-world situation, especially in multidirectional crowd movements. In the real world, pedestrians predict each other's future path and may take action in advance (such as adapting their speed or direction) to avoid any potential collision and are not likely to walk straight (See Figure 4.2). (2) During the experience, more and more characters are added into the virtual scene, and will make it more and more difficult for the users to plan their path (See Figure 4.3). This is because the virtual characters only replay recorded motion and cannot interact with the users. When generating multidirectional crowd motion within the One-Man-Crowd paradigm, the users may get trapped by numerous surrounding characters and will experience unavoidable inter-penetration with several of them. Such phenomena may lead to an unpleasant user experience and force the users to adopt a passive strategy where they always keep large distance with the characters, and finally result in specific crowd motion patterns.

In this chapter, we seek to tackle these two limitations by further exploiting VR's capability of creating immersive virtual experience. To this end, we reconsider the idea of taking advantage of crowd simulation to increase interactivity. As mentioned in Chapter 2, numerous studies have deployed simulated crowds to conduct immersive experiments

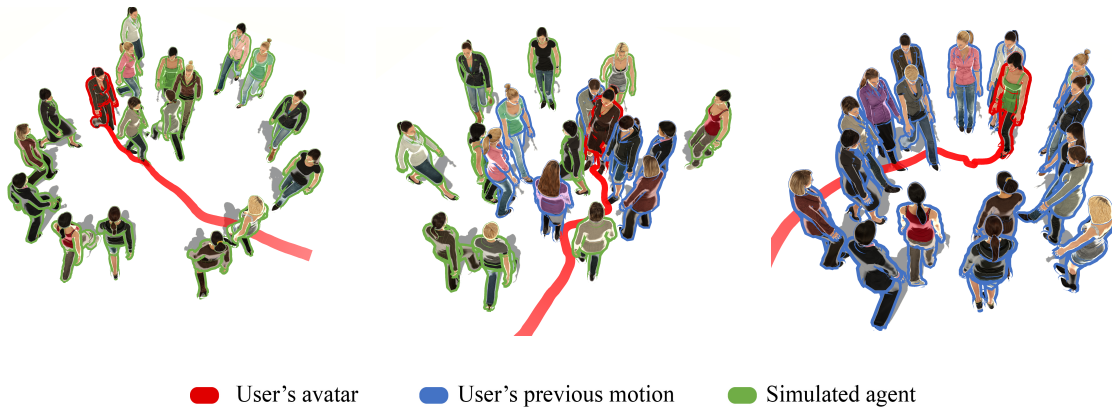


Figure 4.1 – Illustration of our proposed 3R paradigm (Replace-Record-Replay), where a single user (red outline) is initially immersed into a simulated *contextual crowd* whose autonomous agents (green outline) are successively replaced by the user’s captured data (blue outline). The consistency of the user’s behavior in this 3R scenario is compared with 4R (Replace-Record-Replay-Responsive), where the blue agents are made locally responsive.

involving a single user. This type of experiment immerses the recorded participant within a simulated, virtual crowd, consisting of autonomous – and therefore reactive – agents (e.g., [Olivier et al. 2014; Bruneau et al. 2015]), and has been proved to be very appropriate for closely observing the reaction of one (at a time) individual to a given situation within a crowd. This approach is, however, irrelevant for recording full crowd datasets, since it focuses on recording the behavior of several participants at once, and everything but the VR-user is simulated. Meanwhile, the One-Man-Crowd paradigm does not involve any simulation. We believe that such simulated interactions can be valuable in the pursuit of improving interactivity to the One-Man-Crowd paradigm

To this end, we build on and extend the One-Man-Crowd paradigm – which can be categorized as a process of **Record-and-Replay (2R)** – by immersing participants into *contextual crowds*, defined as a number of characters (agents) providing participants with realistic surroundings from the beginning.

Specifically, we propose two versions of this new paradigm: **Replace-Record-Replay (3R)**, where, instead of an empty scene as in 2R, users move among a contextual crowd and progressively replace, one at a time, each of its virtual agent; **Replace-Record-Replay-Responsive (4R)** where, extending 3R, a motion correction strategy is added to agents by locally editing their motion when the user is close. This makes user-recorded characters responsive, e.g. by enabling them to show collision avoidance behaviors. Our

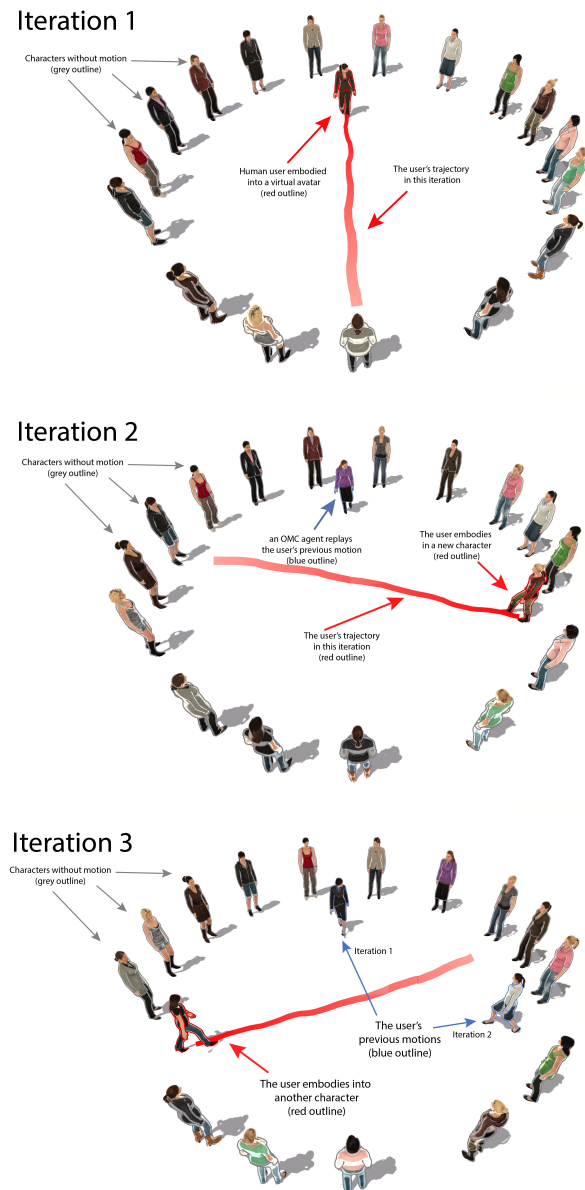


Figure 4.2 – During the first 3 iterations of the **2R** procedure, the virtual scene is nearly empty, and thus the user tends to walk straight towards his/her goal. This phenomenon is not realistic.

exploration of these two paradigms leads to the understanding of: (1) the effect of being immersed in a contextual crowd on the users' behavior during the crowd generation process, and (2) the effect of having responsive characters departing from their original trajectories on the user's behavior, which indeed impacts the quality of the generated

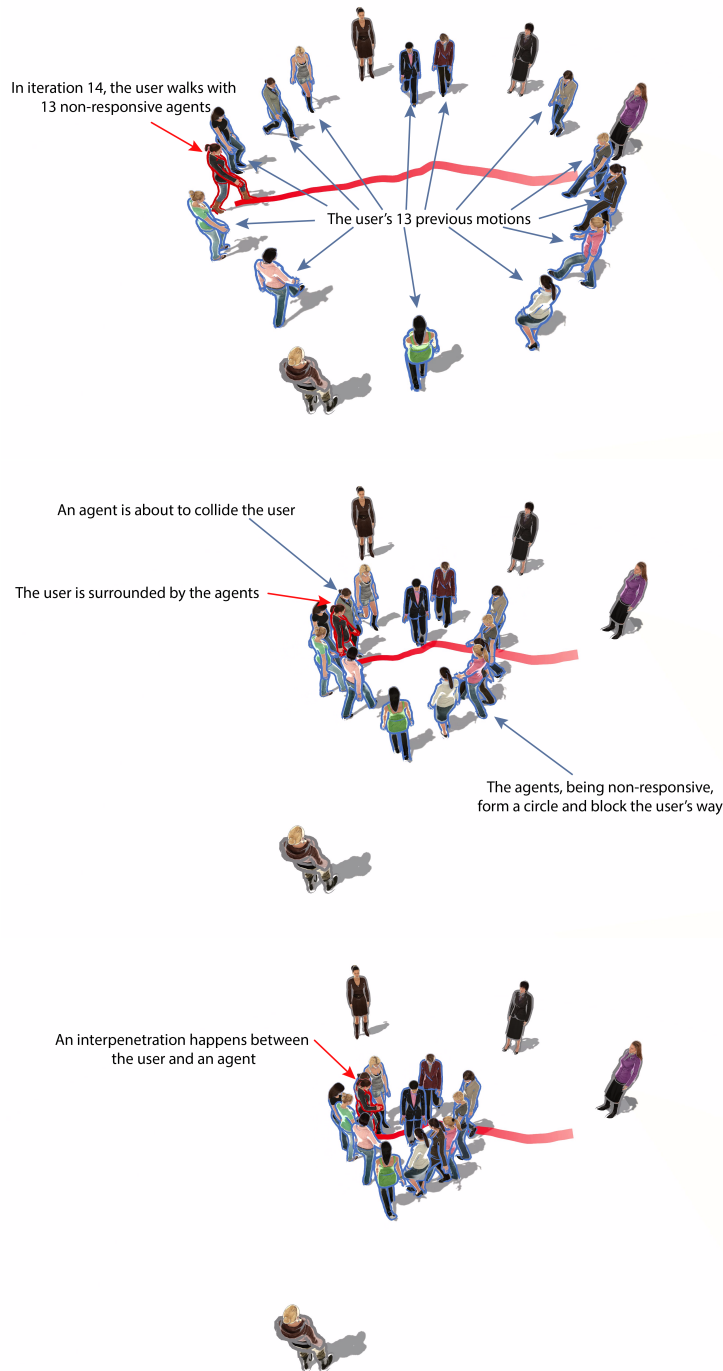


Figure 4.3 – As the generated agents are not responsive in the **2R** condition, these agents' movements might constrain the user's path planning and lead to unavoidable collision between the user and the other agents.

crowd data. The exploration of these effects relies on behavioral metrics to compare the properties of trajectories recorded using various paradigms, or with real data.

The contributions of this chapter are threefold:

1. We introduce the concept of **contextual crowds** to combine the benefits of immersion into a simulated crowd with those of Record-and-Replay methods, towards consistent crowd motion capture in VR. We propose two alternative setups (3R and 4R), and compare their results on two well-studied crowd scenarios.
2. Using the **3R** paradigm, we validate the use of contextual crowds to induce interactions during the crowd generation procedure. We further demonstrate that simulated contexts have similar effects than reusing real ones.
3. Using the **4R** paradigm, we demonstrate that the use of locally responsive agents improves the quality of user’s collision avoidance, at the cost of locally departing from user-captured-only data.

We introduce the key concepts in Section 4.2, the general hypotheses in Section 4.3, and detail our experiments in Section 4.4. In Section 4.5 we present our analysis methodology and the corresponding results. We finally discuss our method and the main findings in Section 4.6, before concluding.

## 4.2 Key Concepts & Notations

This chapter aims at proposing a novel approach for VR-based crowd motion capture, which further exploits the potential of simulation in Virtual Environment, and goes beyond the limitations of our work in the past chapter. To this end, we explore the benefits of using contextual and responsive crowds to improve the consistency of users’ behavior, towards the generation of more realistic crowd datasets. With this novel approach, we introduce multiple new concepts that refine or extend those of the previous chapter. Thus, a clear revision of the concepts and notations is required first. The remainder of this section defines the fundamental concepts underlying our study, i.e., the different models we use for virtual agents, the alternative VR motion-capture processes, and the experimental scenarios we chose for validation.



### 4.2.1 Agents

To populate a virtual scene, we use several types of agents, which differ by the method that drives their **motion**, by their **role** in the data collection process or finally by their **interaction capabilities**.

**Motion.** Agents can be animated using different methods:

- **OMC agents** refer to agents whose movement was recorded by the participant at a previous iteration. In other words, they refer to the agents that are created by following the paradigm introduced in the previous chapter.
- **Captured agents** also move from pre-existing motion data (e.g. crowd dataset), but not recorded by the participant. When 2D only crowd datasets are used, 2D trajectories are augmented into 3D full-body animations using animation techniques.
- **Simulated agents** refer to agents whose motion is fully computed using a real-time simulator (e.g. some microscopic crowd simulation method), enabling them to make real-time decisions in order to reach a predefined destination in the scene while avoiding collisions. As for captured agents, the computed trajectory is to be combined on the fly with other animation techniques to produce 3D full-body animated agents.

**Role.** Our goal remains to enable a single user to create new crowd motion datasets. For such datasets to originate from participants' motion, **OMC agents** data are stored (and can then be edited by the 4R mechanism, as explained below). The two other types of agents (captured or simulated) may be used as **Contextual agents**, i.e., agents intended to provide users with contextual information, while they successively act to capture motion for the OMC agents. Given our new concept of contextual crowds, Contextual agents are to be present in the virtual scene from the first motion capture iteration.

**Interaction Capabilities.** From the agent descriptions given above, we understand that, at any given moment, we may immerse the participant among agents whose movements are pre-recorded and who ignore the participant's actions. Collision avoidance behaviors being expected to be reciprocal, this may cause the user to fail avoiding collisions. Therefore, we investigate the possibility of providing both OMC and Captured agents with collision avoidance capabilities. Therefore, we define:

- **Non-responsive agents** as Captured or OMC Agents, non-responsive as they only move by following pre-recorded data and cannot interact with the user.
- **Responsive agents** as either Simulated (i.e., responsive by nature), or modified Captured or OMC agents that locally adjust their motion upon the detection of a risk of collision with the user.

**Agents implementation details.** In our experimental set-up, the ORCA crowd simulation technique [Van Den Berg et al. 2011] is used to generate the global trajectory for Simulated agents as well as the reactive portions of trajectory for the Responsive agents, using the implementation provided by Toll et al. [2020]. ORCA is always activated for Simulated agents, but needs to be activated and deactivated for the other kinds of Responsive agents according to the risk of collision with the participant. To this end, we simply check the distance  $d$  between any agent and the participant, and trigger activation when  $d < 1m$ . ORCA is active for a minimum of  $0.2s$  to avoid any flickering effects due to frequent activation and deactivation. The preferred velocity of Simulated agents is computed to steer them towards their final goal (see Section 4.4.1). The preferred velocity of the other categories of reactive agents is computed as to recover their pre-recorded trajectory when ORCA is further deactivated. 2D trajectories are augmented to 3D animated characters using a Unity animator module driven by walking speed, in combination with a set of motion-captured walking animations.

## 4.2.2 Crowd-motion-capture Paradigm

### 4.2.2.1 Process

The main purpose of our work is to evaluate the importance of adding contextual agents when a single user records a crowd dataset. To this end, we propose and compare a number of variations of the OMC paradigm that was presented and evaluated in the previous chapter (illustrated in Figure 4.4):

- **2R** stands for “**R**ecord and **R**eplay”. We hereby use this term to represent the process described in the previous chapter, to better distinguish with the following variations of the OMC paradigm. In this version, the user is initially alone in the virtual scene, which is iteratively populated with OMC agents (see Figure 4.4).
- **3R** stands for “**R**eplace, **R**ecord and **R**eplay”. In this version, the virtual scene is initially populated with **non-responsive contextual agents**. The participant

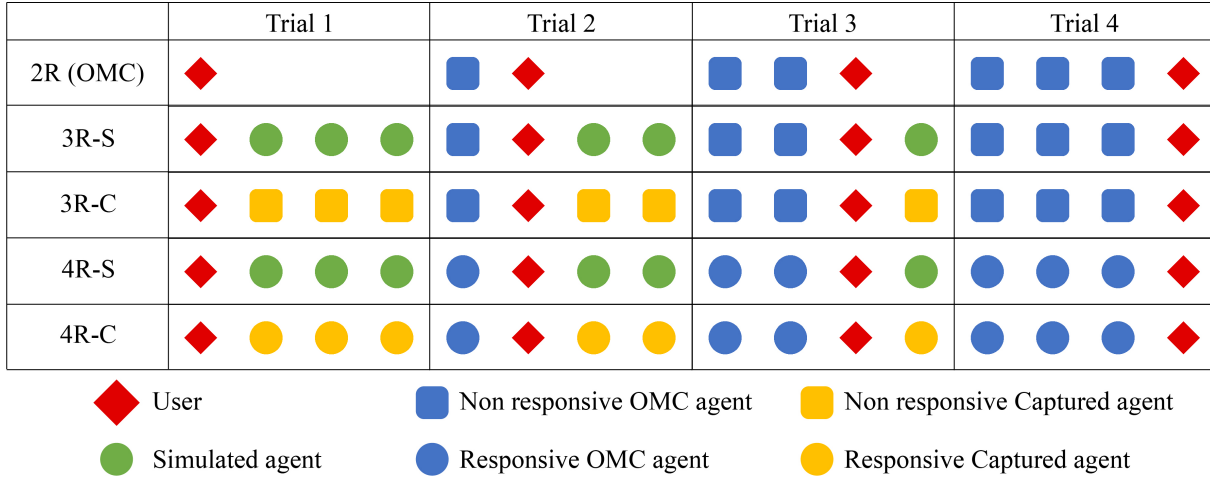


Figure 4.4 – Illustration of the different paradigms in an example with 4 agents. The figure illustrates each agent’s behavior before and after being acted out by the user, which leads to a total of 4 trials to record the exemplar scenario. Since Simulated agents are always responsive, they come in a single version (disc shape).

iteratively embodies each contextual agent, to record a new motion and replace it with an OMC agent. We distinguish 2 versions of the **3R** paradigm: i) **3R-C** where contextual agents are **captured agents**, and ii) **3R-S** where contextual agents are **simulated agents**.

- **4R** stands for “**R**eplace, **R**ecord, **R**eplay and **R**esponsive”. In this version, the virtual scene is initially populated with **responsive contextual agents**. The participant iteratively embodies each contextual agent, to replace it with an OMC agent. We distinguish 2 versions of the **4R** paradigm: i) **4R-C** where contextual agents are **captured agents**, and ii) **4R-S** where contextual agents are **simulated agents**.

In summary, the **2R** condition is our baseline. Since one of its flaws is the lack of contextual information, the **3R** condition enables us to evaluate the importance of adding contextual agents. In addition, the **4R** condition enables us to evaluate the importance of also preventing collisions between agents and participants. Lastly, the **XR-S** conditions enable us to evaluate the use of simulation-based contextual or reactive agents instead of those based on captured data (**XR-C**). We refer the readers to Figure 4.5 and Figure 4.6 for more concrete illustrations about the 3R and 4R process.

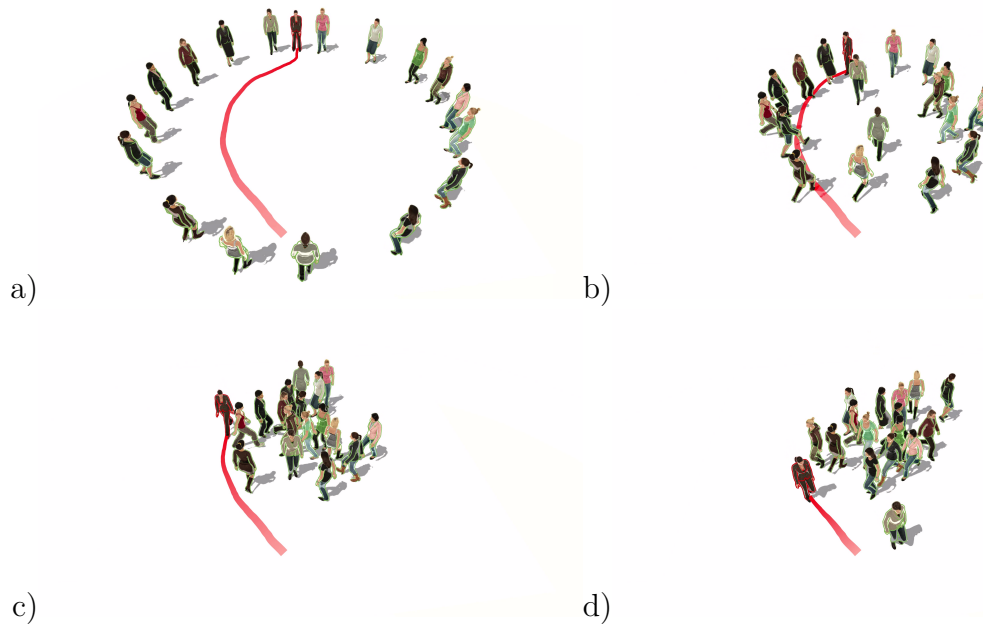


Figure 4.5 – We illustrate the 3R process through an example of the simulated context (i.e., 3R-S). From a) to d), the subfigures depict the evolution of the user’s (red outline) and the simulated agents’ (green outline) position during the first iteration of capture. The red curves on the ground show the user’s future trajectory. To capture a virtual crowd with 18 crowd members, the user will first walk and interact with 17 simulated contextual agents (i.e., from the character  $c_2$  to  $c_{18}$ ), whose movement are computed in real-time by a simulator. While acting out the character  $c_1$ , the user’s motions and paths will thus reflect his/her interaction with those agents. Then, during each iteration, the user will replace the corresponding simulated agent. Afterward, we keep the captured user motion to animate that virtual character and remove the former simulated agent. By the end of the capture, all the simulated agents will be removed, and only user-generated data will be kept. The 3R-C process follows the same process, and the only difference is that the contextual agents are reanimating real-world captured trajectories, and are thus non-responsive.

