

# Speed-Density Analysis in Pedestrian Single-File Experiments

Sarah Paetzke

IAS Series

Band / Volume 68

ISBN 978-3-95806-818-6







Forschungszentrum Jülich GmbH  
Institute for Advanced Simulation (IAS)  
Zivile Sicherheitsforschung (IAS-7)

# **Speed-Density Analysis in Pedestrian Single-File Experiments**

Sarah Paetzke

Schriften des Forschungszentrums Jülich  
IAS Series

Band / Volume 68

ISSN 1868-8489

ISBN 978-3-95806-818-6

Bibliografische Information der Deutschen Nationalbibliothek.  
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der  
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten  
sind im Internet über <http://dnb.d-nb.de> abrufbar.

Herausgeber  
und Vertrieb:      Forschungszentrum Jülich GmbH  
                         Zentralbibliothek, Verlag  
                         52425 Jülich  
                         Tel.: +49 2461 61-5368  
                         Fax: +49 2461 61-6103  
                         [zb-publikation@fz-juelich.de](mailto:zb-publikation@fz-juelich.de)  
                         [www.fz-juelich.de/zb](http://www.fz-juelich.de/zb)

Umschlaggestaltung:      Grafische Medien, Forschungszentrum Jülich GmbH

Druck:                      Grafische Medien, Forschungszentrum Jülich GmbH

Copyright:                Forschungszentrum Jülich 2025

Schriften des Forschungszentrums Jülich  
IAS Series, Band / Volume 68

D 468 (Diss. Wuppertal, Univ., 2025)

ISSN 1868-8489  
ISBN 978-3-95806-818-6

Vollständig frei verfügbar über das Publikationsportal des Forschungszentrums Jülich (JuSER)  
unter [www.fz-juelich.de/zb/openaccess](http://www.fz-juelich.de/zb/openaccess).



This is an Open Access publication distributed under the terms of the [Creative Commons Attribution License 4.0](https://creativecommons.org/licenses/by/4.0/),  
which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

---

# Acknowledgments

First of all, I would like to thank Prof. Dr. Armin Seyfried and Dr. Maik Boltes for giving me the opportunity to work under their supervision and with their support in the field of pedestrian dynamics. I have benefited greatly from their knowledge of fundamental diagrams, stepping behavior, data analysis, and so on, and I enjoyed the very pleasant working atmosphere during our meetings. Furthermore, I am grateful for being a part of the Pedestrian Dynamics—Empiricism research group.

In addition, I would like to thank Dr. Mohcine Chraibi, Prof. Dr. Anna Sieben, and Dr. Mira Beermann for their support in conducting further single-file experiments at the Mitsubishi Electric Halle in Düsseldorf, Germany.

I also wish to thank all my colleagues for an excellent time in Jülich and Wuppertal.

Finally, I thank my family for their continuous support and motivation.

---

# Abstract

In recent years, many studies have been published on how individual characteristics of pedestrians affect the fundamental diagram. These studies compared cumulative data from individuals within groups that were homogeneous in one characteristic, such as age, but heterogeneous in other factors, such as gender. The question arises regarding which factors influence the fundamental diagram in single-file experiments. In addition, some research suggests that group composition, particularly gender, may impact the results. Therefore, three different single-file experiments are considered in this study.

The first step in data analysis is considering the speed-density relationship between homogeneous and heterogeneous group compositions regarding gender. Second, individual fundamental diagrams are introduced and analyzed to investigate the effects of both known and unknown human factors. In addition, the speed-density model that best describes the speed of the individuals is studied.

For a single-file experiment conducted in Germany with groups of participants that are either homogeneous or heterogeneous in terms of gender, a Tukey HSD test is carried out. This is done to determine any differences between these groups in the average speed over various density ranges. A comparison of different group compositions shows that any effect of gender is only observed, if at all, within a small range of densities. For a cultural comparison, the previous experiment is compared with a single-file experiment conducted in Palestine with homogeneous and heterogeneous gender compositions. There are no significant cultural differences.