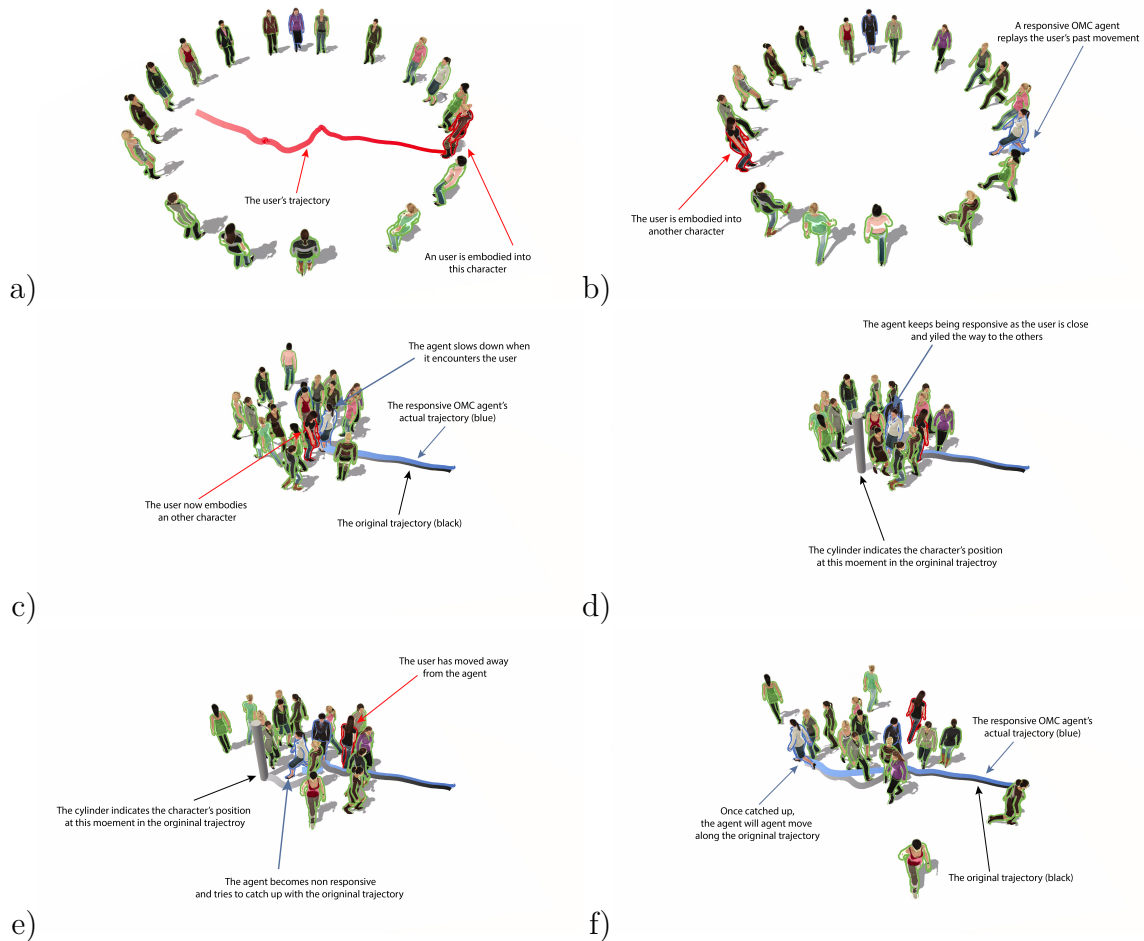


Figure 4.6 – Illustration of the 4R process with an example of the captured contextual agents. a) An user first acts out the concerned character of this illustration during a given iteration. b) During the following iteration, we use an OMC agent to display the recorded movement, while the user embodies another character. c) The OMC replays the recorded movement until it encounters the user, when we make the agent responsive by activating a simulator. d) Driven by the simulator, the OMC agent avoids collision with both the user and the other agents around, by mainly yielding the way to them. Meanwhile, the collision avoidance causes delay compared to the agent’s original movement e) When the user goes away from the agent, we deactivate the simulation, making the agent non-responsive and let it catch up the delay. f) If the OMC agent manages to catch up the delay, it will once again replay the original recorded movement.

#### 4.2.2.2 Apparatus

For the experimentation, we reused the hardware system that has been deployed in the previous chapter (see Section 3.4.3), except that their drivers were updated to the latest version before the experimentation (at the beginning of 2023). Regarding the virtual environment creation, we updated to Unity 2021.3 instead of 2019.4.

### 4.3 General Hypotheses

Our experiment was designed to assess whether these two novel VR crowd motion-capture methods can address the limitations of the 2R method, where users 1) freely walk in an empty scene during the first trials instead of walking among a moving crowd, and 2) are struggling to avoid collisions as more and more non-reactive agents are progressively added to the scene. This was shown to create biased behavior from participants, as they are 1) unlikely to initiate interactions when faced with an empty experimental environment, and 2) likely to experience unavoidable collisions when faced with a dense non-responsive crowd.

Our general hypotheses are

- **H1:** By initializing scenarios with contextual crowds, we can mitigate the bias caused by the fact that users initially walk alone in an empty scene.
- **H2:** In the 3R paradigm, the type of contextual crowd, whether it is created through simulation or reanimation from previously captured data, does not significantly impact the results.
- **H3:** The local responsiveness of surrounding agents can reduce the number of collisions experienced by users and thus improve the consistency of the dataset. However, frequent alterations of the trajectories already recorded by the participant may gradually decrease its similarity to human behavior.

## 4.4 Experiment

### 4.4.1 Experimental Design

Participants first read and signed the experiment consent form, and were presented the task to perform. They were then equipped with the Xsens motion capture system, VR backpack, and Pimax 5K HMD. The motion capture system was then calibrated

with the guidance of the experimenter to guarantee the quality of motion capture and to ensure that avatar size corresponded to the participants’ dimensions. Participants were then immersed in an empty tutorial room and asked to walk six times back and forth from one side of the room to the other, in order to become familiar with walking in a virtual environment. Meanwhile, the participants’ comfort walking speed was measured during this process, and used to adapt the preferred walking speed of simulated characters (see details below).

During the experiment, participants went through the 5 experimental conditions: 2R, 3R-S, 3R-C, 4R-S, 4R-C. The order in which each participant completed these 5 blocks was randomly generated. In each condition, participants were always immersed successively into all the different virtual characters of the sequence in a fixed order. To ensure a smooth and realistic experience, we adjusted the speed of the contextual agents, which prevented participants from falling behind and losing contact with the contextual crowd due to a reduced walking speed. For *captured agents*, we adjusted the speed of real trajectories using the factor  $f = V_U/V_R$ , where  $V_U$  represents the participants’ comfort speed, measured during the tutorial and unique to each individual, and  $V_R = 1.4 \text{ m/s}$  represents the empirical pedestrian comfort walking speed in real-life [Berton et al. 2019]. Similarly, we randomly generated a preferred walking speed for each *simulated agent* within  $1.4 \pm 0.2 \text{ m/s}$  as a constant default speed for each experience, and then changed their preferred walking speed by the factor  $f$ .

## 4.4.2 Scenarios & Tasks

The 2 scenarios that we chose to replicate are similar to those used in the previously mentioned exploratory experiment (Section 3.7), which are indeed of great interest in the understanding of multidirectional crowd movements. They involve however numerous difference, such as character numbers, starting positions, etc. Those details are described below, where we first introduce references to relevant existing research, then present the tasks that the participants accomplished in each scenario.

### 4.4.2.1 Circle

**Context.** The **Circle** scenario corresponds to a scenario where a group of pedestrians walks from a different point on the circle to the antipodal position of the circle. Before

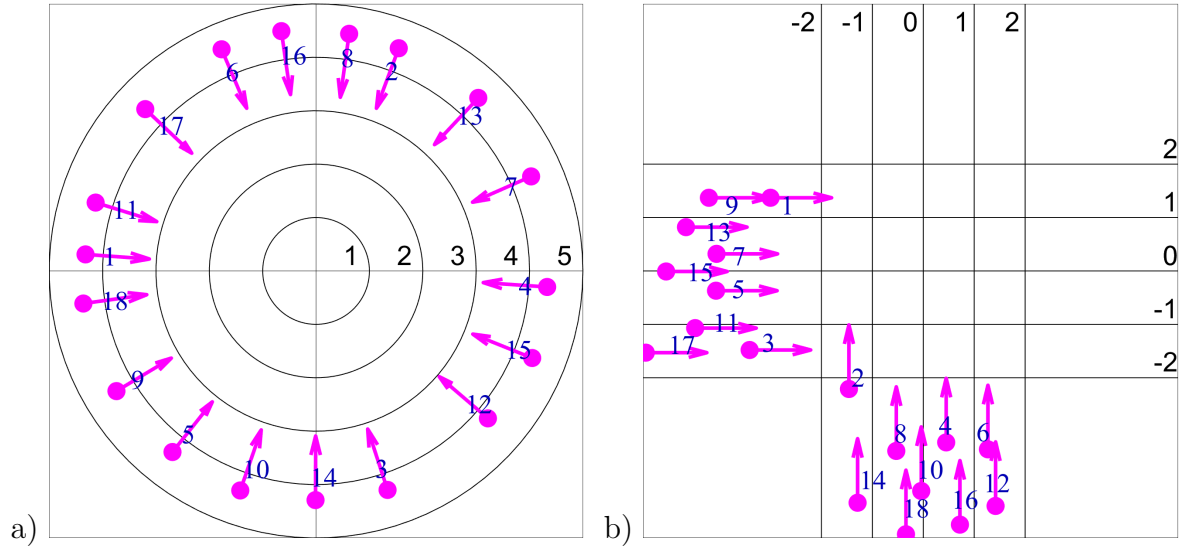


Figure 4.7 – Illustration of the two experimental scenarios, showing each agent’s initial position and direction to go. For each scenario, we use equidistant  $1\text{ m}$  guidelines to visualize the crossing area. a) Circle. b) Crossing Flow.

walking, the pedestrians’ starting positions are uniformly distributed around the circle. Thus, the distance to the closest neighbor is equal to every pedestrian. This scenario is not regular in daily routine. It is, however, a scenario of great interest in crowd study as it maximizes the number of interactions [Ondřej et al. 2010]. Such a scenario is challenging to crowd simulation, as congestion may arise from the agents converging to the center. Previous research leveraged this scenario to validate simulation algorithms [Kwiatkowski et al. 2023; Karamouzas et al. 2017] and calibrate simulation parameters [Wolinski et al. 2014a; Daniel et al. 2021].

**Task.** In our experiments, we replicated the setup presented by Wolinski et al. [2014a], using a set of captured data with 18 pedestrians as depicted in Figure 4.7-a. Our participants were instructed to walk from one side to the antipodal position of a  $4\text{ m}$  radius circle. To make the initial condition of our experiment as close as possible to the real-world experiment, we selected a set of recorded data and precisely replicated the initial positions, which explains why those initial positions were not homogeneously distributed, and the circle is not strictly of  $4\text{ m}$  radius. To help the participants localize the antipodal position of each starting position, we constantly visualized the end position with a green spot on the ground. Then, the participants were instructed to walk and reach the green spot without colliding with any other agents. Let us remind the readers that too strong



a demand on collision avoidance may induce the participants to behave unnaturally (see Section 3.7.3.1). To minimize the influence of our experimental instruction, we instructed the participants to prioritize behaving normally instead of committing unnatural behaviors, if the situation was too complicated for them to avoid collision. Namely, it involves two specific situations: 1) a non-responsive agent approaching the participants from an unperceivable direction, and 2) a group of non-responsive agents walking in conflicted directions surrounding the participants. For all the capture conditions (i.e., 2R, 3R-X, 4R-X), the participants received the same instruction. In addition, the participants were informed that, in general, the agents' behaviors differed during the experiment, that some agents would actively avoid the participants and some agents would not. But we did not provide any specific agent behavior information linked to each condition. The participants ended each iteration when they came within 0.5 meters of the green spot.

#### 4.4.2.2 Crossing Flow

**Context.** This scenario, representative of regular daily crowd movements, corresponds to a flow of pedestrians walking in two orthogonal directions and sharing a crossing area. In crowd study, it plays a vital role in understanding both collective emergent behavior (such as lane and stripe formations [Mullick et al. 2022]), and individual motion characteristics (such as fundamental diagram [Cao et al. 2017]). This scenario is complementary to the Circle scenario, as it involves a continuous human flow in each direction, and therefore continuous interactions.

**Task.** We designed our virtual Crossing Flow scenario to replicate the experiment introduced by Mullick et al. [2022]. This replication, however, faced multiple physical constraints. Controlled experiments of crossing flow scenarios usually involve large-scale movements, where the distance between the start region and the end region can be up to 30 *m*. In our research facility, it was impossible to find an experimental area of such size, not to mention that our VR capture solution did not have the capability to track the participant in such an area. Considering the importance and the research interest of the Crossing Flow scenario, we decided to realize several adaptations when replicating the scenario presented by [Mullick et al. 2022]. We first needed to limit the character number in the scene. To this end, we retained the first 9 pedestrians of each group, and selected the moment when the first pedestrian enters the crossing area, along with the positions of

these 18 individuals at this stage, as the starting point of our experiment (as depicted in Figure 4.7-b.). The participants were instructed to start each iteration from the corresponding character’s start position, and to walk across the virtual scene until they passed an indication line on the ground. The line was located at 8 *m* in front of the participants to visually indicate the general walking direction. The participants were free to pass over any point of the line as they wished. Similar to the Circle scenario, we instructed the participants to avoid collision with the agents, while prioritizing normal behaviors. The participants were equally informed that the agents’ behavior could be different. For all the five experimental conditions (2R, 3R-X, 4R-X), the experimental instructions were the same. Each capture iteration ended when the participants walked over the indication line.

### 4.4.3 Participants

Thirty participants took part in our experiment (average age of  $31.6 \pm 8.4$  *y.o.*, ranging from 22 to 51 *y.o.*). They were recruited through internal emailing lists among students and staffs. All participants had normal or corrected-to-normal vision and gave written informed consent. The study followed the declaration of Helsinki, and was approved by our local ethical committee. Each participant only performed one scenario (but all five conditions for this scenario). In total,  $15 \times 5 \times 2 = 150$  sets of crowd motion were generated, containing 2700 individual trajectories, with an average duration of 14.8 *s*.

Among the 15 participants who participated in the Circle scenario, the estimated comfort walking speed varied from  $0.74 \text{ m.s}^{-1}$  to  $1.28 \text{ m.s}^{-1}$  with a mean value of  $1.04 \text{ m.s}^{-1}$  and a standard deviation of  $0.17 \text{ m.s}^{-1}$ . Among the 15 participants who participated in the Crossing Flow scenario, it varied from  $0.82 \text{ m.s}^{-1}$  to  $1.17 \text{ m.s}^{-1}$  with a mean value of  $0.96 \text{ m.s}^{-1}$  and a standard deviation of  $0.10 \text{ m.s}^{-1}$ .

## 4.5 Evaluation

In this section, we present the analysis and results of the experiment. The two experimental scenarios shared the same analysis while presenting similar results but with difference. Thus, we first introduce the shared analysis framework, then separately provide each scenario’s results.

## 4.5.1 Analysis Framework

We divide our analysis of the experiment into two parts. The first analysis focuses on the effect of contextual crowds, which corresponds to the *Replace* component in both paradigms. The second analysis focuses on the effect of the local correction of user-generated trajectories, corresponding to the additional *Responsive* component of the 4R paradigm.

### 4.5.1.1 Contextual Crowds

To compare trajectories of various lengths and durations across participants, we normalize the captured trajectories over the distance covered. This normalization leads to relative walking progress from 0% (beginning) to 100% (end) of each trajectory. For trajectories with different shapes and lengths, the *relative progress* provides us with a universal “semantic” reference. These normalized trajectories are used to compute the following characteristics.

We then leverage Statistical Parametric Mapping (SPM) [Friston et al. 2007] to study the differences and similarities of these characteristics under different conditions.

**Average Trajectories.** As participants act and replace the contextual agents in a fixed order, we conduct statistical analysis on the generated movements regarding each agent. This is done by overlapping the same agent’s relative walking trajectories of different participants (see above). For each condition  $c$  and each agent  $a$ , we compute the mean trajectory  $\bar{\mathbf{x}}_{c,a}^*$  of all the participants:

$$\bar{\mathbf{x}}_{c,a}^* = \frac{1}{M} \sum_{j=1}^M \mathbf{x}_{c,a,j}^* \quad (4.1)$$

where  $M$  is the number of participants, and  $\mathbf{x}_{c,a,j}^*$  is the trajectory of participant  $j$  for agent  $a$  in condition  $c$ . The average trajectory reveals the mean behavior of participants when acting out each specific agent.

**Trajectory Variation.** In addition to the computation of average trajectories, we visualize the variances on each trial to show to which level participants’ behaviors differed

from each other under these conditions. The variance is defined as follows:

$$\mathbf{s}_{c,a}^* = \frac{1}{M} \sum_{j=1}^M (\mathbf{x}_{c,a,j}^* - \bar{\mathbf{x}}_{c,a}^*)^2, \quad (4.2)$$

where  $M$  is the number of participants, and  $\mathbf{x}_{c,a,j}^*$  is the trajectory of participant  $j$  for agent  $a$  in condition  $c$ , and  $\bar{\mathbf{x}}_{c,a}^*$  is the mean trajectory.

**Normalized Speed.** Generally speaking, participants experience three distinct phases during walking: an initial acceleration phase from the starting point to just before entering the crossing area, a deceleration phase upon entering the crossing area due to interactions and avoidance maneuvers with other agents, and a re-acceleration phase after disengaging from the agents. We thus study the velocity profiles under each condition. As each set of crowd motions was generated by a single participant, the crowd’s moving speed was dependent on the participant’s preferred walking speed. This speed is unique to each participant. Thus, we normalized each participant’s data by their comfort walking speed  $v_p$  before conducting statistical analysis.

With a similar idea, we normalized real data’s speed when making comparisons between the real data and our results. The real data, being a set of trajectories of different individuals, cannot be normalized by one specific person’s preferred speed. We therefore use an average speed value  $v_p^* = 1.4 \text{ m/s}$  provided by previous research [Berton et al. 2019] to compute the normalized speed  $v_n = v_r/v_p$ , where  $v_r$  is the speed from the real data.

**Lateral Movement.** To estimate the level of effort put by participants to avoid collisions, we compute their lateral movement for each trial. In an empty scene, participants would ideally walk straight from their starting point to their destination; we thus assume that lateral movement is the result of avoidance behaviors. It is calculated by projecting the participants’ positions to the line segment between their starting point and destination, and computing the distance to the projected points.

**Local Density.** The local density describes the participant’s surrounding density, and relates to the number of agents who are in their proximity at each given time. This metric reveals both the difficulty of collision avoidance and the size of the personal zone that participants are comfortable with. To compute the local density around each character at

each time frame, it is common to perform a Voronoi tessellation of the space, and then compute the density as the inverse of the character’s Voronoi cell’s area. To avoid artifacts caused by edge effects, we adopted a modified Voronoi tessellation [Nicolas et al. 2019], in which an agent is considered on the edge if its Voronoi cell shares an intersection with the convex hull of all the agents’ positions. Nicolas et al. [2019] define the angle  $\theta$  that describes the proportion of the non-empty space in each cell, such that the local density can be computed as:

$$\rho_i = \frac{\theta}{2\pi} \frac{1}{A_i}, \quad (4.3)$$

where  $A_i$  is the area of the Voronoi cell intersecting with the agents’ convex hull. For agents that are not on the edge,  $\theta = 2\pi$  and thus the local density is the inverse of the area of the Voronoi cell.

**Distance to Nearest Neighbor.** We aim to study participants’ tolerance to proximity with other agents while they are walking within virtual crowds. By assuming that they do not allow any virtual agent to enter their minimal personal zone, we only focus on the positional relationship between the participants and the virtual agent closest to them. To this end, we compute the distance to the participants’ nearest neighbor at each frame.

**First vs. Last Trajectories.** We further investigate whether participants behaved differently between the beginning and the end of the experiment. To do so, we analyze the aforementioned normalized walking speed and lateral movement. These two metrics focus on the individual characteristics demonstrated by each crowd member, and thus can separately reflect the participants’ behavior during each trial. For each trial, we compute a single mean value of both the normalized walking speed and lateral movement, then aggregate all the participants’ results.

#### 4.5.1.2 Local Corrections due to Reactive Agents

To evaluate the effect of local corrections of the user-captured-trajectories due to reactive agents, we adopt a multi-faceted approach. Given the novel nature of the problem (i.e. ideally, evaluating the decrease in realism due to these local corrections), there is no established analysis framework to rely on. To offer a comprehensive understanding, we therefore selected multiple metrics for our analysis. Initially, we investigate the effectiveness of local corrections in reducing collisions between the user and agents. Subsequently,

we delve into its impact on previously generated trajectories. Specifically, we seek to answer the following questions: How frequently is the local correction invoked? In which parts of the agent trajectories do local corrections occur? How do the adjusted trajectories differ from the original user-generated ones?

**User Collision.** We examine the average number of collisions between the participant and OMC agents in each condition. Collisions are determined based on interpersonal distance (0.5 m threshold).

**Correction Frequency.** To examine the frequency of changes in OMC agents' trajectories, we count the number of responsive OMC agents altered in each trial. As the agent count increases with each trial, so does the likelihood of local correction. This makes each trial individually valuable for analysis, leading us to conduct evaluations on a per-trial basis instead of aggregating all trials.

**Correction Location.** We investigate trajectory modifications of all OMC agents within each trial to understand local correction occurrences. To this end, we measure the position along the agent's original trajectory – expressed as the relative walking progress – where the local correction occurs. To discern overarching patterns, we aggregate data from all participants and calculate the average local correction occurrence position for each OMC agent in every trial. This approach allows us to scrutinize the relationship between trial number, agent index, and the triggering of local correction.

**Trajectory Dissimilarities.** We further shift our focus to quantifying the dissimilarities between the OMC agents' modified trajectories and the original ones within each individual trial. To accomplish this, we use Dynamic Time Warping (DTW) as a metric for gauging these differences. DTW offers the advantage of nonlinear matching capabilities, enabling us to align sequences that may differ in time but share similar shapes. It also provides a balanced integration of both local and global information, enhancing the robustness of our analysis. Again, we aggregate the data by taking the average results across all participants to identify possible systematical effects related to scenarios, trial number, and agent index.

## 4.5.2 Scenario IV - Circle

### 4.5.2.1 Results on Contextual Crowds

To isolate the effects of contextual crowds, our analysis in this section only focuses on the comparisons between the 2R and 3R-X (X represents either S or C) conditions, as well as the real data. Additional comparisons to the 4R-X conditions are provided in parallel for completeness, but is not our main focus.

**Average trajectories.** We start by reporting average trajectories in each of these conditions, displayed in Figure 4.8. This figure is useful for visual inspection and understanding. For instance, we see that the 2R and 3R-X conditions present each unique features, with, for 3R-X, lower speeds and more gradual speed changes within the crossing area (central area of each scenario). Further quantitative comparisons are delivered through Figure 4.10. Note that the figure reports quantities as a function of the progression on the trajectory normalized by its length. Thus, 0% is the starting position, 100% corresponds to the end position, and the 40% to 60% portion of the trajectory is where interactions are the most intense.

**Trajectory Variation.** In addition to the average trajectories, we further display the trajectory variation for the Circle scenario, depicted in Figure 4.9. In these two figures, we separately display the mean trajectory and the variation of each iteration to ease the readers' lecture. Meanwhile, exhaustively exhibiting each iteration would be pointless and complicate the reading. Thus, we selected several iterations of interests out of all the 18 iterations.

**Normalized Speed.** Figure 4.10-a illustrates the average evolution of the normalized speed in each condition. In the Circle scenario, we observe that all results follow an accelerate-decelerate-re-accelerate pattern, but the location of the deceleration phase differs: 3R-X results are closer to real data, while the 2R results significantly differ from real data and 3R-X. SPM analysis corroborates these observations by detecting significant differences between the 2R and the real data (47% - 100%), and between the 2R and 3R-X conditions (2R/3R-S: 44% - 88% and 2R/3R-C: 45% - 80%). In comparison, the 3R results only significantly differ from the real data at the end of the trajectory, which suggests that participants did not come to a complete stop upon reaching the endpoint.

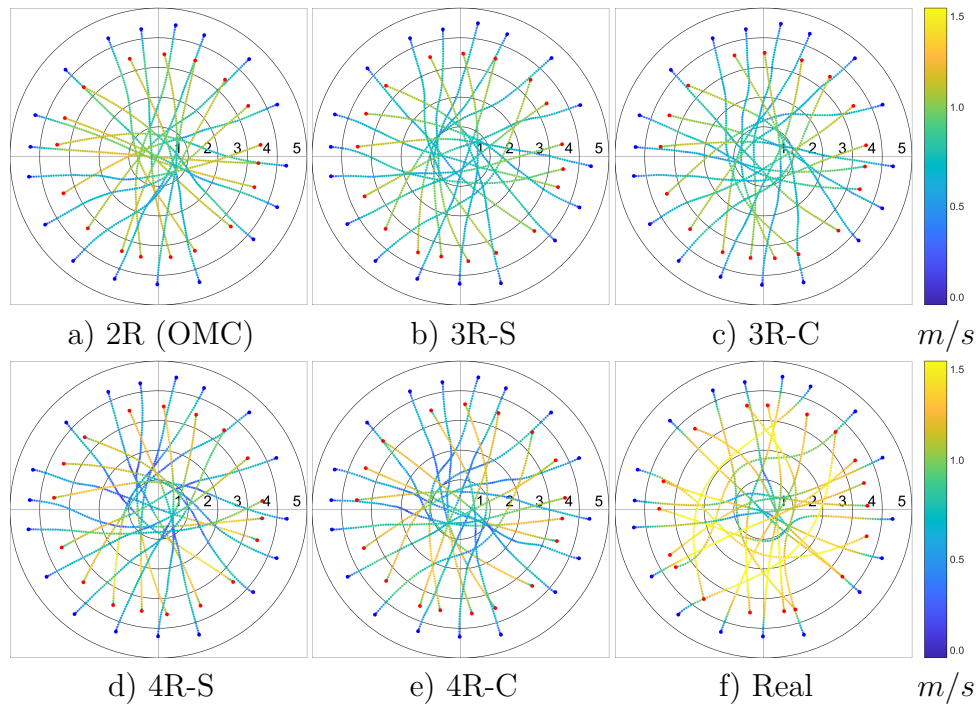


Figure 4.8 – a)-e) Mean trajectories generated under different conditions with all participants aggregated, and colored with the averaged speed. Blue and red discs indicate respectively starting and ending locations. f) The real results from which we replicated the experiment. Equidistant 1 m guidelines (concentric) provide visual information about the areas of crossing.

**Lateral Movement.** Figure 4.10-b depicts the average lateral movement in each condition. 3R-X results present higher lateral movement nearby the crossing area compared to both 2R and real results. For the Circle scenario, while SPM analysis revealed a 100% significant difference of 2R when compared with 3R-X, it did not reveal any significant difference with the real data.

**Density.** While we can observe in Figure 4.10-c that the local density is slightly higher for the real data than for any other condition, SPM analysis only shows significant differences in extremely small areas. In this Circle scenario, these differences occurred around 50% of the walking progress for all 2R and 3R-X conditions.

**Distance to nearest agent.** Figure 4.10-d shows that the analysis of distance to the nearest neighbor did not reveal any meaningful significant difference between the various



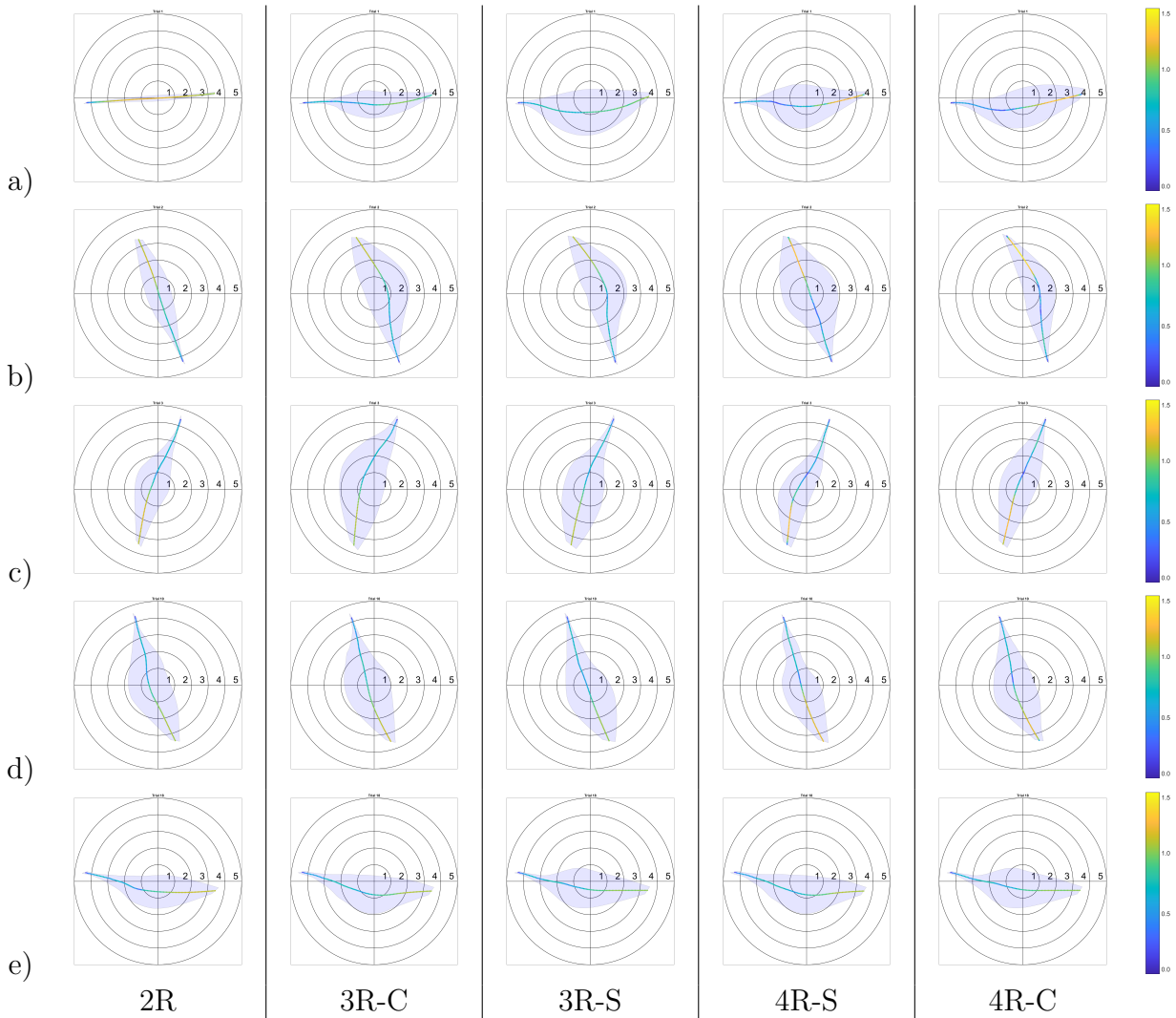


Figure 4.9 – From top to bottom: mean trajectory and variance of iterations a) 1, b) 2, c) 3, d) 10 and e) 18 (last) for the Circle scenario. From left to right: 2R, 3R-S, 3R-C, 4R-S, and 4R-C paradigms. The color bar reveals the average speed ( $m/s$ ) along the trajectories.

experimental conditions.

**First vs. Last Trajectories.** The iterative process set by 2R or 3R-X paradigms results in an order for each recorded trajectories. Figure 4.11 shows that the 2R paradigm introduces more differences between first and last recorded trajectories, both on normalized speed and lateral movement. In particular, the Circle scenario exhibits a noticeable trend of higher speed during the first 6 iterations and lower lateral movement during the first 4 iterations, indicating that participants tended to walk straight forward across the

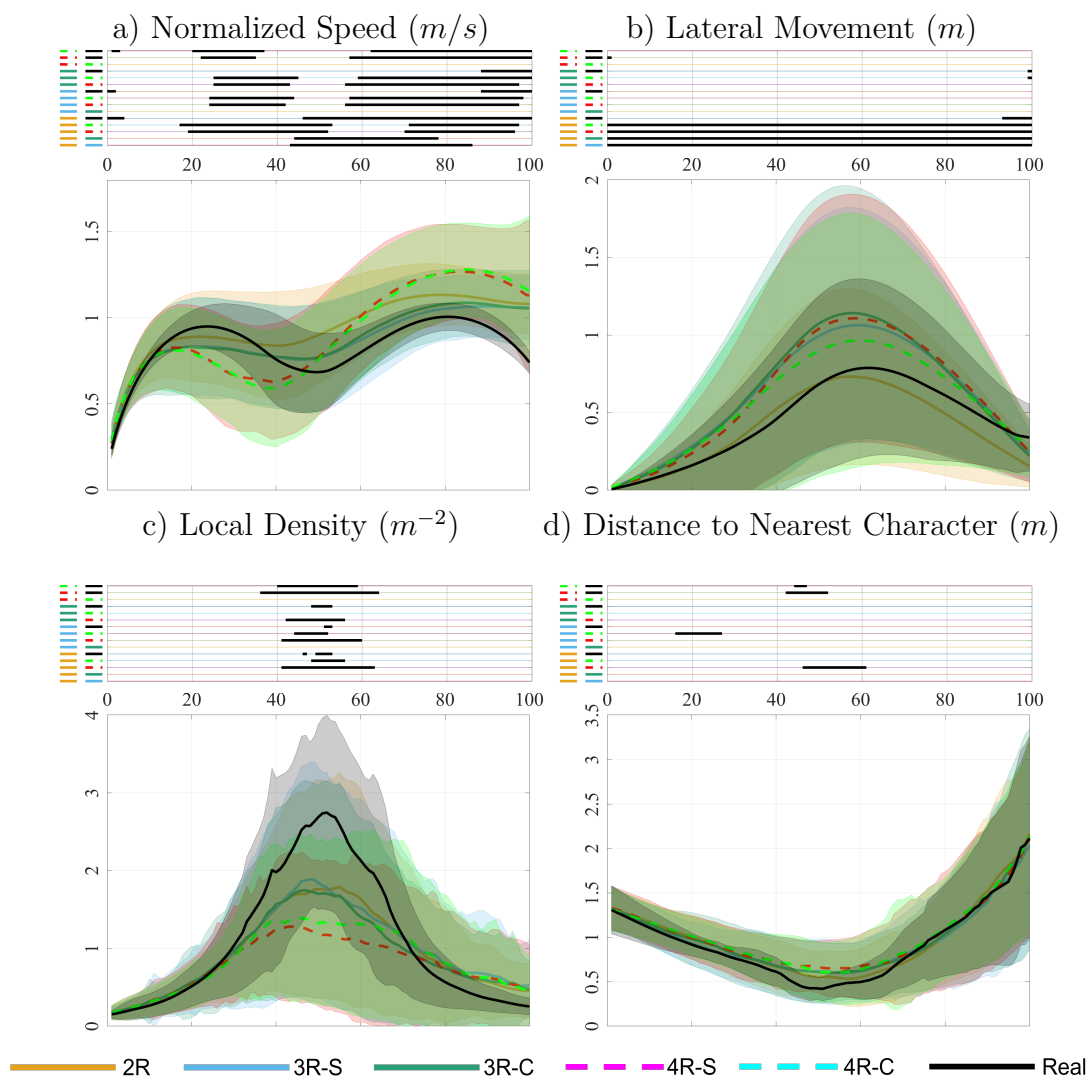


Figure 4.10 – Evolution along the distance that participants walked (in percentage) of participants’ a) normalized speed, b) lateral movement, c) local density, and d) distance to the nearest character. SPM significant differences are displayed above the figure using thick black bars.

scene, as the scene was not very populated in the beginning. The 3R-X results, on the other hand, present consistency among different iterations. This implies that participants tended to make more similar efforts across iteration.

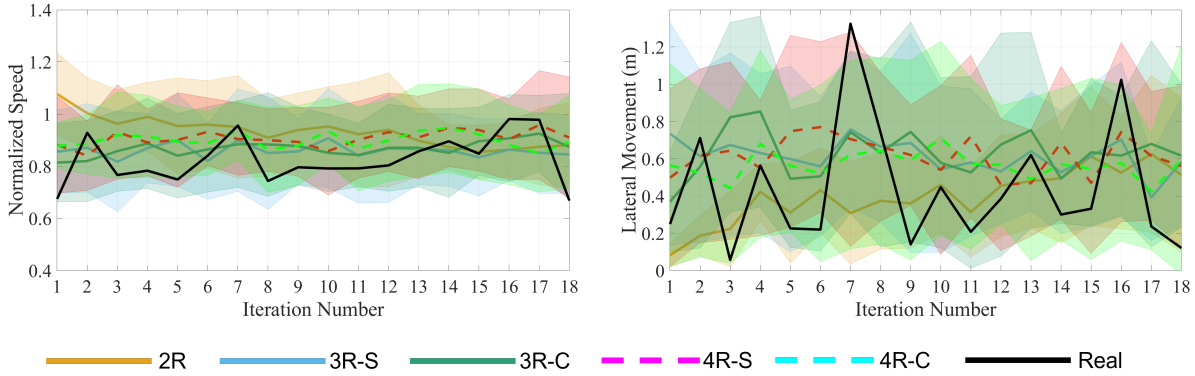


Figure 4.11 – Mean normalized speed and lateral movement over all participants when acting different characters under different conditions in the Circle scenario, sorted by number of the iteration. While participants were walking together in the real scenarios, and therefore the real data did not have any order, it is overlapped on the figure for illustrative purposes. **Change Figure->Include the 4R results, and remove the flow results**

#### 4.5.2.2 Results on Local Corrections

**User Collision.** As shown in Table 4.1, the local correction strategy effectively reduced the number of collisions between participants and OMC agents. Although there are still some collisions, our method generally reduced the number by half in both scenarios. This observation suggests that participants experienced less avoidance difficulties in the 4R conditions than the 3R ones. However, because of implementation limitations the final generated crowd presents more collisions (4R-S vs. 3R-S, 4R-C vs. 3R-C). This suggests the presence of irregular trajectories in 4R-X results, which is discussed in Section 4.6.2.

Table 4.1 – Average number of collisions between participants and OMC agents across trials and participants (top), and total number of collisions between all individuals in the final generated crowd, averaged over participants (bottom). The values are presented in the format  $mean \pm SD$ .

2R (OMC)	3R-S	3R-C	4R-S	4R-C
<b>Number of collisions with OMC agents</b>				
1.10 ± 0.85	1.30 ± 1.07	1.10 ± 0.96	0.50 ± 0.47	0.47 ± 0.43
<b>Number of collisions in the final generated crowd</b>				
13.2 ± 11.9	14.5 ± 14.9	11.5 ± 11.2	16.6 ± 10.1	12.7 ± 9.2

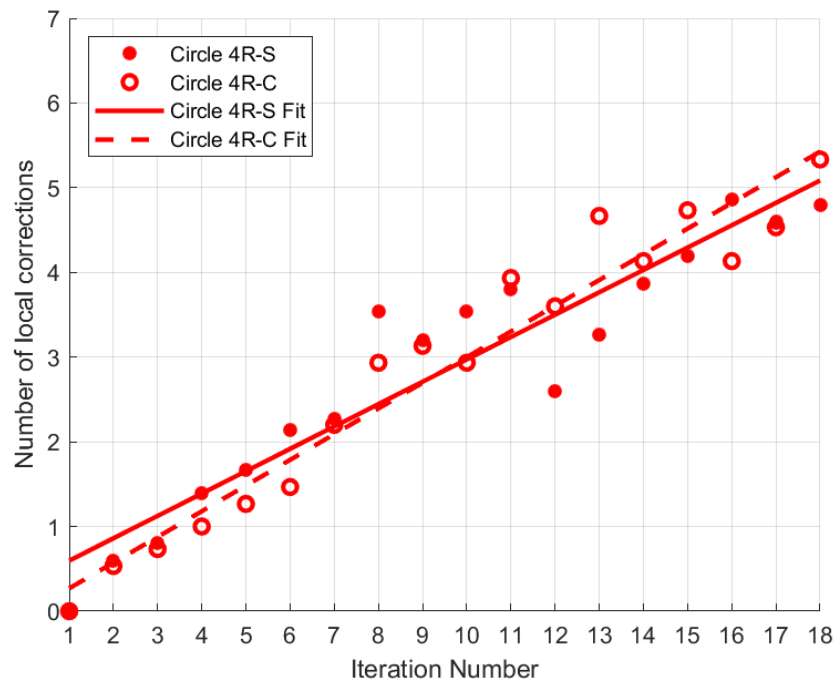


Figure 4.12 – Number of OMC agents whose trajectories are modified regarding the trial number, in the Circle scenario, given that  $i - 1$  OMC agents are present in trial  $i$ , separated for the 4R-S and 4R-C conditions. Linear regressions show that participants triggered local corrections on approximately 30% of the OMC agents in each trial, in both scenarios and conditions.

**Correction Frequency.** Figure 4.12 shows the number of OMC agents whose trajectories were modified, regarding the number of iteration, during the 4R process. As the recording process advances, participants encounter more OMC agents, leading to a rise in the number of OMC agents triggering local corrections. However, it is important to note that participants did not trigger local corrections in all OMC agents in every trial. Through a linear regression, we found that on average 30% of OMC agents experienced local corrections in each iteration, a finding that holds true across both 4R-S and 4R-C conditions.

**Correction Location.** Figure 4.13 displays the local correction occurrence position across responsive OMC agents, depending on their relative walking progress, for the circle scenario. Most local corrections occur around the middle of the trajectory (30 to 50%),

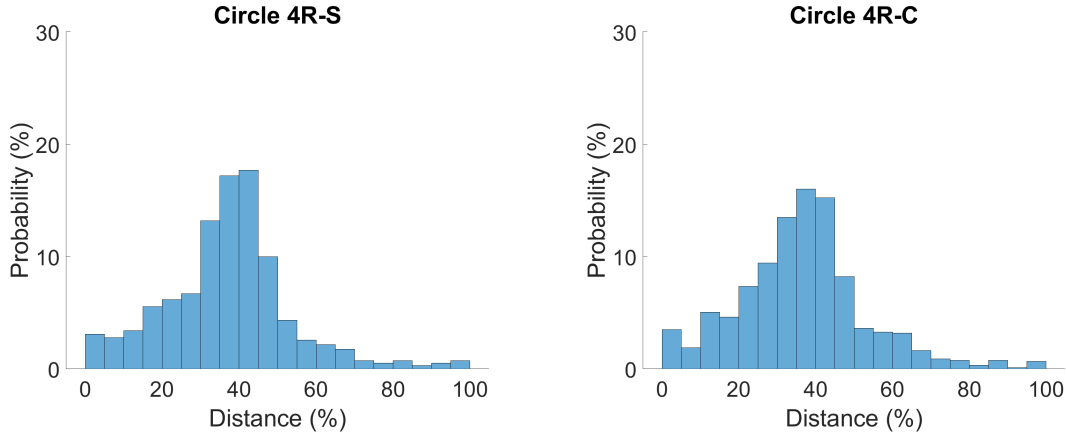


Figure 4.13 – Histograms of local correction occurrence position of responsive OMC agents, for the 4R-S (left) and 4R-C (right) conditions. For both conditions, most local corrections happen around 30% to 50% of the agents’ trajectories in the Circle scenario, suggesting that most local corrections happen in the crossing area.

suggesting that encounters with users happen mostly near the center of the interaction area. We observed similar results between the 4R-S and 4R-C conditions.

**Trajectory Dissimilarities.** As OMC agents become responsive when upcoming collisions with users are detected, we measured after each trial the difference between each OMC agent and its original user-recorded trajectory using DTW. Figure 4.14 shows these differences averaged over participants for each OMC agent and trial, using a matrix-based visualization, which reveals a gradual amplification of this difference as the recording process advances.

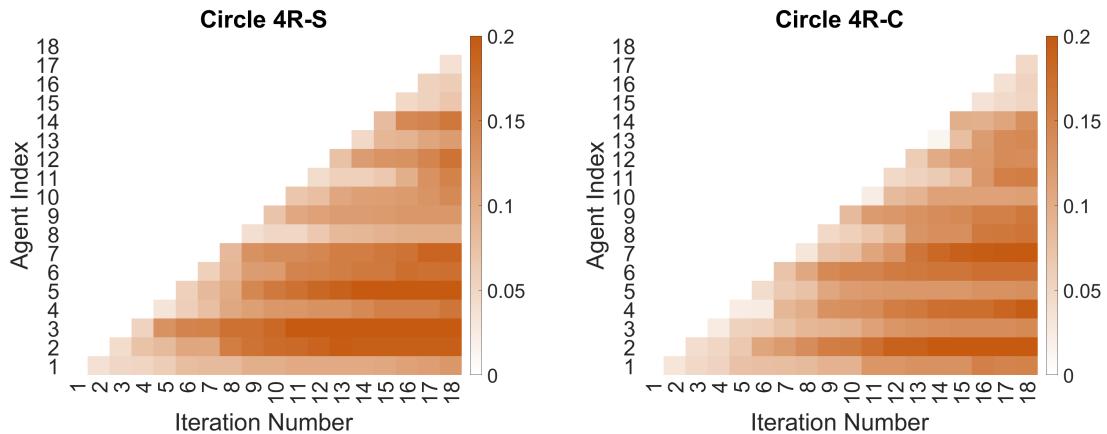


Figure 4.14 – Average trajectory differences (computed using DTW) compared to each OMC agent’s original trajectory for the Circle scenario, for both 4R-S (left) and 4R-C (right) conditions. The color scale represents the error in meters.

### 4.5.3 Scenario V - Crossing Flow

In this section, we present the results of the Crossing Flow scenario. The analysis of the Crossing Flow scenario shared the same framework with the Circle scenario. To avoid redundancy, we highlight the specific results of this scenario without repeating the general interpretation of each analysis.

#### 4.5.3.1 Results on Contextual Crowds

Following the analysis framework (Section 4.5.1), in the study of the contextual crowd results, we focus our illustrations and interpretations on the 2R and 3R-X conditions. Similar to Section 4.5.2.1, the 4R-X results are once again presented in parallel for completeness.

**Average trajectories.** The averaged trajectories obtained in the Crossing Flow scenario are displayed in Figure 4.15. Overall speaking, one can observe that the 2R results present higher walking speed than the 3R-S and 3R-C conditions. The 3R-X conditions, in comparison, present lower speed, especially in the crossing area ( $(x, y) \in [-2, 2]^2$ ). Specifically, the 3R-S average trajectories show higher curvature than the 3R-C condition. This observation is further confirmed by quantitative comparisons that are delivered through Figure 4.17. We remark that the real data presents higher walking speed than our results in all the conditions, being coherent with previous research observations [Banton et al.

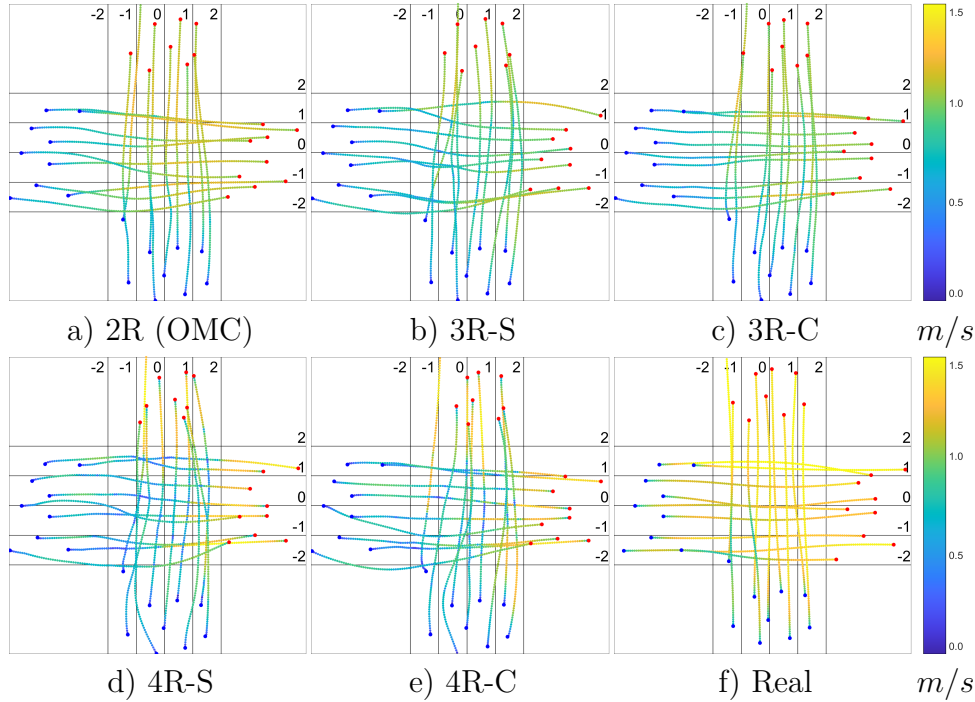


Figure 4.15 – a)-e) Mean trajectories generated under different conditions with all participants aggregated, and colored with the averaged speed. Blue and red discs indicate respectively starting and ending locations. f) The real results from which we replicated the experiment. Equidistant 1 m guidelines (horizontal and vertical) provide visual information about the areas of crossing.

2005; Fink et al. 2007a] on the reduced walking speed in VR environment. We remind our reader that the Figure reports quantities as a function of the progression on the trajectory, normalized by its length. One can consider 0% as the starting position, and 100% as the end position. For most of the agents, 40% to 60% portion of the trajectory can be roughly considered the region where interactions are the most intense. However, this rule does not fit for the agents indexed 2 and 3, as can be seen in Figure 4.7-b, the agents indexed 2 and 3 started walking from a position very close to the area of interaction, indicating an earlier interaction in their walking progress.

**Trajectory Variation.** Figure 4.16 displays the trajectory variation in several milestone iterations during the capture process under each condition. During the first 3 iterations, the 2R results show lower variation compared to the 3R-X conditions. In the 2R results, specifically, we observe extreme convergence in the 2nd and 3rd iterations’ results. This implies that all of our participants have chosen very similar paths when playing these two

iterations during the 2R experience.

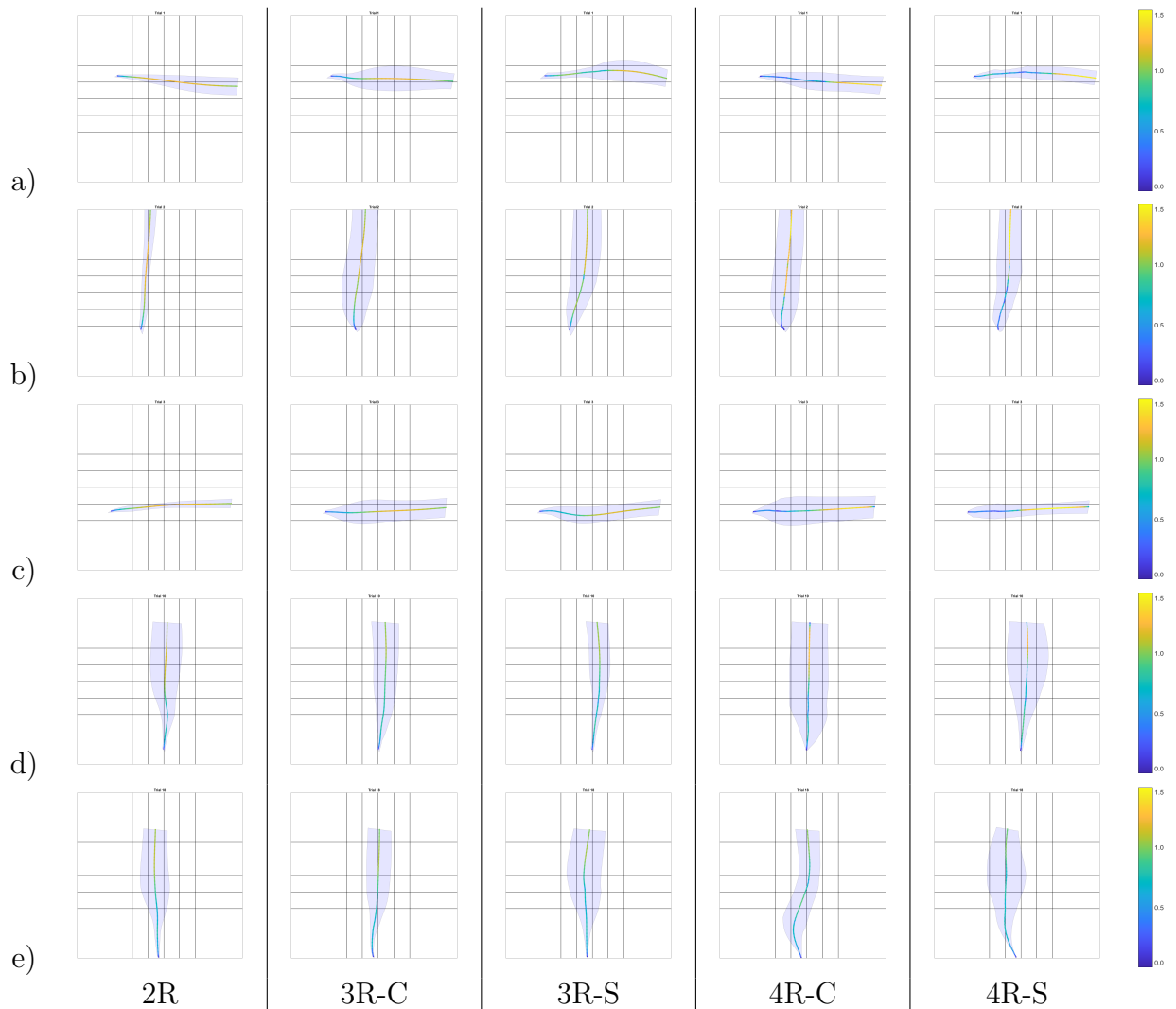


Figure 4.16 – From top to bottom: mean trajectory and variance of iterations a) 1, b) 2, c) 3, d) 10 and e) 18 (last) for the Crossing Flow scenario. From left to right: 2R, 3R-S, 3R-C, 4R-C, and 4R-S paradigms. The color bar reveals the average speed ( $m/s$ ) along the trajectories.

**Normalized Speed.** Figure 4.17-a illustrates the average evolution of the normalized speed in each condition. The curve of our real data reference presents the baseline speed characteristics in this Crossing Flow scenario: the overall pedestrian walking speed continuously increased before arriving in the interaction area, at approximately 20% of the walking progress. Then, we observe speed variation inside the crossing area due to the



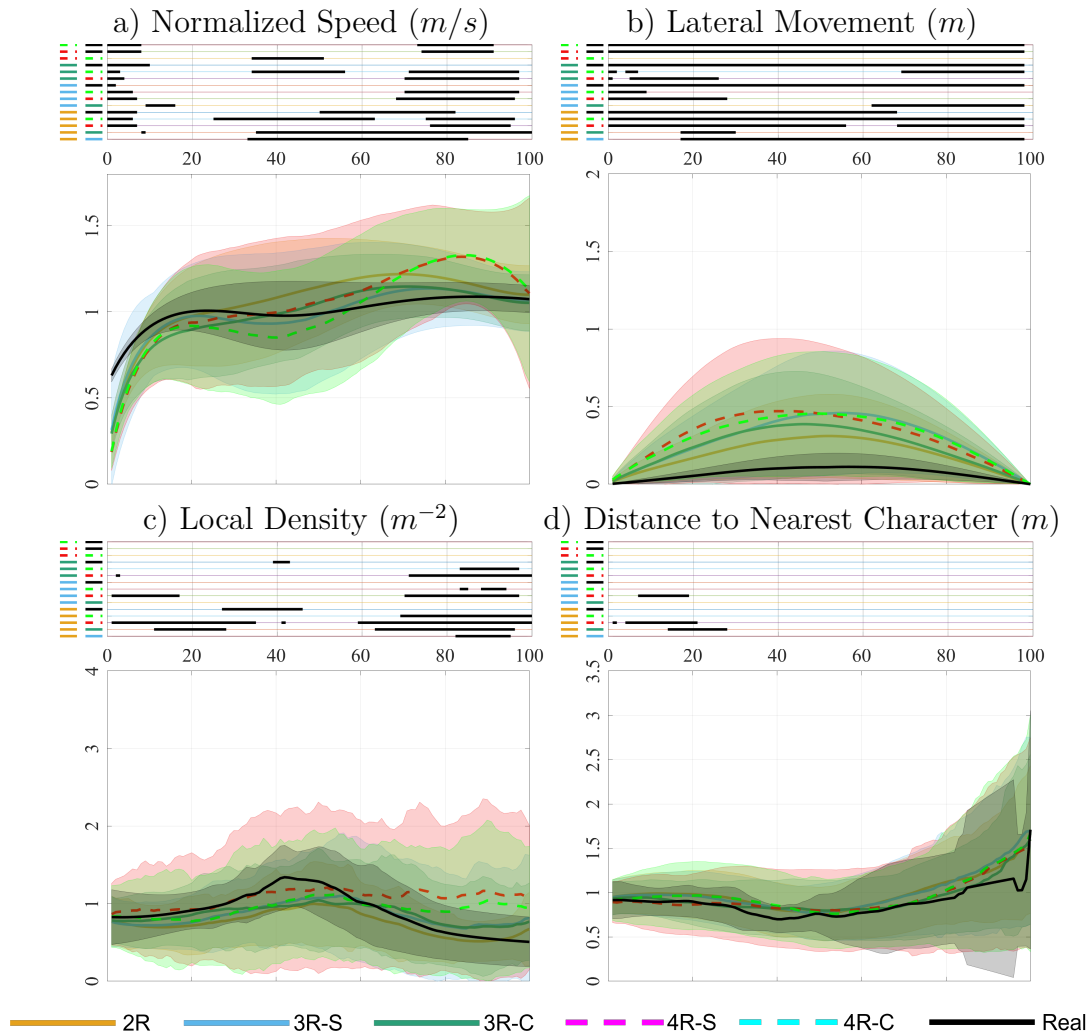


Figure 4.17 – Evolution along the distance that participants walked (in percentage) of participants’ a) normalized speed, b) lateral movement, c) local density, and d) distance to the nearest character. SPM significant differences are displayed above the figure using thick black bars.

interactions with several pedestrians from the other group, resulting in a slightly lower speed in the middle (approximately 45%) of the interaction area. In the 3R-X conditions, despite the quantitative differences in the absolute speed value, the speed variation tendency remains similar to the real data. The 2R results, on the contrary, present a relatively irregular speed evolution curve, where the acceleration phase appears to end around the 70% position. SPM analysis also reveals significant differences between the 2R condition and the real data (51% - 84%), and between the 2R and 3R-X conditions (2R/3R-S: 34%

- 87% and 2R/3R-C: 36% - 100%). Few significant differences were detected between the real and 3R-X results.

**Lateral Movement.** In Figure 4.17-b, where the average lateral movement in each condition is depicted, we observe that while the real data presents extremely low lateral movements, both the 2R and 3R-X results present visibly higher lateral movements. In addition, the 3R-X results are higher than the 2R condition. The SPM analysis indicates that 3R-X results exhibit a 100% significant difference from the real data, while the 2R condition is only significantly different in the 0%-70% range. Other partial differences were observed between the 2R and 3R-X conditions. We highlight that 3R-S shows greater lateral movement compared to 3R-C in the 63% to 100% phase, suggesting that participants tend to make larger lateral movements, and to more significantly deviate from their goal, when faced with simulated contextual crowds.

**Density.** In the Crossing Flow scenario (Figure 4.17-c), the overall density is centered around 1 *ped/m*<sup>2</sup>. The real data shows a slightly higher density in the interaction area when compared to both the 2R and 3R-X conditions (2R/Real: 25%-48%; 3R-C/Real: 40%-45%). SPM analysis also showed some significant differences between the 2R and 3R-X conditions (2R/3R-S: 83%-97%; 2R/3R-C: 12%-30% and 64%-98%).

**Distance to nearest agent.** Figure 4.17-d indicates that the investigation into the nearest neighbor distance did not yield any substantial significant differences among the diverse experimental conditions.

**First vs. Last Trajectories.** Similar to our observation in Section 4.5.2.1, the 2R condition leads to greater disparities between the initial and final recorded trajectories in terms of normalized speed and lateral movement. Specifically, we observe in Figure 4.18, that the 2R results exhibit higher but decreasing average speed than the 3R-X results during the first 5 iterations, and stays in the same speed range of the 3R-X conditions starting from the 6th iteration. In terms of the lateral movement, although the overall lateral movements were restricted to a low level (approximately 0.3 *m*), the 2R results present remarkably lower lateral movements in the first 4 iterations.

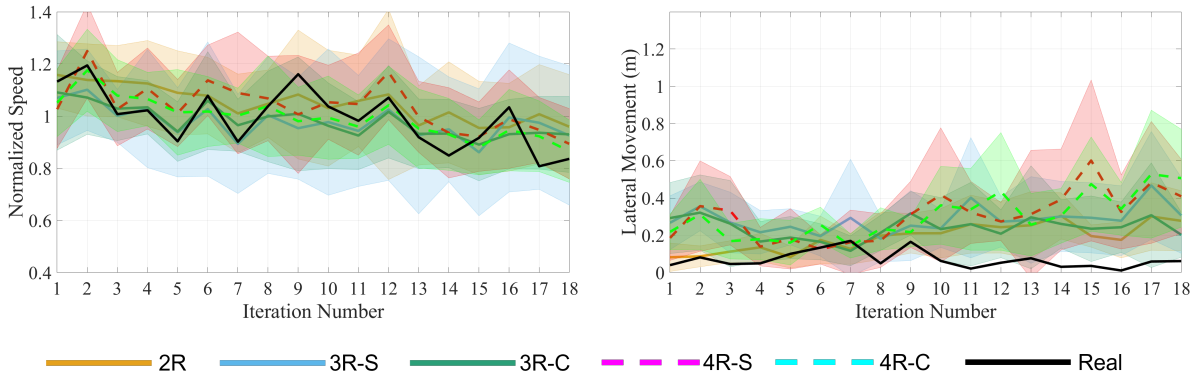


Figure 4.18 – Mean normalized speed and lateral movement over all participants when acting different characters under different conditions in the Crossing Flow scenario, sorted by trial number. Similar to Figure 4.11, the real data is overlapped on the figure for illustrative purposes.

#### 4.5.3.2 Results on Local Correction

**User Collision.** In this Crossing Flow scenario, agents entered the interaction area in a sequential order, meaning that the collision avoidance that participants needed to face was simpler than the Circle scenario. Thus, in Table 4.2, we observed that the number of collisions between participants and OMC agents are lower than in the Circle scenario even for the 2R and 3R-X conditions. Still, we observed reduced numbers of collisions between participants and OMC agents in the 4R-X results compared to the 2R and 3R-X conditions.

Regarding the total collision number at the end of the capture procedure, however, notable higher collision numbers are observed in the 4R-X conditions than in the Circle scenarios. While such phenomena were similarly caused by irregular trajectories, our observations implied the irregular trajectories had more impact on the Crossing Flow scenario than on the Circle scenario.

**Correction Frequency.** Similar to Section 4.5.2.2, we observe in Figure 4.19 that approximately 30% of OMC agents experienced local corrections in each trial.

**Correction Location.** For the scenario of Crossing Flow, although we expected to see the majority of local corrections happening in the middle (40%-60%) of the agents' walking progress (Figure 4.20), we observed that many local corrections happened very close to

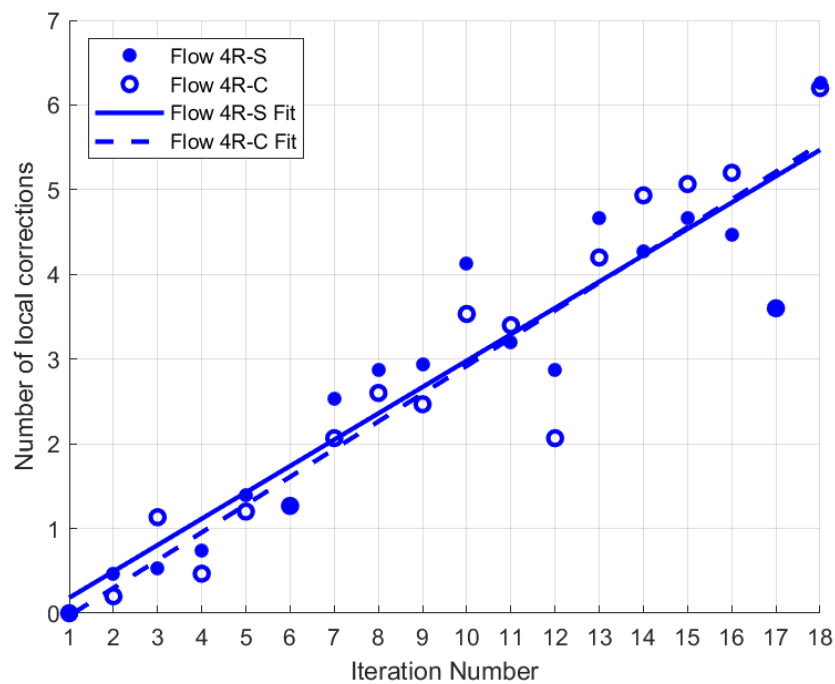


Figure 4.19 – Number of OMC agents whose trajectories are modified regarding the iteration number, in the Crossing Flow scenario, given that  $i - 1$  OMC agents are present in iteration  $i$ , separated for the 4R-S and 4R-C conditions. Linear regressions show that participants triggered local corrections on approximately 30% of the OMC agents in each trial, in both scenarios and conditions.

Table 4.2 – Average number of collisions between participants and OMC agents across trials and participants (top), and total number of collisions between all individuals in the final generated crowd, averaged over participants (bottom). The values are presented in the format *mean ± SD*.

2R (OMC)	3R-S	3R-C	4R-S	4R-C
<b>Number of collisions with OMC agents</b>				
0.23 ± 0.19	0.37 ± 0.30	0.30 ± 0.19	0.18 ± 0.15	0.17 ± 0.17
<b>Number of collisions in the final generated crowd</b>				
1.9 ± 1.7	3.6 ± 3.8	2.7 ± 2.5	6.5 ± 3.1	10.1 ± 4.2

the trajectory’s starting point. This suggests that numerous OMC agents trigger local corrections almost immediately as the trial begins (approximately 40% of OMC agents become reactive in the first 10% of their performed trajectory). This is potentially caused by the close initial proximity of characters in the Crossing Flow scenario, which is further discussed in Section 4.6.2.

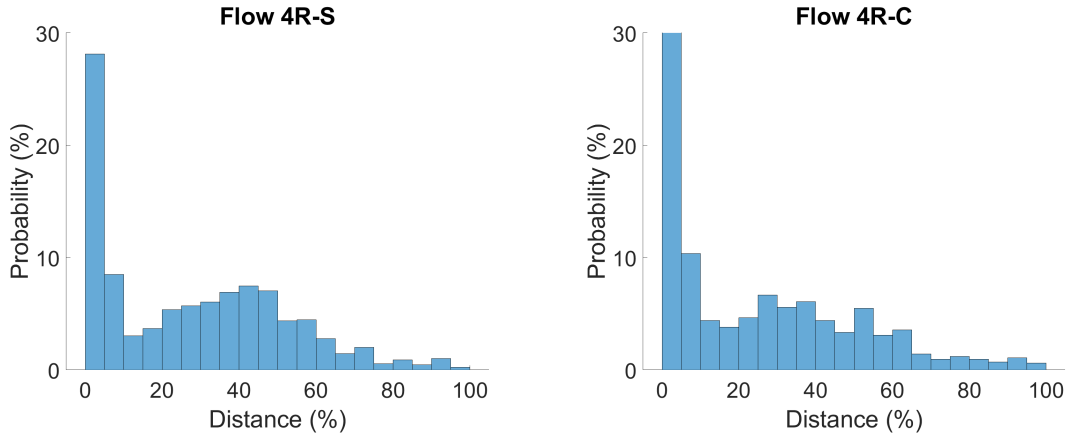


Figure 4.20 – Histograms of local correction occurrence position of responsive OMC agents, in the Crossing Flow scenario, for the 4R-S (left) and 4R-C (right) conditions. Contrary to the Section 4.5.2.2, nearly half of local corrections happen at the beginning of the agents’ trajectories (first 10%) in the Crossing Flow scenario for both 4R-S and 4R-C, suggesting that local corrections happen largely with the agents of the same flow of the participants.

**Trajectory Dissimilarities.** Figure 4.21 shows the differences averaged over participants for each OMC agent and iteration, in the Crossing Flow scenario. The visualization is again based on a matrix representation. We observe a similar evolution in both 4R-S and 4R-C

results. The OMC agents' original trajectories are, in general, gradually degraded by the local correction. However, we remark that, for the agent No.6, the averaged difference between its original and final modified trajectories is tiny compared to the other agents. As this situation can be observed in both of the 4R-X conditions, we believe it is related to the experimental configuration (including the capture order and the condition of triggering local corrections).

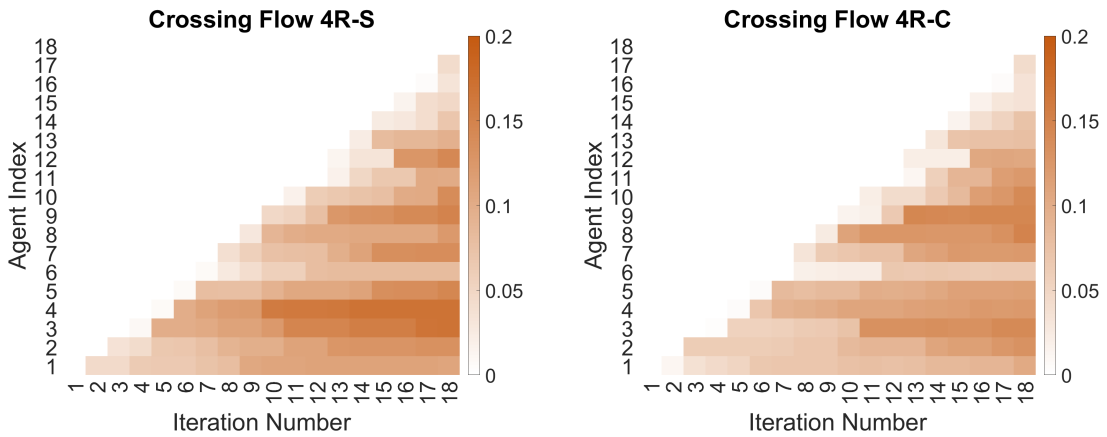


Figure 4.21 – Average trajectory differences (computed using DTW) compared to each OMC agent's original trajectory in the Crossing Flow scenario, for both 4R-S (left) and 4R-C (right) conditions. Although the error is lower than the Circle (see Figure 4.14), we keep the same color scale here for the consistency. The errors are represented in meters.

## 4.6 General Discussion

In this section, we discuss the results of the analyses presented in the previous section. We obtained similar results in both the 4R-S and 4R-C conditions, as well as comparable observations in both the Circle and Crossing Flow scenarios. Thus, our general discussion jointly covers both the two conditions and the two scenarios. The discussion is separated into two parts. In the first part, we discuss the effects of contextual crowd. In the second part, we discuss the local correction strategies.

### 4.6.1 Effect of Contextual Crowds

We first hypothesized (**H1**) that we could mitigate the bias caused by the sequential generation of trajectories inherent to the Record-and-Replace (2R) paradigm by initial-

izing scenarios with contextual crowds, while simultaneously generating trajectories exhibiting characteristics more similar to real data. The results presented in Section 4.5.2.1 and 4.5.3.1 validate **H1**, despite some differences. Results demonstrated that trajectories generated with our Replace-Record-Replace (3R) paradigm are more similar to real trajectories in terms of speed (both scenarios) and lateral deviations (Circle scenario) than those generated using a simpler 2R paradigm. Other characteristics, such as local density and distance to nearest character, were however similar between all the tested conditions.

While some characteristics are seemingly more similar to real data for the 2R conditions, in particular for lateral movements in the Crossing Flow scenario, our results suggest that these similarities might be caused by biases typically observed when relying on the 2R paradigm. Figure 4.11 and 4.18 both illustrate such biases: in the early iterations, where few characters are present in the scene, walking speed is typically high and lateral movement is low, as users rarely interact with other characters. Figure 4.9 and Figure 4.16 provide more concrete illustration of such phenomenon: under the 2R condition, the trajectories generated for the first iteration have extremely low variance and tend to converge on a straight line between the start and end position. As the process unfolds, the walking speed of generated characters drops, while lateral movement significantly increases. While average lateral movement in the 2R paradigm is visually more similar to real data, this higher similarity appears to be due to extremely low deviations in initial trials, biasing the results. In comparison, both walking speed and lateral movement are more consistent throughout the generation process using 3R paradigms. These results suggest that contextual crowds enable us to maintain constant user interactions through the generation process, therefore ensuring that participants undergo pedestrian interactions more similar to those experienced in real-life.

Let us also stress that previous research already demonstrated an increase in users' avoidance distance in Virtual Reality [Gérin-Lajoie et al. 2008; Sanz et al. 2015], compared to real situations. This raises the question of the evaluation of the characteristics of crowd datasets generated in VR. Ideally, evaluations should be based on a comparison of our 3R/4R approaches against  $n$  participants immersed simultaneously to record together a  $n$ -trajectory dataset, which is technically difficult. Nevertheless, a very interesting related question is to search for the number  $m$  of simultaneously immersed participants to record such a  $n$ -trajectory dataset in a 3R fashion that maximizes the dataset quality tradeoff between efficiency ( $m \ll n$ ) and realism ( $m \approx n$ ).

Second, we hypothesized that the type of contextual crowd, whether it is created

through simulation or reanimation from existing data, does not significantly impact the results (**H2**). This is important, because the scarcity of crowd datasets would make difficult the use of XR-C paradigms in practice, and defeat the purpose of capturing complex crowd datasets using only a limited number of persons. Our results indeed show that similar results are obtained between the 3R-C and 3R-S versions we studied, as well as between 4R-C and 4R-S: no significant difference was found between those pairs in the Circle scenario, and only slight ones in the Crossing Flow scenario; where participants showed smaller lateral movements and slower speeds among captured than simulated agents. Our explanation is that participants took advantage of larger adaptations capabilities of simulated agents to adjust their strategy accordingly. These differences, however, remain subtle. Our general conclusion about **H2** is that, since simulated contextual crowds offer a larger flexibility, XR-S paradigms are more relevant in practice. An interesting direction for future work would be to explore more deeply how simulation parameters can actually affect participants self behaviors. For instance a contextual crowd of hurried agents would probably influence a participant differently compared to a crowd of lazy wandering agents.

### 4.6.2 Effects of Local Corrections

As participants are likely to experience unavoidable collisions when faced with non-responsive crowds generated using 2R, our last hypothesis was that providing recorded agents with responsive capabilities would reduce the number of collisions experienced by users, while introducing some differences in the recorded trajectories (**H3**).

First, the results presented in Table 4.1 and 4.2 demonstrate that using responsive agents indeed reduces the number of collisions experienced by users by approximately 50%, which partially validates **H3**. Interestingly, through the comparison between Figure 4.13 and 4.19, we can see the location of these collisions in time is influenced by the scenario. In the Circle scenario, most of the collisions occurred around the middle of the trajectory, i.e., the virtual agents started to avoid users more towards the center of the interaction area where they all converged. In comparison, most of the collisions in the Crossing Flow scenario occurred early while recording new movements, which suggests that responsive capacities were activated for agents initially close to the user, and probably among the same walking group. Our explanation is that, as characters started very close in the Crossing Flow scenario, collision avoidance was activated from the start of the trial. This suggests that the parameters chosen for local corrections (distance of 1 *m* to activate the Responsive component), might have been too conservative for this scenario. This raises



the question whether responsive simulation should be tailored to each scenario, vs. using fixed parameters, which is a question left open for future studies.

In addition, the DTW analysis illustrated in Figure 4.14 and 4.21 shows increasing differences between user-recorded and responsive-agent trajectories in both scenarios. This indicates that local corrections degrade progressively the similarity to the original user-generated trajectories, which fully validates **H3**. At the moment, our understanding of what actually happens when local corrections are triggered is still limited, which is something that we plan to further study in the future. In particular, it would be interesting to more finely quantify such differences, potentially relying on existing trajectory metrics (e.g. [Daniel et al. 2021]), or by including more body-movement characteristics, which are, to our knowledge, currently rarely accounted for in the literature for such questions.

As combining collision avoidance approaches with recorded agent trajectories is still a challenging problem, responsive agents sometimes displayed unrealistic behaviors compared to what we would expect from a real person, such as sudden stops instead of more natural reactions (e.g., slight detours or side-steps). Such behaviors impacted the final trajectories, as well as the studied characteristics. For instance, as responsive agents came to a halt to give priority to participants, thereby reducing the number of potential collisions, this resulted in delays relative to the agents' original trajectory. While the agents proceeded to catch up with their recorded trajectory at their maximum permitted speed, this led to increased average speed in the 4R-X. Simultaneously, only the vicinity of participants activated local corrections (towards both participants and other agents), which were deactivated when agents were catching-up with the recording trajectory (if participants were sufficiently far). This led to an increased number of collisions in the final dataset, as can be seen in Table 4.1 and 4.2. However, as we believe that solving such local collisions is important for interactivity in such a VR-based crowd motion capture context, it would be interesting to explore whether more complex simulation models would provide better local corrections in responsive OMC agents, or to explore the benefits of using post-processing local corrections. Overall, such observations raise novel challenging questions about the general evaluation of VR-based crowd motion capture datasets, their comparison to baseline examples, and how to most appropriately use or adapt to their corresponding real contexts.

## 4.7 Limitations

As a further step of the One-Man-Crowd paradigm, our study still has limitations. First, as the behaviors of the simulation-based autonomous agents are highly related to the simulation algorithm and the parameter configuration, we still lack tools and preliminary knowledge to find the contextual crowds' optimal configuration, not even to conclude whether such an optimal configuration exists for every participant. Although previous research has explored human-agent interaction in VR from several perspectives, and has examined how a human user would perceive simulated virtual agents, universal and adequate metrics still need to be defined to measure the quality of those interactions, as well as to evaluate the *performance* of a virtual crowd.

Second, our study is also limited by the physical constraints from the hardware side. Our tracking system is built using SteamVR-based HMD and base stations. According to our practical experience, this system only allows stable tracking within a  $10\text{ m} \times 10\text{ m}$  area. Thus, we were obliged to adjust the real-world Crossing Flow scenario regarding the physical constraints. As a result, in our experiment, the participants only walked for approximately  $8\text{ m}$  while starting from a position very close to the crossing area. Thus, when acting out some virtual characters, the participants had no time to observe the other group's movement, and tended to let the other group pass first. In comparison, in our real-world reference, the participants walked through approximately  $30\text{ m}$ , starting at  $10\text{ m}$  from the crossing area. Those pedestrians could easily accelerate to their comfort walking speed within such a distance, and had enough time to gather the other group's motion information, to predict the other group's movement, and to plan their path. We believe that this difference is one important reason that led to higher lateral movement in our results than in the real-world results.

## 4.8 Future Work

Regarding our analysis and discovery in this chapter, we identify several possible directions for further exploration or improvements. As mentioned in 4.6, a first interesting direction for future work would be to explore how the participants' behaviors can be influenced by the simulations, and what's the link between the participants' behaviors and the quality of a simulated crowd. For this, one may consider creating an evaluation system of crowd simulation (e.g. Quality function). Although ORCA has a deep influence in the

research society and has been widely used to simulate crowd motion and even to synthesize crowd dataset, previous research has demonstrated numerous simulation algorithms perform better in certain crowd scenarios. A second direction would be to evaluate how the participants' movements retrospectively influence the quality of simulation. During our experiences, we observed that when a human participant walked along the simulated agents, sometimes the simulation performs worse than a pure simulation without human participation. In fact, during the experiment, the captured human trajectories usually present more noise and more irregularity compared to the simulated trajectories. While the human participant makes his/her own decision in real time, he/she is considered a non-responsive obstacle by the simulator. The participants' irregular trajectories make it difficult for the simulator to predict their motion, which possibly influence the entire simulation.

## 4.9 Conclusion

In this chapter, we proposed and compared several new variations for the One-Man-Crowd paradigm for recording increasingly complex and realistic crowd dataset. By leveraging a contextual crowd, we enable human users to evolve among a virtual crowd of autonomous agents at the beginning of the crowd motion capture, and gradually replace each of them to build a complete dataset from his/her own locomotion data. We evaluated these different variations (3R-C, 3R-S, 4R-C, 4R-S), and our results imply that 3R-S is, for the moment, the recommended version in the regard of both the fidelity and flexibility.

Despite the limitations that have been discussed in Section 4.6, our results are promising, and already enable us to recommend the use of the 3R-S paradigm in practice, which opens up novel research directions (Section 4.8).

# GENERAL CONCLUSION & FUTURE WORKS

---

In this thesis, we first reviewed the existing approaches of crowd motion capture, and observed that capturing crowd motion data in the real world still faces multiple technical and ethical challenges. Meanwhile, although VR has exhibited its potential in capturing human behavior, especially when focusing on capturing only one single user's behaviors. To further explore the potential of VR in crowd motion capture, and to avoid the challenges of crowd motion capture in the real world, we proposed the One-Man-Crowd paradigm. This paradigm enables to capture crowd motion in VR with a single user, by iteratively capturing the user's motion and displaying different captured motions in the same virtual scene. Furthermore, we gradually extended the paradigm's usage to more and more complex crowd scenarios. Our research introduces a paradigm capable of meticulously capturing the behavior of every individual in a crowd, a pioneering achievement in the realm of crowd capture techniques. To ascertain the high authenticity of the crowd behavior our method constructs, we employed a validation approach that compares with real experimental data. The findings reveal that the One-Man-Crowd paradigm, despite generating specific crowd behaviors seldom seen in reality under certain circumstances, can replicate crowd behaviors akin to those in the real world using a single user. To further enhance the quality of pedestrian behavior captured by the One-Man-Crowd method, we attempted to introduce contextual crowds and responsive crowds into the original OMC process, encouraging users to exhibit more interactive behaviors. The experimental results indicate that employing contextual crowds indeed elevates the realism of user behaviors. However, our attempt to augment the virtual crowd's interactivity and reduce collisions by locally adjusting the generated trajectories did not yield the anticipated success. This suggests that local modifications of human trajectories within the existing simulation framework remain challenging. Our experimentation and analyses have not only deepened the understanding of this paradigm but also provided valuable insights for its further enhancement.

Our approach proposed in this thesis is, however, only a small step towards a promising solution to high-quality crowd motion capture solution. One can easily see the potential in this approach: the capture of high-quality human interaction motion data in a crowded scene, despite heavy occlusion between agents (from a computer vision perspective). At the time of writing this thesis, mature solutions for human behavior capture, including body motion, hand motion, (HMD-based) eye gaze and (HMD-based) facial expression capture, were already available in the market. Combining all these capture solutions, it is not difficult to imagine that the One-Man-Crowd method and its variants can easily create highly realistic virtual crowd animations, imbuing each virtual pedestrian with a sense of life, and further utilizing such data to study the interaction between subjects and the virtual crowd. Gibson [1958] and Warren [2006] have proposed the concept of the perception-action loop, and in VR-based crowd studies, the behavior of subjects depends on their perception of virtual characters. To our knowledge, existing crowd simulation methods still have not reached the capability of generating highly realistic paths, body movements, eye contacts, and facial expressions simultaneously. Therefore, the behavior of subjects facing unnatural virtual characters is inevitably affected. In contrast, the One-Man-Crowd method allows experimenters to construct a virtual crowd that is close to a real crowd, as its data is directly collected from real human behavior. However, within the scope of this thesis, the crowds' responsiveness of the OMC paradigm is still limited, and its current application scenarios are still limited by this flaw. On the other hand, we note that crowd data constructed by a single user usually shows homogenized behavior characteristics, such as a general tendency to jaywalk or yield. We have tested involving multiple users in the same crowd capture process one by one, successfully obtaining more diverse behaviors, but this also increased the burden of capture operations. Therefore, whether generating crowd data by a single user is the optimal solution remains a question. If multiple users are needed to complete the capture together, how many users are the most appropriate number is also a question that needs further experimental exploration.

## 5.1 Technical Limitations

In this thesis, our research primarily focuses on validating the effectiveness and applicability of the One-Man-Crowd (OMC) paradigm in generating unidirectional and multidirectional motion datasets for crowd simulation. To this end, we replicated several classic controlled experiments within a ten by ten meter space. For certain scenarios,

due to physical constraints of the experimental venues available to us, we had to adjust related experimental parameters to carry out and complete these studies. This implies that further development of the OMC method necessitates overcoming several technical challenges:

1. Higher Motion Capture Accuracy: In our experiments, we employed a tracking solution combining Head-Mounted Displays (HMDs) with a motion capture system. Although Xsens motion capture equipment provides native positioning capabilities, its inertia-based positioning mode can drift in scenarios where the user moves frequently. To ensure the quality of tracking in our experiments, we utilized the optical tracking solution of the HMD to determine the user’s location, and we manually coupled full-body skeletal motion information from Xsens with positional data from the HMD in the virtual environment, achieving acceptable capture results. However, this combined positioning solution has limited accuracy, thus restricting the quality of the animations we could capture. Despite being barely perceptible to participants, the recorded character animations inevitably exhibited feet sliding on the ground, which decreases the realism of the recorded motion data. We believe the use of optical motion tracking system such as Vicon can lead to a better body motion capture results.

2. Large-Scale Motion Capture: Due to our reliance on the SteamVR tracking solution for HMD positioning, our experimental space could not exceed ten by ten meters. Our practical experience indicates that HMD tracking is not effective beyond a 10-meter range. However, controlled experiments in the real world are often conducted on a larger scale, and bigger experimental spaces can accommodate more participants or allow subjects to perform longer, accelerated movements over greater distances, providing them with extended observation times to anticipate virtual crowds from various directions. We believe that improvement of VR systems not relying on external tracking devices will enable to study new and more complex scenarios in the future.

3. Multi-User Experiments: As discussed in Chapter 3, we envisage the M2N concept, where  $M$  ( $M \ll N$ ) users are simultaneously embedded in a virtual environment. This allows for the capture of local real-time interactions, thereby enhancing interactivity during the capture process. We hypothesize that by simultaneously immersing multiple users into the same virtual environment, natural real-time interaction behaviors will emerge during the experience, which will contribute to improving the captured data by reducing the behavioral biases. Indeed, this hypothesis needs to be validated by further experiments. In addition, it would be valuable to explore what value of  $M$  corresponds to

the best compromise between the quality of captured data and the efforts needed in the preparation of the experiment.

4. Impact of Autonomous Agents' Animation Quality on User Experience and Behavior: In this thesis, we assigned a very basic, motion-capture-based animation system to autonomous agents. This system tends to produce non-human-like behaviors in scenarios involving low-speed movements, hesitation, and sudden change of direction. During the experiments, the participants might exhibit defensive strategies and behaviors when they cannot clearly predict the movements of the autonomous agents. According to our empirical observation, the participants broadly behaved more relaxed when facing the non-responsive OMC. Even though the non-responsive OMC agents did not answer to the participants' movements, their human-like behaviors seemed to make it easy for the participants to make their own decisions. Thus, studying how the autonomous agents' motion influences the participants' behavior, as well as in which condition will the participants behave more naturally, will bring valuable insights to improve the human-agent interaction experiences. Indeed, some studies have already started to explore similar directions [Koiliias et al. 2020a; Trivedi and Mousas 2023]. Nevertheless, in addition to simply studying the effects of the autonomous agents' behavior, we believe it is necessary to explore how we can improve the agents' behavior to make the human participants' behave more naturally.

## 5.2 Discussion

In this thesis, we highlight an innovative method for data acquisition, which holds the potential to advance research in related fields significantly. Traditionally, many studies have been constrained by the lack of specific data, limiting the breadth and depth of scientific inquiry. This is particularly true in the domain of crowd interaction behavior simulation, where the absence of a deep understanding of actual behavior patterns often leads to research outcomes that do not accurately reflect reality. Previous studies on the data extraction from video recordings have demonstrated that in the absence of high-quality data, inferring pedestrian behavior using prior knowledge from simulations is an effective method for reconstructing complex pedestrian dynamics. However, this approach also relies on the interpretation of existing data.

The data acquisition method proposed in this research not only opens up new avenues for collecting highly realistic crowd behavior data but also paves the way for new

possibilities in simulation studies. The application of this method has the potential to significantly enhance the realism and predictive capacity of models, thereby advancing the fields of crowd simulation, virtual reality, computer animation, and human-computer interaction. For instance, human pedestrians typically negotiate their paths nonverbally through subtle body language interactions during collision avoidance behavior. However, few studies have delved deeply into this phenomenon. The OMC paradigm makes it possible to collect data on these interaction processes, shedding light on the nuanced communications that govern pedestrian dynamics. By providing researchers with unprecedented quality and richness of data, our method enables a deeper exploration of the complexities of crowd behavior, marking the beginning of a new chapter in research in these domains.

Readers may have noticed that while the research conducted in this thesis has yielded experimental data containing participants' body motion data, we have not delved into the study and analysis of the collected motion data. This constitutes one of our limitations. At the same time, it is important to note that our primary research objective was to validate the comparability of crowd data generated by the OMC paradigm at both the collective and individual levels with real-world experimental outcomes. However, to date, the majority of related research has been predominantly focused on the analysis of trajectories, and even the data we attempted to compare did not include body motion data. This absence of data again limits our ability to further evaluate and analyze individual motion data within crowds, which we hope to explore in the future

Even so, we believe that the research work in this thesis, as well as the proposed crowd motion paradigm, is at the forefront of a new phase in crowd motion studies. Recently, with the rapid development of human body motion analysis, a vast amount of research has utilized extensive human motion datasets to achieve remarkable advancements in the fields of individual motion reconstruction and generation. However, the reconstruction and generation of human-human interactions have received relatively less attention, a major reason being the lack of relevant data. In contrast, crowd movement encompasses a wealth of interaction data, but as previously mentioned, recovering this data still poses significant challenges with existing methods. Recent research in crowds has increasingly focused on detailed individual body motion data [Feldmann and Adrian 2023b; Boomers et al. 2023]. Our method fills this critical gap by introducing an innovative framework that enables the extraction and analysis of intricate body motions within crowds. This not only paves the way for a deeper understanding of individual behaviors within a collective context but also enhances the realism and accuracy of crowd simulation models, significantly contributing



to fields such as virtual reality, urban planning, and emergency evacuation procedures.

## 5.3 Future Works

The One-Man-Crowd (OMC) paradigm represents a significant step forward in capturing and analyzing crowd motion. As we consider the future of OMC, it is crucial to delineate our research into short-term and long-term perspectives.

### 5.3.1 Short-Term Perspectives

From the short-term perspective, the One-Man-Crowd (OMC) paradigm opens up avenues for innovative experimental designs and data enhancement methodologies, fostering a deeper understanding of human-crowd interactions and enriching the quality of crowd motion data.

Firstly, leveraging the OMC framework to design novel experiments provides an intuitive application of this research. A pertinent research topic is the investigation of user responses to virtual crowds with varying levels of realism. The OMC method facilitates the rapid construction of moving virtual crowds as ground truth. Subsequently, simulations with different levels of realism can replicate this ground truth, enabling a comparative analysis of user reactions. Moreover, the OMC approach can swiftly generate specific virtual pedestrian motions in designated scenarios, offering contextual depth to studies such as VR-based investigations of single-user responses in crowd evacuation situations. Essentially, employing the OMC paradigm in experimental setups replaces the labor-intensive process of manually crafting virtual character animations, thereby streamlining research workflows and enhancing experimental fidelity.

Secondly, the data captured via the OMC method holds untapped potential for enhancing the quality of crowd motion data. In our initial contribution, we observed a tendency for a single user to repeat strategies while generating an entire crowd motion, leading to a lack of behavioral diversity in the generated data. If we consider the different crowd datasets produced by various participants, this observation implies that each dataset exhibits a high level of consistency at strategic, tactical, and operational levels, likely rooted in individual personality traits. This leads to an intriguing question: can we extract these personality-based characteristics from each dataset and integrate them into other crowd datasets? Such a fusion could synthesize data with heightened behavioral

diversity, thereby more closely mirroring the nuances of real crowd dynamics. In other words, can we harness the data already collected to perform style transfer across different OMC crowd datasets, enriching the datasets with a broader spectrum of behavioral patterns and interactions?

In conclusion, the short-term research horizons facilitated by the OMC paradigm are twofold: enriching the realm of experimental design in human-virtual crowd interactions and harnessing the potential of existing datasets to foster a richer, more diverse representation of crowd motion. These pursuits not only align with the core objectives of the OMC paradigm but also contribute significantly to the broader field of human-crowd interaction research.

### 5.3.2 Long-Term Perspectives

In the long-term vision for the One-Man-Crowd (OMC) paradigm, a critical goal is to address the current limitations in the realism of simulated agents used as contextual crowds. Notably, these agents exhibit shortcomings in both path planning and behavior dynamics. To surmount these challenges and progressively enhance the verisimilitude of crowd simulations, a multi-faceted approach is essential.

One promising direction involves refining the quality of simulations by leveraging advanced techniques such as motion matching [Holden et al. 2020]. However, it is crucial to recognize that motion matching, like many other cutting-edge methods, is data-driven. Similarly, generative models, which hold the potential to simultaneously generate paths and animations, are also rooted in data-driven methodologies. The scarcity of data pertaining to multi-person interactions in academic settings poses a significant hurdle, as the effectiveness of these data-driven approaches is inherently tied to the richness and relevance of the underlying data. If the available data lacks granular details about individual-level interactions within crowds, the resulting simulations will inevitably fall short of capturing the intricacies of real-world crowd dynamics.

Drawing parallels from the evolution of human motion reconstruction and generation, we can envisage a similar trajectory for crowd motion reconstruction and generation. Central to this vision is the utilization of the authentic user data captured via the OMC paradigm. Despite the current limitations in the simulated agents' behavior and motion quality, the user data remains a trove of real, actionable insights. The contrast between the users' authentic motions and the agents' simulated behaviors presents a unique opportunity. By treating the users' observational data and behavioral patterns as foundational

inputs, we can employ data-driven approaches, especially deep learning, to craft new simulated agents that exhibit markedly improved realism.

These enhanced agents, when deployed as contextual crowds in subsequent experiments, are expected to elicit more genuine and nuanced interactions from users. This, in turn, enriches the user data collected, creating a virtuous cycle. Each iteration of this process not only refines the simulated agents but also brings user behaviors closer to how they would naturally unfold in real-world crowd scenarios. Envisioning this iterative cycle of improvement, we foresee a future where the virtual realm is populated by highly realistic virtual pedestrians, thereby elevating the fidelity of crowd simulations to unprecedented levels.

# BIBLIOGRAPHY

---

- Adrian, Juliane, Maik Boltes, Stefan Holl, Anna Sieben, and Armin Seyfried (2018), « Crowding and queuing in entrance scenarios: influence of corridor width in front of bottlenecks », *in: arXiv preprint arXiv:1810.07424*.
- Adrian, Juliane, Armin Seyfried, and Anna Sieben (2020), « Crowds in front of bottlenecks at entrances from the perspective of physics and social psychology », *in: Journal of the Royal Society Interface* 17.165, p. 20190871.
- Alahi, Alexandre, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese (2016), « Social LSTM: Human Trajectory Prediction in Crowded Spaces », *in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 961–971, URL: <https://api.semanticscholar.org/CorpusID:9854676>.
- Amirian, Javad, Jean-Bernard Hayet, and Julien Pettré (2019), « Social Ways: Learning Multi-Modal Distributions of Pedestrian Trajectories With GANs », *in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2964–2972, URL: <https://api.semanticscholar.org/CorpusID:128346817>.
- Andriyenko, Anton and Konrad Schindler (2011), « Multi-target tracking by continuous energy minimization », *in: CVPR 2011*, pp. 1265–1272, URL: <https://api.semanticscholar.org/CorpusID:3042496>.
- Argelaguet, Ferran, Anne-Hélène Olivier, G. Bruder, Julien Pettré, and Anatole Lécuyer (2015), « Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments », *in: 2015 IEEE Virtual Reality (VR)*, pp. 75–80, URL: <https://api.semanticscholar.org/CorpusID:2592495>.
- Armbrüster, C., M. Wolter, T. Kuhlen, W. Spijkers, and B. Fimm (Feb. 2008), « Depth Perception in Virtual Reality: Distance Estimations in Peri- and Extrapersonal Space », *in: CyberPsychology & Behavior* 11.1, pp. 9–15, DOI: 10.1089/cpb.2007.9935, URL: <http://www.liebertpub.com/doi/10.1089/cpb.2007.9935> (visited on 02/18/2019).
- Asano, Miho, Takamasa Iryo, and Masao Kuwahara (2010), « Microscopic pedestrian simulation model combined with a tactical model for route choice behaviour », *in:*

- 
- Transportation Research Part C-emerging Technologies* 18, pp. 842–855, URL: <https://api.semanticscholar.org/CorpusID:109861821>.
- Al-Azzawi, Marwan and Robert Raeside (2007), « Modeling Pedestrian Walking Speeds on Sidewalks », *in: Journal of Urban Planning and Development-asce* 133, pp. 211–219, URL: <https://api.semanticscholar.org/CorpusID:109515507>.
- Bailenson, Jeremy N, Jim Blascovich, Andrew C Beall, and Jack M Loomis (2001), « Equilibrium theory revisited: Mutual gaze and personal space in virtual environments », *in: Presence: Teleoperators & Virtual Environments* 10.6, pp. 583–598.
- Banton, Tom, Jeanine Stefanucci, Frank Durgin, Adam Fass, and Dennis Proffitt (2005), « The perception of walking speed in a virtual environment », *in: Presence* 14.4, pp. 394–406.
- Berg, Jur P. van den, Ming C. Lin, and Dinesh Manocha (2008), « Reciprocal Velocity Obstacles for real-time multi-agent navigation », *in: 2008 IEEE International Conference on Robotics and Automation*, pp. 1928–1935, URL: <https://api.semanticscholar.org/CorpusID:971270>.
- Berg, van den (2016), « The Influence of Herding on Departure Choice in Case of an Evacuation : Design and Analysis of a Serious Gaming Experimental Set-up », *in: URL: https://api.semanticscholar.org/CorpusID:135017381*.
- Bernardini, Gabriele, Enrico Quagliarini, and Marco D’Orazio (2016), « Towards creating a combined database for earthquake pedestrians’ evacuation models », *in: Safety Science* 82, pp. 77–94, URL: <https://api.semanticscholar.org/CorpusID:109001736>.
- Berseth, Glen, Mubbasir Kapadia, Brandon Haworth, and Petros Faloutsos (2014), « Steer-Fit: automated parameter fitting for steering algorithms », *in: Symposium on Computer Animation*, URL: <https://api.semanticscholar.org/CorpusID:2930328>.
- Berton, Florian, Fabien Grzeskowiak, Alexandre Bonneau, Alberto Jovane, Marco Aggravi, Ludovic Hoyet, Anne-Hélène Olivier, Claudio Pacchierotti, and Julien Pettré (2020a), « Crowd Navigation in VR: Exploring Haptic Rendering of Collisions », *in: IEEE Transactions on Visualization and Computer Graphics* 28, pp. 2589–2601, URL: <https://api.semanticscholar.org/CorpusID:227246203>.
- Berton, Florian, Ludovic Hoyet, Anne-Hélène Olivier, Julien Bruneau, Olivier Le Meur, and Julien Pettré (2020b), « Eye-Gaze Activity in Crowds: Impact of Virtual Reality and Density », *in: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 322–331, URL: <https://api.semanticscholar.org/CorpusID:218560771>.

- 
- Berton, Florian, Anne-Hélène Olivier, Julien Bruneau, Ludovic Hoyet, and Julien Pettre (2019), « Studying Gaze Behaviour during Collision Avoidance with a Virtual Walker: Influence of the Virtual Reality Setup », *in: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 717–725, DOI: 10.1109/VR.2019.8798204.
- Boatright, Cory D., Mubbasir Kapadia, Jennie M. Shapira, and Norman I. Badler (2015), « Generating a multiplicity of policies for agent steering in crowd simulation », *in: Computer Animation and Virtual Worlds* 26, pp. 483–494, URL: <https://api.semanticscholar.org/CorpusID:14371042>.
- Bode, Nikolai W. F., Stefan Holl, Wolfgang Mehner, and Armin Seyfried (2015a), « Disentangling the Impact of Social Groups on Response Times and Movement Dynamics in Evacuations », *in: PLoS ONE* 10, URL: <https://api.semanticscholar.org/CorpusID:12664287>.
- Bode, Nikolai W. F., Armel U. Kemloh Wagoum, and Edward A. Codling (2015b), « Information use by humans during dynamic route choice in virtual crowd evacuations », *in: Royal Society Open Science* 2, URL: <https://api.semanticscholar.org/CorpusID:88261>.
- Boltes, Maik, Armin Seyfried, Bernhard Steffen, and Andreas Schadschneider (2010), « Automatic Extraction of Pedestrian Trajectories from Video Recordings », *in:* URL: <https://api.semanticscholar.org/CorpusID:51919433>.
- Boomers, Ann Katrin, Maik Boltes, Juliane Adrian, Mira Beermann, Mohcine Chraibi, Sina Feldmann, Frank Fiedrich, Niklas Frings, Arne Graf, Alica Kandler, Deniz Kilic, Krisztina Konya, Mira Küpper, Andreas Lotter, Helena Lügering, Francesca Müller, Sarah Paetzke, Anna-Katharina Raytarowski, Olga Sablik, Tobias Schrödter, Armin Seyfried, Anna Sieben, and Ezel Üsten (2023), « Pedestrian Crowd Management Experiments: A Data Guidance Paper », *in: Collective Dynamics*, URL: <https://api.semanticscholar.org/CorpusID:257365844>.
- Bruneau, Julien, Anne-Hélène Olivier, and Julien Pettré (2015), « Going Through, Going Around: A Study on Individual Avoidance of Groups », *in: IEEE Transactions on Visualization and Computer Graphics* 21, pp. 520–528, URL: <https://api.semanticscholar.org/CorpusID:18033483>.
- Burghardt, Sebastian, Armin Seyfried, and Wolfram Klingsch (2013), « Performance of stairs: fundamental diagram and topographical measurements », *in: Transportation Research Part C-emerging Technologies* 37, pp. 268–278, URL: <https://api.semanticscholar.org/CorpusID:108765408>.

- 
- Cao, Lijun, Jing Lin, and Nan Li (2019a), « A virtual reality based study of indoor fire evacuation after active or passive spatial exploration », *in: Comput. Hum. Behav.* 90, pp. 37–45, URL: <https://api.semanticscholar.org/CorpusID:53729163>.
- Cao, Shuchao, Libi Fu, Peng Wang, Guang Zeng, and Weiguo Song (2018), « Experimental and modeling study on evacuation under good and limited visibility in a supermarket », *in: Fire Safety Journal*, URL: <https://api.semanticscholar.org/CorpusID:116495972>.
- Cao, Shuchao, Armin Seyfried, Jun Zhang, Stefan Holl, and Weiguo Song (2017), « Fundamental diagrams for multidirectional pedestrian flows », *in: Journal of Statistical Mechanics: Theory and Experiment* 2017.3, p. 033404.
- Cao, Shuchao, Peng Wang, Ming Yao, and Weiguo Song (2019b), « Dynamic analysis of pedestrian movement in single-file experiment under limited visibility », *in: Communications in Nonlinear Science and Numerical Simulation* 69, pp. 329–342.
- Cao, Shuchao, Jun Zhang, Dan R. Salden, Jian Ma, Chang’an Shi, and Ruifang Zhang (2016), « Pedestrian dynamics in single-file movement of crowd with different age compositions. », *in: Physical review. E* 94 1-1, p. 012312, URL: <https://api.semanticscholar.org/CorpusID:30575579>.
- Carstens, RL and SL Ring (1970), « Pedestrian capacities of shelter entrances », *in: Traffic Engineering, Inst Traffic Engr.*
- Casadiago, Luiselena and Nuria Pelechano (2015), « From One to Many: Simulating Groups of Agents with Reinforcement Learning Controllers », *in: International Conference on Intelligent Virtual Agents*, URL: <https://api.semanticscholar.org/CorpusID:1034675>.
- Charalambous, Panayiotis and Yiorgos Chrysanthou (2014), « The PAG Crowd: A Graph Based Approach for Efficient Data-Driven Crowd Simulation », *in: Computer Graphics Forum* 33, URL: <https://api.semanticscholar.org/CorpusID:18818295>.
- Charalambous, Panayiotis, Ioannis Karamouzas, Stephen J. Guy, and Yiorgos Chrysanthou (2014), « A Data-Driven Framework for Visual Crowd Analysis », *in: Computer Graphics Forum* 33, URL: <https://api.semanticscholar.org/CorpusID:13549632>.
- Chattaraj, Ujjal, Armin Seyfried, and Partha Chakroborty (2009), « Comparison of pedestrian fundamental diagram across cultures », *in: Advances in Complex Systems* 12 (3), pp. 393–405.

- 
- Chen, Juan, SM Lo, and Jian Ma (2017a), « Pedestrian ascent and descent fundamental diagram on stairway », *in: Journal of Statistical Mechanics: Theory and Experiment* 2017.8, p. 083403.
- Chen, Yu Fan, Miao Liu, Michael Everett, and Jonathan P. How (2017b), « Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning », *in: 2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 285–292, DOI: 10.1109/ICRA.2017.7989037.
- Cirio, Gabriel, Anne-Hélène Olivier, Maud Marchal, and Julien Pettré (2013), « Kinematic Evaluation of Virtual Walking Trajectories », *in: IEEE Transactions on Visualization and Computer Graphics* 19.4, pp. 671–680, DOI: 10.1109/TVCG.2013.34.
- Corbetta, Alessandro, Jasper Meeusen, Chung-Min Lee, and Federico Toschi (2016), « Continuous measurements of real-life bidirectional pedestrian flows on a wide walkway », *in: URL: <https://api.semanticscholar.org/CorpusID:55863612>*.
- D. Yanagisawa, A. Tomoeda, and K. Nishinari (2012), « Improvement of pedestrian flow by slow rhythm », *in: Phys. Rev. E* 85, p. 016111.
- D’Orazio, Marco, Gabriele Bernardini, Silvia Tacconi, Valentina Arteconi, and Enrico Quagliarini (2016), « Fire safety in Italian-style historical theatres: How photoluminescent wayfinding can improve occupants’ evacuation with no architecture modifications », *in: Journal of Cultural Heritage* 19, pp. 492–501, URL: <https://api.semanticscholar.org/CorpusID:110473345>.
- D’Orazio, Marco, Luca Spalazzi, Enrico Quagliarini, and Gabriele Bernardini (2014), « Agent-based model for earthquake pedestrians’ evacuation in urban outdoor scenarios: Behavioural patterns definition and evacuation paths choice », *in: Safety Science* 62, pp. 450–465, URL: <https://api.semanticscholar.org/CorpusID:951097>.
- Daamen, W and SP Hoogendoorn (2003a), « Qualitative results from pedestrian laboratory experiments », *in: Pedestrian and evacuation dynamics 2003, Greenwich, CMS Press*, pp. 121–132.
- Daamen, Winnie and Serge Paul Hoogendoorn (2003b), « Experimental Research of Pedestrian Walking Behavior », *in: Transportation Research Record* 1828, pp. 20–30, URL: <https://api.semanticscholar.org/CorpusID:18027517>.
- (2010), « Capacity of doors during evacuation conditions », *in: Procedia Engineering* 3, pp. 53–66, URL: <https://api.semanticscholar.org/CorpusID:110727166>.



- 
- Daamen, Winnie, Serge Paul Hoogendoorn, Mario Campanella, and D. M. Versluis (2014), « Interaction Behavior Between Individual Pedestrians », *in*: URL: <https://api.semanticscholar.org/CorpusID:168421267>.
- Daniel, Beatriz Cabrero, Ricardo Marques, Ludovic Hoyet, Julien Pettré, and Josep Blat (2021), « A Perceptually-Validated Metric for Crowd Trajectory Quality Evaluation », *in*: *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 4, pp. 1–18, URL: <https://api.semanticscholar.org/CorpusID:237347076>.
- Ding, Zhong-Jun, Zhiwei Shen, Ning Guo, Kongjin Zhu, and Jiancheng Long (2020), « Evacuation through area with obstacle that can be stepped over: experimental study », *in*: *Journal of Statistical Mechanics: Theory and Experiment* 2020, URL: <https://api.semanticscholar.org/CorpusID:214049314>.
- Do, Truong Lam, Milad Haghani, and Majid Sarvi (2016), « Group and Single Pedestrian Behavior in Crowd Dynamics », *in*: *Transportation Research Record* 2540, pp. 13–19, URL: <https://api.semanticscholar.org/CorpusID:112910873>.
- Duives, Dorine C., Winnie Daamen, and Serge Paul Hoogendoorn (2013), « State-of-the-art crowd motion simulation models », *in*: *Transportation Research Part C-emerging Technologies* 37, pp. 193–209, URL: <https://api.semanticscholar.org/CorpusID:111331383>.
- (2014), « Influence of Group Size and Group Composition on the Adhered Distance Headway », *in*: *Transportation research procedia* 2, pp. 183–188, URL: <https://api.semanticscholar.org/CorpusID:167654870>.
- Dutra, Teofilo Bezerra, Ricardo Marques, Joaquim B. Cavalcante Neto, Creto Augusto Vidal, and Julien Pettré (2017), « Gradient-based steering for vision-based crowd simulation algorithms », *in*: *Computer Graphics Forum* 36, URL: <https://api.semanticscholar.org/CorpusID:11316267>.
- Evans, Gary W and Roger B Howard (1973), « Personal space. », *in*: *Psychological bulletin* 80.4, p. 334.
- Ezaki, Takahiro, Kazumichi Ohtsuka, Daichi Yanagisawa, and Katsuhiro Nishinari (2015), « Inflow process: A counterpart of evacuation », *in*: *Traffic and Granular Flow'13*, Springer, pp. 227–231.
- Ezaki, Takahiro, Daichi Yanagisawa, Kazumichi Ohtsuka, and Katsuhiro Nishinari (2012), « Simulation of space acquisition process of pedestrians using proxemic floor field model », *in*: *Physica A: Statistical Mechanics and its Applications* 391.1-2, pp. 291–299.

- 
- Fang, Zhi Ming, Li Xue Jiang, Xiao Lian Li, Wen Qi, and Ling Zhu Chen (2019), « Experimental study on the movement characteristics of 5–6 years old Chinese children when egressing from a pre-school building », *in: Safety science* 113, pp. 264–275.
- Fang, Zhiming, Weiguo Song, Jun Zhang, and Hao Wu (2010), « Experiment and modeling of exit-selecting behaviors during a building evacuation », *in: Physica A-statistical Mechanics and Its Applications* 389, pp. 815–824, URL: <https://api.semanticscholar.org/CorpusID:122115736>.
- Feldmann, Sina and Juliane Adrian (2023a), « Forward propagation of a push through a row of people », *in: Safety Science*, URL: <https://api.semanticscholar.org/CorpusID:258502930>.
- (2023b), « Forward propagation of a push through a row of people », *in: Safety Science* 164, p. 106173, DOI: <https://doi.org/10.1016/j.ssci.2023.106173>, URL: <https://www.sciencedirect.com/science/article/pii/S0925753523001157>.
- Feng, Yan, Dorine C. Duives, Winnie Daamen, and Serge Paul Hoogendoorn (2021), « Data collection methods for studying pedestrian behaviour: A systematic review », *in: Building and Environment*, URL: <https://api.semanticscholar.org/CorpusID:224971390>.
- Feng, Yan, Dorine C. Duives, and Serge Paul Hoogendoorn (2020), « Developing a VR tool for studying pedestrian movement and choice behavior », *in: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 815–816, URL: <https://api.semanticscholar.org/CorpusID:218597634>.
- Feng, Yan and Dewei Li (2016), « An Empirical study and a Conceptual Model on Heterogeneity of Pedestrian Social Groups for Friend-group and Family-group », *in:* URL: <https://api.semanticscholar.org/CorpusID:226991178>.
- Fink, Philip W, Patrick S Foo, and William H Warren (2007a), « Obstacle avoidance during walking in real and virtual environments », *in: ACM Transactions on Applied Perception (TAP)* 4.1, 2–es.
- (2007b), « Obstacle avoidance during walking in real and virtual environments », *in: ACM Trans. Appl. Percept.* 4, p. 2, URL: <https://api.semanticscholar.org/CorpusID:12373068>.
- Fridolf, Karl, Daniel Nilsson, and Håkan Frantzich (2016), « Evacuation of a Metro Train in an Underground Rail Transportation System: Flow Rate Capacity of Train Exits, Tunnel Walking Speeds and Exit Choice », *in: Fire Technology* 52, pp. 1481–1518, URL: <https://api.semanticscholar.org/CorpusID:110214152>.

- 
- Fridolf, Karl, Enrico Ronchi, Daniel Nilsson, and Håkan Frantzich (2013), « Movement speed and exit choice in smoke-filled rail tunnels », *in: Fire Safety Journal* 59, pp. 8–21, URL: <https://api.semanticscholar.org/CorpusID:111365473>.
- Friston, Karl, John Ashburner, Stefan Kiebel, Thomas Nichols, and William Penny (2007), « Statistical Parametric Mapping », *in: Academic Press*.
- Friston, Sebastian, Ben J. Congdon, David Swapp, Lisa Izzouzi, Klara Brandstätter, Daniel Archer, Otto Olkkonen, Felix J. Thiel, and Anthony Steed (2021), « Ubiq: A System to Build Flexible Social Virtual Reality Experiences », *in: Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, URL: <https://api.semanticscholar.org/CorpusID:244731608>.
- Fruin, John J (1971a), *Pedestrian planning and design*, tech. rep.
- (1971b), « DESIGNING FOR PEDESTRIANS: A LEVEL-OF-SERVICE CONCEPT », *in: Highway Research Record*, URL: <https://api.semanticscholar.org/CorpusID:112610444>.
- Galea, Edwin R., Hui Xie, Steven Deere, David Cooney, and Lazaros Filippidis (2017), « Evaluating the effectiveness of an improved active dynamic signage system using full scale evacuation trials », *in: Fire Safety Journal* 91, pp. 908–917, URL: <https://api.semanticscholar.org/CorpusID:114738584>.
- Garcimartín, Angel, Iker Zuriguel, José Martín Pastor, César Martín-Gómez, and Daniel R. Parisi (2014), « Experimental evidence of the "Faster Is Slower" effect », *in: Transportation research procedia* 2, pp. 760–767, URL: <https://api.semanticscholar.org/CorpusID:55521491>.
- Gérin-Lajoie, Martin, Carol L Richards, Joyce Fung, and Bradford J McFadyen (2008), « Characteristics of personal space during obstacle circumvention in physical and virtual environments », *in: Gait & posture* 27.2, pp. 239–247.
- Geuss, Michael, Jeanine Stefanucci, Sarah Creem-Regehr, and William B Thompson (2010), « Can I pass? Using affordances to measure perceived size in virtual environments », *in: Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pp. 61–64.
- Gibson, James J (1958), « Visually controlled locomotion and visual orientation in animals », *in: British journal of psychology* 49.3, pp. 182–194.
- Gu, Tianpei, Guangyi Chen, Junlong Li, Chunze Lin, Yongming Rao, Jie Zhou, and Jiwen Lu (2022), « Stochastic Trajectory Prediction via Motion Indeterminacy Diffusion », *in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*

- 
- (*CVPR*), pp. 17092–17101, URL: <https://api.semanticscholar.org/CorpusID:247748591>.
- Gulhare, Siddhartha, Ashish Verma, and Partha Chakroborty (2018), « Comparison of pedestrian data of single file movement collected from controlled pedestrian experiment and from field in mass religious gathering », *in: Collective Dynamics* 3, pp. 1–14.
- Guo, Renyong, Haijun Huang, and Sze Chun Wong (2012), « Route choice in pedestrian evacuation under conditions of good and zero visibility: Experimental and simulation results », *in: Transportation Research Part B-methodological* 46, pp. 669–686, URL: <https://api.semanticscholar.org/CorpusID:153600852>.
- Gupta, Agrim, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi (2018), « Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks », *in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2255–2264, URL: <https://api.semanticscholar.org/CorpusID:4461350>.
- Guy, Stephen J, Jatin Chhugani, Sean Curtis, Pradeep Dubey, Ming C Lin, and Dinesh Manocha (2010), « PLEdestrians: A Least-Effort Approach to Crowd Simulation. », *in: Symposium on computer animation*, pp. 119–128.
- Guy, Stephen J., Jur P. van den Berg, Wenxi Liu, Rynson W. H. Lau, Ming C. Lin, and Dinesh Manocha (2012), « A statistical similarity measure for aggregate crowd dynamics », *in: ACM Transactions on Graphics (TOG)* 31, pp. 1–11, URL: <https://api.semanticscholar.org/CorpusID:13866273>.
- Guy, Stephen J., Sujeong Kim, Ming C. Lin, and Dinesh Manocha (2011), « Simulating heterogeneous crowd behaviors using personality trait theory », *in: Symposium on Computer Animation*, URL: <https://api.semanticscholar.org/CorpusID:1478678>.
- Haghani, Milad (2020), « Empirical methods in pedestrian, crowd and evacuation dynamics: Part II. Field methods and controversial topics », *in: Safety Science* 129, p. 104760, URL: <https://api.semanticscholar.org/CorpusID:218963983>.
- Haghani, Milad and Majid Sarvi (2016), « Pedestrian crowd tactical-level decision making during emergency evacuations », *in: Journal of Advanced Transportation* 50, pp. 1870–1895, URL: <https://api.semanticscholar.org/CorpusID:113697420>.
- (2017a), « Following the crowd or avoiding it? Empirical investigation of imitative behaviour in emergency escape of human crowds », *in: Animal Behaviour* 124, pp. 47–56, URL: <https://api.semanticscholar.org/CorpusID:53159554>.

- 
- Haghani, Milad and Majid Sarvi (2017b), « Stated and revealed exit choices of pedestrian crowd evacuees », *in: Transportation Research Part B-methodological* 95, pp. 238–259, URL: <https://api.semanticscholar.org/CorpusID:157762168>.
- (2018), « Crowd behaviour and motion: Empirical methods », *in: Transportation Research Part B-methodological* 107, pp. 253–294, URL: <https://api.semanticscholar.org/CorpusID:157168967>.
- Haghani, Milad, Majid Sarvi, and Abbas Rajabifard (2017), « Common Misconceptions about Herd-type Behavior in Emergency Evacuations of Pedestrian Crowds », *in:* URL: <https://api.semanticscholar.org/CorpusID:67418722>.
- Haworth, Brandon, Glen Berseth, Seonghyeon Moon, Petros Faloutsos, and Mubbasir Kapadia (2020), « Deep Integration of Physical Humanoid Control and Crowd Navigation », *in: Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games*, URL: <https://api.semanticscholar.org/CorpusID:227121236>.
- He, Feixiang, Yuanhang Xiang, Xi Zhao, and He Wang (2020), « Informative scene decomposition for crowd analysis, comparison and simulation guidance », *in: ACM Transactions on Graphics (TOG)* 39, 50:1–50:13, URL: <https://api.semanticscholar.org/CorpusID:216641644>.
- Helbing, Dirk, Lubos Buzna, Anders Johansson, and Torsten Werner (2005), « Self-Organized Pedestrian Crowd Dynamics: Experiments, Simulations, and Design Solutions », *in: Transp. Sci.* 39, pp. 1–24, URL: <https://api.semanticscholar.org/CorpusID:36666996>.
- Helbing, Dirk, Anders Johansson, and Habib Z. Al-Abideen (2007), « Dynamics of crowd disasters: an empirical study. », *in: Physical review. E, Statistical, nonlinear, and soft matter physics* 75 4 Pt 2, p. 046109, URL: <https://api.semanticscholar.org/CorpusID:21398961>.
- Helbing, Dirk, Anders Johansson, Joachim Mathiesen, Mogens H Jensen, and Alex Hansen (2006), « Analytical approach to continuous and intermittent bottleneck flows », *in: Physical review letters* 97.16, p. 168001.
- Helbing, Dirk and Peter Molnar (May 1995), « Social Force Model for Pedestrian Dynamics », *en, in: Physical Review E* 51.5, arXiv: cond-mat/9805244, pp. 4282–4286, DOI: 10.1103/PhysRevE.51.4282, URL: <http://arxiv.org/abs/cond-mat/9805244> (visited on 02/18/2019).

- 
- Helbing, Dirk and Péter Molnár (May 1995), « Social force model for pedestrian dynamics », *in: Phys. Rev. E* 51 (5), pp. 4282–4286, DOI: 10.1103/PhysRevE.51.4282, URL: <https://link.aps.org/doi/10.1103/PhysRevE.51.4282>.
- Helbing, Dirk and Pratik Mukerji (2012), « Crowd disasters as systemic failures: analysis of the Love Parade disaster », *in: EPJ Data Science* 1, pp. 1–40, URL: <https://api.semanticscholar.org/CorpusID:25531666>.
- Heliövaara, Simo, Juha-Matti Kuusinen, Tuomo Rinne, Timo O. Korhonen, and Harri Ehtamo (2012), « Pedestrian behavior and exit selection in evacuation of a corridor – An experimental study », *in: Safety Science* 50, pp. 221–227, URL: <https://api.semanticscholar.org/CorpusID:110921030>.
- Holden, Daniel, Oussama Kanoun, Maksym Perepichka, and Tiberiu Popa (2020), « Learned motion matching », *in: ACM Transactions on Graphics (TOG)* 39, 53:1–53:12, URL: <https://api.semanticscholar.org/CorpusID:221105494>.
- Hoogendoorn, Serge Paul and Piet H. L. Bovy (2004), « Pedestrian route-choice and activity scheduling theory and models », *in: Transportation Research Part B-methodological* 38, pp. 169–190, URL: <https://api.semanticscholar.org/CorpusID:14252376>.
- Huang, Shenshi, Teng Zhang, Siuming Lo, Shouxiang Lu, and Changhai Li (2018), « Experimental study of individual and single-file pedestrian movement in narrow seat aisle », *in: Physica A: Statistical Mechanics and its Applications* 509, pp. 1023–1033.
- Huber, Markus, Yi-Huang Su, Melanie Krüger, Katrin Faschian, Stefan Glasauer, and Joachim Hermsdörfer (2014), « Adjustments of Speed and Path when Avoiding Collisions with Another Pedestrian », *in: PLoS ONE* 9, URL: <https://api.semanticscholar.org/CorpusID:16738247>.
- Huo, Feizhou, Weiguo Song, Lei Chen, Chi Liu, and Kim Meow Liew (2016), « Experimental study on characteristics of pedestrian evacuation on stairs in a high-rise building », *in: Safety Science* 86, pp. 165–173, URL: <https://api.semanticscholar.org/CorpusID:112007670>.
- « Improved Social Force Models Considering Heterogenous Characteristics among Social Groups » (2018), *in:* URL: <https://api.semanticscholar.org/CorpusID:264609881>.
- Ivanovic, B. and Marco Pavone (2018), « The Trajectron: Probabilistic Multi-Agent Trajectory Modeling With Dynamic Spatiotemporal Graphs », *in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2375–2384, URL: <https://api.semanticscholar.org/CorpusID:129945241>.

- 
- Jelić, A., C. Appert-Rolland, S. Lemercier, and J. Pettré (2012a), « Properties of pedestrians walking in line – Fundamental diagrams », *in: Phys. Rev. E* 85, p. 036111, DOI: 10.1103/PhysRevE.85.036111.
- (2012b), « Properties of pedestrians walking in line. II. Stepping behavior », *in: Phys. Rev. E* 86, p. 046111.
- Jeon, Gyu-Yeob, Ju-Young Kim, Won-Hwa Hong, and Godfried Augenbroe (2011), « Evacuation performance of individuals in different visibility conditions », *in: Fuel and Energy Abstracts*, URL: <https://api.semanticscholar.org/CorpusID:93080577>.
- Johansson, Anders, Dirk Helbing, Habib Z. Al-Abideen, and Salim Al-Bosta (2008), « From Crowd Dynamics to Crowd Safety: a Video-Based Analysis », *in: Adv. Complex Syst.* 11, pp. 497–527, URL: <https://api.semanticscholar.org/CorpusID:8058328>.
- Jovane, Alberto, Pierre Raimbaud, Katja Zibrek, Claudio Pacchierotti, Marc Christie, Ludovic Hoyet, Anne-Hélène Olivier, and Julien Pettré (2022), « Warping character animations using visual motion features », *in: Comput. Graph.* 110, pp. 38–48, URL: <https://api.semanticscholar.org/CorpusID:253920304>.
- Karamouzas, Ioannis, Peter Heil, Pascal van Beek, and Mark H. Overmars (2009), « A Predictive Collision Avoidance Model for Pedestrian Simulation », *in: Motion in Games*, URL: <https://api.semanticscholar.org/CorpusID:14427744>.
- Karamouzas, Ioannis, Brian Skinner, and Stephen J. Guy (2014), « Universal power law governing pedestrian interactions. », *in: Physical review letters* 113 23, p. 238701, URL: <https://api.semanticscholar.org/CorpusID:14431563>.
- Karamouzas, Ioannis, Nick Sohre, Rahul Narain, and Stephen J. Guy (2017), « Implicit crowds », *in: ACM Transactions on Graphics (TOG)* 36, pp. 1–13, URL: <https://api.semanticscholar.org/CorpusID:541809>.
- Kim, Chanho, Fuxin Li, Arridhana Ciptadi, and James M. Rehg (2015), « Multiple Hypothesis Tracking Revisited », *in: 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 4696–4704, URL: <https://api.semanticscholar.org/CorpusID:8671079>.
- Kinateder, Max, Brittany A. Comunale, and William H. Warren (2018), « Exit choice in an emergency evacuation scenario is influenced by exit familiarity and neighbor behavior », *in: Safety Science*, URL: <https://api.semanticscholar.org/CorpusID:117448264>.
- Kinateder, Max, Mathias Müller, Michael Jost, Andreas Mühlberger, and Paul Pauli (2014a), « Social influence in a virtual tunnel fire—influence of conflicting information

- 
- on evacuation behavior. », *in: Applied ergonomics* 45 6, pp. 1649–59, URL: <https://api.semanticscholar.org/CorpusID:28036289>.
- Kinateder, Max, Enrico Ronchi, Daniel Gromer, Mathias Müller, Michael Jost, Markus Nehfischer, Andreas Mühlberger, and Paul Pauli (2014b), « Social influence on route choice in a virtual reality tunnel fire », *in: Transportation Research Part F-traffic Psychology and Behaviour* 26, pp. 116–125, URL: <https://api.semanticscholar.org/CorpusID:146475674>.
- Kinateder, Max and William H. Warren (2016), « Social Influence on Evacuation Behavior in Real and Virtual Environments », *in: Frontiers Robotics AI* 3, p. 43, URL: <https://api.semanticscholar.org/CorpusID:7014414>.
- Kinateder, Max, William H. Warren, and Karen B. Schloss (2019), « What color are emergency exit signs? Egress behavior differs from verbal report. », *in: Applied ergonomics* 75, pp. 155–160, URL: <https://api.semanticscholar.org/CorpusID:54565845>.
- Kobes, Margrethe, Ira Helsloot, Bauke de Vries, and Joseph Post (2010a), « Exit choice, (pre-)movement time and (pre-)evacuation behaviour in hotel fire evacuation - Behavioural analysis and validation of the use of serious gaming in experimental research », *in:* URL: <https://api.semanticscholar.org/CorpusID:65089400>.
- Kobes, Margrethe, Ira Helsloot, Bauke de Vries, Joseph Post, N Oberije, and Karin H Groenewegen (2010b), « Way finding during fire evacuation; an analysis of unannounced fire drills in a hotel at night », *in: Building and Environment* 45, pp. 537–548, URL: <https://api.semanticscholar.org/CorpusID:111250153>.
- Koilias, Alexandros, Christos Mousas, and Christos-Nikolaos E. Anagnostopoulos (2020a), « I feel a moving crowd surrounds me: Exploring tactile feedback during immersive walking in a virtual crowd », *in: Computer Animation and Virtual Worlds* 31, URL: <https://api.semanticscholar.org/CorpusID:222278545>.
- Koilias, Alexandros, Michael G. Nelson, Christos-Nikolaos E. Anagnostopoulos, and Christos Mousas (2020b), « Immersive walking in a virtual crowd: The effects of the density, speed, and direction of a virtual crowd on human movement behavior », *in: Computer Animation and Virtual Worlds* 31, URL: <https://api.semanticscholar.org/CorpusID:225837175>.
- Konya, Krisztina and Anna Sieben (2023), « Waiting and walking with strangers: a socio-psychological pedestrian experiment on joint action in anonymous situations », *in: Royal Society Open Science* 10, URL: <https://api.semanticscholar.org/CorpusID:259091648>.



- 
- Krausz, Barbara and Christian Bauckhage (2012), « Loveparade 2010: Automatic video analysis of a crowd disaster », *in: Comput. Vis. Image Underst.* 116, pp. 307–319, URL: <https://api.semanticscholar.org/CorpusID:20249873>.
- Kretz, Tobias, Anna Grünebohm, and Michael Schreckenberg (2006), « Experimental study of pedestrian flow through a bottleneck », *in: Journal of Statistical Mechanics: Theory and Experiment* 2006, P10014–P10014, URL: <https://api.semanticscholar.org/CorpusID:17271988>.
- Kruchten, Cornelia von and Andreas Schadschneider (2017), « Empirical study on social groups in pedestrian evacuation dynamics », *in: Physica A-statistical Mechanics and Its Applications* 475, pp. 129–141, URL: <https://api.semanticscholar.org/CorpusID:119336974>.
- Kwiatkowski, Ariel, Vicky Kalogeiton, Julien Pettré, and Marie-Paule Cani (2023), « Understanding reinforcement learned crowds », *in: Computers & Graphics* 110, pp. 28–37, DOI: <https://doi.org/10.1016/j.cag.2022.11.007>, URL: <https://www.sciencedirect.com/science/article/pii/S0097849322002035>.
- Lam, William H. K., John F. Morrall, and Herbert H. Ho (1995), « PEDESTRIAN FLOW CHARACTERISTICS IN HONG KONG », *in: Transportation Research Record*, pp. 56–62, URL: <https://api.semanticscholar.org/CorpusID:108128987>.
- Larsson, Anna, Elin Ranudd, Enrico Ronchi, Aoife L. E. Hunt, and Steve M. V. Gwynne (2021), « The impact of crowd composition on egress performance », *in: Fire Safety Journal*, URL: <https://api.semanticscholar.org/CorpusID:219763136>.
- Lee, Jaedong, Jungdam Won, and Jehee Lee (2018), « Crowd simulation by deep reinforcement learning », *in: Proceedings of the 11th ACM SIGGRAPH Conference on Motion, Interaction and Games*, URL: <https://api.semanticscholar.org/CorpusID:53222730>.
- Lemercier, Samuel, Asja Jelic, Richard Kulpa, Jiale Hua, Jérôme Fehrenbach, Pierre Degond, Cécile Appert-Rolland, Stéphane Donikian, and Julien Pettré (2012), « Realistic following behaviors for crowd simulation », *in: Computer Graphics Forum*, vol. 31, 2pt2, Wiley Online Library, pp. 489–498.
- Lerner, Alon, Yiorgos Chrysanthou, and Dani Lischinski (2007), « Crowds by Example », *in: Computer Graphics Forum* 26, URL: <https://api.semanticscholar.org/CorpusID:17374844>.

- 
- Lerner, Alon, Yiorgos Chrysanthou, Ariel Shamir, and Daniel Cohen-Or (2009), « Data Driven Evaluation of Crowds », *in: Motion in Games*, URL: <https://api.semanticscholar.org/CorpusID:2965732>.
- Li, Hongliu, Jun Zhang, Long Xia, Weiguo Song, and Nikolai W. F. Bode (2019), « Comparing the route-choice behavior of pedestrians around obstacles in a virtual experiment and a field study », *in: Transportation Research Part C: Emerging Technologies*, URL: <https://api.semanticscholar.org/CorpusID:202091779>.
- Liao, Weichen, Armin Seyfried, Jun Zhang, Maik Boltes, Xiaoping Zheng, and Ying Zhao (2014), « Experimental Study on Pedestrian Flow through Wide Bottleneck », *in: Transportation Research Procedia 2*, The Conference on Pedestrian and Evacuation Dynamics 2014 (PED 2014), 22-24 October 2014, Delft, The Netherlands, pp. 26–33, DOI: <https://doi.org/10.1016/j.trpro.2014.09.005>, URL: <https://www.sciencedirect.com/science/article/pii/S2352146514000416>.
- Liao, Weichen, Armel U. Kemloh Wagoum, and Nikolai W. F. Bode (2017), « Route choice in pedestrians: determinants for initial choices and revising decisions », *in: Journal of The Royal Society Interface* 14, URL: <https://api.semanticscholar.org/CorpusID:1081478>.
- Liddle, Jack, Armin Seyfried, Wolfram Klingsch, Tobias Rupprecht, Andreas Schadschneider, and Andreas Winkens (2009), « An experimental study of pedestrian congestions: influence of bottleneck width and length », *in: arXiv preprint arXiv:0911.4350*.
- Lin, Jing, Lijun Cao, and Nan Li (2020), « How the completeness of spatial knowledge influences the evacuation behavior of passengers in metro stations: A VR-based experimental study », *in: Automation in Construction* 113, p. 103136, URL: <https://api.semanticscholar.org/CorpusID:213493791>.
- Liu, Xiaodong, Weiguo Song, Libi Fu, and Zhiming Fang (2016a), « Experimental study of pedestrian inflow in a room with a separate entrance and exit », *in: Physica A: Statistical Mechanics and its Applications* 442, pp. 224–238.
- Liu, Xiaodong, Weiguo Song, Libi Fu, Wei Lv, and Zhiming Fang (2016b), « Typical features of pedestrian spatial distribution in the inflow process », *in: Physics Letters A* 380.17, pp. 1526–1534.
- Llobera, Joan, Bernhard Spanlang, Giulio Ruffini, and Mel Slater (2010), « Proxemics with multiple dynamic characters in an immersive virtual environment », *in: ACM Transactions on Applied Perception (TAP)* 8.1, pp. 1–12.

- 
- Loomis, Jack M. and Joshua M. Knapp (2003), « Visual perception of egocentric distance in real and virtual environments. », *in*: URL: <https://api.semanticscholar.org/CorpusID:60717474>.
- López, Axel, François Chaumette, Éric Marchand, and Julien Pettré (2019), « Character navigation in dynamic environments based on optical flow », *in*: *Computer Graphics Forum* 38, URL: <https://api.semanticscholar.org/CorpusID:181694047>.
- Lovreglio, Ruggiero, Dino Borri, Luigi dell’Olio, and Angel Ibeas (2014), « A discrete choice model based on random utilities for exit choice in emergency evacuations », *in*: *Safety Science* 62, pp. 418–426, URL: <https://api.semanticscholar.org/CorpusID:109855979>.
- Lovreglio, Ruggiero, Achille Fonzone, Luigi dell’Olio, and Dino Borri (2016), « A study of herding behaviour in exit choice during emergencies based on random utility theory. », *in*: *Safety Science* 82, pp. 421–431, URL: <https://api.semanticscholar.org/CorpusID:110964177>.
- Lovreglio, Ruggiero, Enrico Ronchi, and Daniel Nilsson (2015), « A model of the decision-making process during pre-evacuation », *in*: *Fire Safety Journal* 78, pp. 168–179, URL: <https://api.semanticscholar.org/CorpusID:109812359>.
- Ma, J., Wei Guo Song, S. M. Lo, and Zheng Fang (2013), « New insights into turbulent pedestrian movement pattern in crowd-quakes », *in*: *Journal of Statistical Mechanics: Theory and Experiment* 2013, URL: <https://api.semanticscholar.org/CorpusID:121348155>.
- Ma, Jian, Dongdong Shi, Tao Li, Xiaofei Li, Tengfei Xu, and Peng Lin (2020), « Experimental study of single-file pedestrian movement with height constraints », *in*: *Journal of Statistical Mechanics: Theory and Experiment* 2020, URL: <https://api.semanticscholar.org/CorpusID:225629115>.
- Moussaïd, Mehdi, Dirk Helbing, Simon Garnier, Anders Johansson, Maud Combe, and Guy Theraulaz (2009), « Experimental study of the behavioural mechanisms underlying self-organization in human crowds », *in*: *Proceedings of the Royal Society B: Biological Sciences* 276.1668, pp. 2755–2762.
- Moussaïd, Mehdi, Mubbasir Kapadia, Tyler Thrash, Robert W Sumner, Markus Gross, Dirk Helbing, and Christoph Hölscher (2016), « Crowd behaviour during high-stress evacuations in an immersive virtual environment », *in*: *Journal of The Royal Society Interface* 13.122, p. 20160414.

- 
- Moussaïd, Mehdi, Niriasca Perozo, Simon Garnier, Dirk Helbing, and Guy Theraulaz (2010), « The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics », *in: PLoS ONE* 5, URL: <https://api.semanticscholar.org/CorpusID:1216498>.
- Mullick, Pratik, Sylvain Fontaine, Cécile Appert-Rolland, Anne-Hélène Olivier, William H Warren, and Julien Pettré (2022), « Analysis of emergent patterns in crossing flows of pedestrians reveals an invariant of ‘stripe’formation in human data », *in: PLoS computational biology* 18.6, e1010210.
- Narang, Sahil, Andrew Best, and Dinesh Manocha (2017), « Interactive simulation of local interactions in dense crowds using elliptical agents », *in: Journal of Statistical Mechanics: Theory and Experiment* 2017, URL: <https://api.semanticscholar.org/CorpusID:40505345>.
- Natapov, Asya and Dafna Fisher-Gewirtzman (2016), « Visibility of urban activities and pedestrian routes: An experiment in a virtual environment », *in: Comput. Environ. Urban Syst.* 58, pp. 60–70, URL: <https://api.semanticscholar.org/CorpusID:7292492>.
- Navin, Francis P and Robert J Wheeler (1969), « Pedestrian flow characteristics », *in: Traffic Engineering, Inst Traffic Engr* 39.
- Nelson, Michael G., Alexandros Koiliias, Christos-Nikolaos E. Anagnostopoulos, and Christos Mousas (2022), « Effects of Rendering Styles of a Virtual Character on Avoidance Movement Behavior », *in: 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 594–599, URL: <https://api.semanticscholar.org/CorpusID:253463908>.
- Nelson, Michael G., Alexandros Koiliias, Sahana Gubbi, and Christos Mousas (2019), « Within a Virtual Crowd: Exploring Human Movement Behavior during Immersive Virtual Crowd Interaction », *in: Proceedings of the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, URL: <https://api.semanticscholar.org/CorpusID:208032694>.
- Nelson, Michael G., Alexandros Koiliias, Dominic Kao, and Christos Mousas (2023), « Effects of Speed of a Collocated Virtual Walker and Proximity Toward a Static Virtual Character on Avoidance Movement Behavior », *in: 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 930–939, URL: <https://api.semanticscholar.org/CorpusID:264045210>.

- 
- Nicolas, Alexandre, Sebastián Bouzat, and Marcelo N. Kuperman (2016), « Pedestrian flows through a narrow doorway: Effect of individual behaviours on the global flow and microscopic dynamics », *in: Transportation Research Part B-methodological* 99, pp. 30–43, URL: <https://api.semanticscholar.org/CorpusID:119191927>.
- Nicolas, Alexandre, Marcelo Kuperman, Santiago Ibañez, Sebastián Bouzat, and Cécile Appert-Rolland (2019), « Mechanical response of dense pedestrian crowds to the crossing of intruders », *in: Scientific reports* 9.1, p. 105.
- Nilsson, Daniel and Anders Johansson (2009), « Social influence during the initial phase of a fire evacuation—Analysis of evacuation experiments in a cinema theatre », *in: Fire Safety Journal* 44, pp. 71–79, URL: <https://api.semanticscholar.org/CorpusID:108509508>.
- Older, SJ (1968), « Movement of pedestrians on footways in shopping streets », *in: Traffic engineering & control* 10.4.
- Olivier, Anne-Hélène, Julien Bruneau, Gabriel Cirio, and Julien Pettré (2014), « A Virtual Reality Platform to Study Crowd Behaviors », *in: Transportation Research Procedia* 2, The Conference on Pedestrian and Evacuation Dynamics 2014 (PED 2014), 22-24 October 2014, Delft, The Netherlands, pp. 114–122, DOI: <https://doi.org/10.1016/j.trpro.2014.09.015>, URL: <https://www.sciencedirect.com/science/article/pii/S2352146514000519>.
- Ondřej, Jan, Julien Pettré, Anne-Hélène Olivier, and Stéphane Donikian (2010), « A synthetic-vision based steering approach for crowd simulation », *in: ACM Transactions on Graphics (TOG)* 29.4, pp. 1–9.
- Osimo, Sofia Adelaide, Rodrigo Pizarro, Bernhard Spanlang, and Mel Slater (2015), « Conversations between self and self as Sigmund Freud—A virtual body ownership paradigm for self counselling », *in: Scientific reports* 5.1, pp. 1–14.
- Paetzke, Sarah, Maik Boltes, and Armin Seyfried (2023), « Influence of Gender Composition in Pedestrian Single-File Experiments », *in: Applied Sciences*, URL: <https://api.semanticscholar.org/CorpusID:257079196>.
- Paris, Sébastien, Julien Pettré, and Stéphane Donikian (2007), « Pedestrian Reactive Navigation for Crowd Simulation: a Predictive Approach », *in: Computer Graphics Forum* 26, URL: <https://api.semanticscholar.org/CorpusID:12141470>.
- Parisi, Daniel R., Pablo Negri, and Luciana Bruno (2016), « Experimental characterization of collision avoidance in pedestrian dynamics. », *in: Physical review. E* 94 2-1, p. 022318, URL: <https://api.semanticscholar.org/CorpusID:11373044>.

- 
- Patotskaya, Yuliya, Ludovic Hoyet, Anne-Hélène Olivier, Julien Pettré, and Katja Zibrek (2023), « Avoiding virtual humans in a constrained environment: Exploration of novel behavioural measures », *in: Comput. Graph.* 110, pp. 162–172, URL: <https://api.semanticscholar.org/CorpusID:255881427>.
- Pettré, Julien, Jan Ondřej, Anne-Hélène Olivier, Armel Crétual, and Stéphane Donikian (2009), « Experiment-based modeling, simulation and validation of interactions between virtual walkers », *in: Symposium on Computer Animation*, URL: <https://api.semanticscholar.org/CorpusID:3280806>.
- Porzycki, Jakub, Natalia Schmidt-Polończyk, and Jarosław Wąs (2018), « Pedestrian behavior during evacuation from road tunnel in smoke condition—Empirical results », *in: PLoS ONE* 13, URL: <https://api.semanticscholar.org/CorpusID:52117225>.
- Raimbaud, Pierre, Alberto Jovane, Katja Zibrek, Claudio Pacchierotti, Marc Christie, Ludovic Hoyet, Julien Pettré, and Anne-Hélène Olivier (2022), « The Stare-in-the-Crowd Effect in Virtual Reality », *in: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 281–290, DOI: 10.1109/VR51125.2022.00047.
- (2023), « The Stare-in-the-Crowd Effect When Navigating a Crowd in Virtual Reality », *in: ACM Symposium on Applied Perception (SAP)*.
- Rempe, Davis, Zhengyi Luo, Xue Bin Peng, Ye Yuan, Kris Kitani, Karsten Kreis, Sanja Fidler, and Or Litany (2023), « Trace and Pace: Controllable Pedestrian Animation via Guided Trajectory Diffusion », *in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13756–13766, URL: <https://api.semanticscholar.org/CorpusID:257921535>.
- Ren, Jiaping, Wei Xiang, Yangxi Xiao, Ruigang Yang, Dinesh Manocha, and Xiaogang Jin (2018), « Heter-Sim: Heterogeneous Multi-Agent Systems Simulation by Interactive Data-Driven Optimization », *in: IEEE Transactions on Visualization and Computer Graphics* 27, pp. 1953–1966, URL: <https://api.semanticscholar.org/CorpusID:54438724>.
- Ren, Xiangxia, Jun Zhang, Weiguo Song, and Shuchao Cao (2019), « The fundamental diagrams of elderly pedestrian flow in straight corridors under different densities », *in: Journal of Statistical Mechanics: Theory and Experiment* 2019.2, p. 023403.
- Renner, Rebekka S, Boris M Velichkovsky, and Jens R Helmert (2013), « The perception of egocentric distances in virtual environments—a review », *in: ACM Computing Surveys (CSUR)* 46.2, pp. 1–40.

- 
- Rio, Kevin W, Gregory C Dachner, and William H Warren (2018), « Local interactions underlying collective motion in human crowds », *in: Proceedings of the Royal Society B: Biological Sciences* 285.1878, p. 20180611.
- Ríos, Alejandro, Daniel Bujosa Mateu, and Nuria Pelechano (2018), « Follower Behavior in a Virtual Environment », *in: URL: <https://api.semanticscholar.org/CorpusID:198988655>*.
- Ríos, Alejandro and Nuria Pelechano (2020), « Follower behavior under stress in immersive VR », *in: Virtual Reality* 24.4, pp. 683–694.
- Ronchi, Enrico, Karl Fridolf, Håkan Frantzich, Daniel Nilsson, Anders Lindgren Walter, and Henric Modig (2017), « A tunnel evacuation experiment on movement speed and exit choice in smoke », *in: Fire Safety Journal* 97, pp. 126–136, URL: <https://api.semanticscholar.org/CorpusID:113642060>.
- Salzmann, Tim, B. Ivanovic, Punarjay Chakravarty, and Marco Pavone (2020), « Trajec-tron++: Dynamically-Feasible Trajectory Forecasting with Heterogeneous Data », *in: European Conference on Computer Vision*, URL: <https://api.semanticscholar.org/CorpusID:214802528>.
- Sanz, Ferran Argelaguet, Anne-Helene Olivier, Gerd Bruder, Julien Pettré, and Anatole Lécuyer (2015), « Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments », *in: 2015 IEEE Virtual Reality (VR)*, pp. 75–80, DOI: 10.1109/VR.2015.7223327.
- Seyfried, A., B. Steffen, W. Klingsch, and M. Boltes (2005a), « The Fundamental Diagram of Pedestrian Movement Revisited », *in: Journal of statistical mechanics: theory and experiment*, Record converted from VDB: 12.11.2012, P10002, DOI: 10.1088/1742-5468/2005/10/P10002, URL: <https://juser.fz-juelich.de/record/48274>.
- Seyfried, Armin, Maik Boltes, Jens Kahler, Wolfram Klingsch, Andrea Portz, Tobias Rupprecht, Andreas Schadschneider, Bernhard Steffen, and Andreas Winkens (2010a), « Enhanced Empirical Data for the Fundamental Diagram and the Flow Through Bottlenecks », *in: Pedestrian and evacuation dynamics 2008*, ed. by W.W.F. Klingsch, A. Schadschneider, and M. Schreckenberg, Springer-Verlag Berlin, pp. 145–156, DOI: 10.1007/978-3-642-04504-2\_11.
- Seyfried, Armin, Maik Boltes, Jens Kähler, Wolfram Klingsch, Andrea Portz, Tobias Rupprecht, Andreas Schadschneider, Bernhard Steffen, and Andreas Winkens (2010b), « Enhanced empirical data for the fundamental diagram and the flow through bottle-necks », *in: Pedestrian and Evacuation Dynamics 2008*, pp. 145–156.

- 
- Seyfried, Armin, Oliver Passon, Bernhard Steffen, Maik Boltes, Tobias Rupprecht, and Wolfram Klingsch (2009), « New insights into pedestrian flow through bottlenecks », *in: Transportation Science* 43.3, pp. 395–406.
- Seyfried, Armin, Bernhard Steffen, Wolfram Klingsch, and Maik Boltes (2005b), « The Fundamental Diagram of Pedestrian Movement Revisited », *in: J. Stat. Mech.*, P10002.
- Shafique, Khurram, Mun Wai Lee, and Niels C. Haering (2008), « A rank constrained continuous formulation of multi-frame multi-target tracking problem », *in: 2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, URL: <https://api.semanticscholar.org/CorpusID:12783525>.
- Shah, Jiten, Gaurang Joshi, and Purnima Parida (2013), « Behavioral Characteristics of Pedestrian Flow on Stairway at Railway Station », *in: Procedia - Social and Behavioral Sciences* 104, pp. 688–697, URL: <https://api.semanticscholar.org/CorpusID:109961735>.
- Shahhoseini, Zahra, Majid Sarvi, Meead Saberi, and Milad Haghani (2017), « Pedestrian Crowd Dynamics Observed at Merging Sections: Impact of Designs on Movement Efficiency », *in: Transportation Research Record* 2622, pp. 48–57, URL: <https://api.semanticscholar.org/CorpusID:115638423>.
- Shi, Xiaomeng, Zhirui Ye, Nirajan Shiwakoti, and Offer Grembek (2018), « A State-of-the-Art Review on Empirical Data Collection for External Governed Pedestrians Complex Movement », *in: Journal of Advanced Transportation*.
- Shields, TJ and KE Boyce (2000), « A study of evacuation from large retail stores », *in: Fire Safety Journal* 35, pp. 25–49, URL: <https://api.semanticscholar.org/CorpusID:110872718>.
- Shiwakoti, Nirajan (2016), « Understanding differences in emergency escape and experimental pedestrian crowd egress through quantitative comparison », *in: International journal of disaster risk reduction* 20, pp. 129–137, URL: <https://api.semanticscholar.org/CorpusID:114421554>.
- Shiwakoti, Nirajan, Yan-cheng Gong, Xiaomeng Shi, and Zhirui Ye (2015), « Examining influence of merging architectural features on pedestrian crowd movement », *in: Safety Science* 75, pp. 15–22, URL: <https://api.semanticscholar.org/CorpusID:109146907>.
- Sieben, Anna, Jette Schumann, and Armin Seyfried (2017), « Collective phenomena in crowds—Where pedestrian dynamics need social psychology », *in: PLoS ONE* 12, URL: <https://api.semanticscholar.org/CorpusID:10473143>.



- 
- Singh, Shawn, Mubbasir Kapadia, Petros Faloutsos, and Glenn D. Reinman (2009), « SteerBench: a benchmark suite for evaluating steering behaviors », *in: Computer Animation and Virtual Worlds* 20, URL: <https://api.semanticscholar.org/CorpusID:16462382>.
- Singh, V.K., Bo Wu, and Ramakant Nevatia (2008), « Pedestrian Tracking by Associating Tracklets using Detection Residuals », *in: 2008 IEEE Workshop on Motion and video Computing*, pp. 1–8, URL: <https://api.semanticscholar.org/CorpusID:16234308>.
- Slater, Mel, Martin Usoh, Razia Geeas, and Anthony Steed (1995), « Creating animations using virtual reality ThatcherWorld: a case study », *in: Proceedings Computer Animation'95*, pp. 50–55, URL: <https://api.semanticscholar.org/CorpusID:31727548>.
- Sommer, Robert (1959), « Studies in personal space », *in: Sociometry* 22.3, pp. 247–260.
- Steffen, Bernhard and Armin Seyfried (2010), « Methods for measuring pedestrian density, flow, speed and direction with minimal scatter », *in: Physica A: Statistical mechanics and its applications* 389.9, pp. 1902–1910.
- Sun, Jinlu, Shouxiang Lu, Siuming Lo, Jian Ma, and Qimiao Xie (2018), « Moving characteristics of single file passengers considering the effect of ship trim and heeling », *in: Physica A: Statistical Mechanics and its Applications* 490, pp. 476–487.
- Sun, Zhihong, Jun Chen, Liang Chao, Weijian Ruan, and Mithun Mukherjee (2021), « A Survey of Multiple Pedestrian Tracking Based on Tracking-by-Detection Framework », *in: IEEE Transactions on Circuits and Systems for Video Technology* 31, pp. 1819–1833, URL: <https://api.semanticscholar.org/CorpusID:226691027>.
- Tanaboriboon, Yordphol and Jocelyn A Guyano (1991), « ANALYSIS OF PEDESTRIAN MOVEMENTS IN BANGKOK », *in: Transportation Research Record*, URL: <https://api.semanticscholar.org/CorpusID:108231795>.
- Tanaboriboon, Yordphol, Sim Siang Hwa, and Chin Hoong Chor (1986), « Pedestrian Characteristics Study in Singapore », *in: Journal of Transportation Engineering-asce* 112, pp. 229–235, URL: <https://api.semanticscholar.org/CorpusID:110732674>.
- Tang, Chieh-Hsin, Wu-Tai Wu, and Ching-Yuan Lin (2009), « Using virtual reality to determine how emergency signs facilitate way-finding. », *in: Applied ergonomics* 40 4, pp. 722–30, URL: <https://api.semanticscholar.org/CorpusID:205586003>.
- Tang, Siyu, Mykhaylo Andriluka, Bjoern Andres, and Bernt Schiele (2017), « Multiple People Tracking by Lifted Multicut and Person Re-identification », *in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3701–3710, URL: <https://api.semanticscholar.org/CorpusID:9615667>.

- 
- Tian, Wei, Weiguo Song, Jian Ma, Zhiming Fang, Armin Seyfried, and Jack Liddle (2012), « Experimental study of pedestrian behaviors in a corridor based on digital image processing », *in: Fire Safety Journal* 47, pp. 8–15, URL: <https://api.semanticscholar.org/CorpusID:109698746>.
- Toll, Wouter van, Fabien Grzeskowiak, Axel López Gandía, Javad Amirian, Florian Berton, Julien Bruneau, Beatriz Cabrero Daniel, Alberto Jovane, and Julien Pettré (2020), « Generalized microscopic crowd simulation using costs in velocity space », *in: Symposium on Interactive 3D Graphics and Games*, pp. 1–9.
- Toll, Wouter van and Julien Pettré (2021), « Algorithms for Microscopic Crowd Simulation: Advancements in the 2010s », *in: Computer Graphics Forum* 40, URL: <https://api.semanticscholar.org/CorpusID:235337856>.
- Trivedi, Hritik and Christos Mousas (2023), « Human-virtual crowd interaction: Towards understanding the effects of crowd avoidance proximity in an immersive virtual environment », *in: Computer Animation and Virtual Worlds*, e2169.
- Van Den Berg, Jur, Stephen J Guy, Ming Lin, and Dinesh Manocha (2011), « Reciprocal n-body collision avoidance », *in: Robotics Research: The 14th International Symposium ISRR*, Springer, pp. 3–19.
- Virkler, Mark R. and Sathish Elayadath (1994), « PEDESTRIAN SPEED-FLOW-DENSITY RELATIONSHIPS », *in: Transportation Research Record*, URL: <https://api.semanticscholar.org/CorpusID:88510769>.
- Wagoum, Armel Ulrich Kemloh, Antoine Tordeux, and Weichen Liao (2017), « Understanding human queuing behaviour at exits: an empirical study », *in: Royal Society Open Science* 4, URL: <https://api.semanticscholar.org/CorpusID:696820>.
- Wang, He, Jan Ondřej, and Carol O’Sullivan (2017), « Trending Paths: A New Semantic-Level Metric for Comparing Simulated and Real Crowd Data », *in: IEEE Transactions on Visualization and Computer Graphics* 23, pp. 1454–1464, URL: <https://api.semanticscholar.org/CorpusID:3022061>.
- Wang, Jia, Shunjiang Ni, Shifei Shen, and Shu-ying Li (2019), « Empirical study of crowd dynamic in public gathering places during a terrorist attack event », *in: Physica A: Statistical Mechanics and its Applications*, URL: <https://api.semanticscholar.org/CorpusID:126781340>.
- Warren, William H. (2006), « The dynamics of perception and action. », *in: Psychological review* 113 2, pp. 358–89, URL: <https://api.semanticscholar.org/CorpusID:1703806>.

- 
- Wąs, Jarosław, Bartłomiej Gudowski, and Paweł J Matuszyk (2006), « Social distances model of pedestrian dynamics », *in: International Conference on Cellular Automata*, Springer, pp. 492–501.
- Wolinski, David, Stephen J. Guy, Anne-Hélène Olivier, Ming C. Lin, Dinesh Manocha, and Julien Pettré (2014a), « Parameter estimation and comparative evaluation of crowd simulations », *in: Computer Graphics Forum* 33, URL: <https://api.semanticscholar.org/CorpusID:1226761>.
- Wolinski, David, S J. Guy, A-H Olivier, Ming Lin, Dinesh Manocha, and Julien Pettré (2014b), « Parameter estimation and comparative evaluation of crowd simulations », *in: Computer Graphics Forum*, vol. 33, 2, Wiley Online Library, pp. 303–312.
- Xu, Yuanlu, Lei Qin, Xiaobai Liu, Jianwen Xie, and Song-Chun Zhu (2017), « A Causal And-Or Graph Model for Visibility Fluent Reasoning in Tracking Interacting Objects », *in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2178–2187, URL: <https://api.semanticscholar.org/CorpusID:4437705>.
- Yamori, Katsuya (1998), « Going with the flow: Micro–macro dynamics in the macrobehavioral patterns of pedestrian crowds. », *in: Psychological review* 105.3, p. 530.
- Yang, Bo, Chang Huang, and Ramakant Nevatia (2011a), « Learning affinities and dependencies for multi-target tracking using a CRF model », *in: CVPR 2011*, pp. 1233–1240, URL: <https://api.semanticscholar.org/CorpusID:2633157>.
- Yang, Xiaolin, Zhongliang Wu, and Yingchun Li (2011b), « Difference between real-life escape panic and mimic exercises in simulated situation with implications to the statistical physics models of emergency evacuation: The 2008 Wenchuan earthquake », *in: Physica A-statistical Mechanics and Its Applications* 390, pp. 2375–2380, URL: <https://api.semanticscholar.org/CorpusID:121235909>.
- Zanlungo, Francesco, Tetsushi Ikeda, and Takayuki Kanda (2011), « Social force model with explicit collision prediction », *in: EPL (Europhysics Letters)* 93, p. 68005, URL: <https://api.semanticscholar.org/CorpusID:119611371>.
- Zeng, Guang, Andreas Schadschneider, Jun Zhang, Shibo Wei, Weiguo Song, and Rui Ba (2019), « Experimental study on the effect of background music on pedestrian movement at high density », *in: Physics Letters A* 383.10, pp. 1011–1018.
- Zeng, Weiliang, Peng Chen, Hideki Nakamura, and Miho Iryo-Asano (2014), « Application of social force model to pedestrian behavior analysis at signalized crosswalk », *in: Transportation Research Part C-emerging Technologies* 40, pp. 143–159, URL: <https://api.semanticscholar.org/CorpusID:110190038>.

- 
- Zhang, Jun, Wolfram Klingsch, Andreas Schadschneider, and Armin Seyfried (2011), « Transitions in pedestrian fundamental diagrams of straight corridors and T-junctions », *in: Journal of Statistical Mechanics: Theory and Experiment* 2011, P06004, URL: <https://api.semanticscholar.org/CorpusID:118409549>.
- Zhang, Jun and Armin Seyfried (2012), « Empirical Characteristics of Different Types of Pedestrian Streams », *in: Procedia Engineering* 62, pp. 655–662, URL: <https://api.semanticscholar.org/CorpusID:111167834>.
- Zhang, Li, Yuan Li, and Ramakant Nevatia (2008), « Global data association for multi-object tracking using network flows », *in: 2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, URL: <https://api.semanticscholar.org/CorpusID:32640>.
- Zhang, X. L., W G Weng, and H Y Yuan (2012), « Empirical study of crowd behavior during a real mass event », *in: Journal of Statistical Mechanics: Theory and Experiment* 2012, URL: <https://api.semanticscholar.org/CorpusID:122058572>.
- Zhang, X. L., Wenguo Weng, Hongyong Yuan, and Jing Chen (2013), « Empirical study of a unidirectional dense crowd during a real mass event », *in: Physica A-statistical Mechanics and Its Applications* 392, pp. 2781–2791, URL: <https://api.semanticscholar.org/CorpusID:120509016>.
- Zhao, Mingbi, Stephen John Turner, and Wentong Cai (2013), « A Data-Driven Crowd Simulation Model Based on Clustering and Classification », *in: 2013 IEEE/ACM 17th International Symposium on Distributed Simulation and Real Time Applications*, pp. 125–134, URL: <https://api.semanticscholar.org/CorpusID:5544537>.
- Zhao, Tianyang, Yifei Xu, Mathew Monfort, Wongun Choi, Chris Baker, Yibiao Zhao, Yizhou Wang, and Ying Nian Wu (2019), « Multi-Agent Tensor Fusion for Contextual Trajectory Prediction », *in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12118–12126, DOI: 10.1109/CVPR.2019.01240.
- Zhu, Kongjin and Qin Shi (2016), « Experimental Study on Choice Behavior of Pedestrians During Building Evacuation », *in: Procedia Engineering* 135, pp. 207–216, URL: <https://api.semanticscholar.org/CorpusID:112729605>.
- Zhu, Runhe, Jing Lin, Burçin Becerik-Gerber, and Nan Li (2020), « Influence of architectural visual access on emergency wayfinding: A cross-cultural study in China, United Kingdom and United States », *in: Fire Safety Journal* 113, p. 102963, URL: <https://api.semanticscholar.org/CorpusID:213123609>.

---

Ziemer, Verena, Armin Seyfried, and Andreas Schadschneider (2016), « Congestion Dynamics in Pedestrian Single-File Motion », *in: arXiv: Physics and Society*, pp. 89–96, URL: <https://api.semanticscholar.org/CorpusID:119233677>.





---

**Titre :** l'Homme-Foule : Vers des captures de mouvements de foules basées sur un seul utilisateur immergé en réalité virtuelle

**Mot clés :** Simulation de foule, interaction humaine, réalité virtuelle, capture de mouvement

**Résumé :** Les données de mouvements de foule sont essentielles pour comprendre et simuler leurs comportements. Ces données sont toutefois rares en raison des multiples difficultés liées à leur collecte. Récemment, la réalité virtuelle (RV) a commencé à être utilisée pour étudier le comportement individuel dans les foules, généralement en immergeant les utilisateurs dans des foules virtuelles et en capturant leur comportement. Dans cette thèse, nous proposons et évaluons une nouvelle approche utilisant la RV, qui lève les limites des expériences réelles pour l'acquisition de données de mouvements de foules. Nous appelons ce nouveau paradigme « One-Man-Crowd ». Nous proposons tout d'abord de capturer les mouvements de la foule avec un seul utilisateur. En enregistrant les trajectoires passées et les mouvements du corps

de l'utilisateur, et en les affichant sur des personnages virtuels, ceux-ci construisent progressivement le comportement global de la foule. Nous proposons ensuite un nouveau concept de foule contextuelle qui s'appuie sur l'utilisation de la simulation de foule pour atténuer les biais comportementaux des utilisateurs lors de la procédure de capture. Nous mettons en œuvre deux stratégies différentes, à savoir un processus Replace-Record-Replay (3R) et un processus Replace-Record-Replay-Responsive (4R). Nous évaluons et validons l'approche proposée en reproduisant et en comparant au total cinq expériences de foules réelles. Nos résultats suggèrent que le paradigme One-Man-Crowd est une approche prometteuse pour l'acquisition de données réalistes de mouvements de foule dans des environnements virtuels.

---

**Title:** The One-Man-Crowd: Towards Single-User Capture of Collective Motions using Virtual Reality

**Keywords:** Crowd Simulation, Human Interaction, Virtual Reality, Motion Capture

**Abstract:** Crowd motion data is fundamental for understanding and simulating realistic crowd behaviors. Such data is, however, scarce because of multiple challenges and difficulties involved in its gathering. Virtual Reality (VR) has been leveraged to study individual behavior in crowds, typically by immersing users into simulated virtual crowds and capturing their behavior. In this thesis, we propose and evaluate a novel VR-based approach, lifting the limitations of real-world experiments for the acquisition of crowd motion data. We refer to this approach as the One-Man-Crowd paradigm. In this thesis, we first propose to capture crowd motion with a single user. By recording the past trajectories and body move-

ments of the user, and displaying them on virtual characters, the users progressively build the overall crowd behavior by themselves. Then, we propose the new concept of contextual crowds that leverage crowd simulation to mitigate the users' behavioral bias during the capture procedure. We implement two different strategies, namely a Replace-Record-Replay (3R) process and a Replace-Record-Replay-Responsive (4R) process. We evaluate and validate the proposed approach by replicating and comparing with in total five real crowd experiments. Our results suggest that the One-Man-Crowd paradigm offers a promising approach for acquiring realistic crowd motion data in virtual environments.