Two single-file experiments in Germany are considered to investigate differences in minimum distances and reaction times between individuals in different densities and to analyze the influence of the gender of neighboring pedestrians. It is also studied how human factors such as height, age, and gender and unknown individual effects such as motivation or attention affect individual speed. Regression analysis is performed for this. One experiment is a school experiment, and the results based on

a simple linear regression analysis show that the differences in minimum distances and reaction times are not very significant and are less pronounced. Using a multiple linear regression analysis, the human factors analysis suggested that age could be neglected as there is a strong correlation between the students' age and height. In addition, the study highlights that headway has the most significant influence on speed. At the same time, gender plays a minor role, and other non-measurable individual characteristics have a more significant impact than height. The results of the other experiment with different gender compositions in Germany show no correlation between the genders of neighboring pedestrians in minimum distances and reaction times. In the human factor analysis, the model stays consistent compared to the other experiment in Germany when additional factors are included.

---

# Zusammenfassung

In den letzten Jahren gab es zahlreiche Veröffentlichungen, die den Einfluss individueller Merkmale von Fußgängern auf das Fundamentaldiagramm untersucht haben. Dabei wurden Daten von Personen verglichen, die in einem Merkmal, wie beispielsweise dem Alter, homogen, in anderen Merkmalen, wie beispielsweise dem Geschlecht, jedoch heterogen sind. Dadurch ist unklar, welche Faktoren in single-file Experimenten einen Einfluss auf das Fundamentaldiagramm haben. Außerdem deuten Studien darauf hin, dass die Gruppenzusammensetzung, insbesondere hinsichtlich des Geschlechts, einen Einfluss auf die Ergebnisse haben kann. Für die Analysen werden in dieser Arbeit insgesamt drei verschiedene single-file Experimente herangezogen.

Zuerst wird das Geschwindigkeits-Dichte Verhältnis zwischen geschlechtshomogenen und geschlechtsheterogenen Gruppen betrachtet. Um die Auswirkungen sowohl bekannter als auch unbekannter menschlicher Faktoren zu untersuchen, werden individuelle Fundamentaldiagramme eingeführt und analysiert. Darüber hinaus wird untersucht, welches Geschwindigkeits-Dichte Modell die Geschwindigkeit von Personen am Besten beschreibt.

Für ein single-file Experiment, welches in Deutschland durchgeführt wurde und bei dem die Gruppenzusammensetzungen homogen oder heterogen bezüglich des Geschlechts sind, wird ein Tukey HSD-Test durchgeführt. Dabei soll festgestellt werden, ob es Unterschiede zwischen diesen Gruppen, in der mittleren Geschwindigkeit, in verschiedenen Dichtebereichen, gibt. Ein Vergleich der verschiedenen Gruppenzusammensetzungen zeigt, dass ein Effekt des Geschlechts lediglich in einem kleinen Dichtebereich zu beobachten ist. Hinsichtlich eines kulturellen Vergleichs wird das zuvor genannte Experiment mit einem single-file Experiment, mit geschlechtshomogenen und geschlechtsheterogenen Gruppenzusammensetzungen, verglichen, welches in Palestina durchgeführt wurde. Hier gibt es keine signifikanten kulturellen Unterschiede.

Zwei single-file Experimente, die in Deutschland durchgeführt wurden, werden

herangezogen, um Unterschiede in minimalen Abständen und in den Reaktionszeiten zwischen Individuen, in verschiedenen Dichten, zu untersuchen. Zudem wird betrachtet, ob der Abstand zwischen Personen, bei hohen Dichten, durch das Geschlecht benachbarter Fußgänger beeinflusst wird. Des Weiteren wird berücksichtigt, wie menschliche Eigenschaften wie Körpergröße, Alter und Geschlecht sowie unbekannte individuelle Effekte wie Motivation oder Aufmerksamkeit die individuelle Geschwindigkeit beeinflussen. Dazu werden Regressionsanalysen durchgeführt. Bei einem Experiment handelt es sich um ein Schulexperiment. Die Ergebnisse der einfachen Regressionsanalyse zeigen, dass die Unterschiede bei minimalen Abständen und Reaktionszeiten weniger stark ausgeprägt sind. Aus der Analyse der menschlichen Faktoren, mittels multilinearer Regressionsanalyse, geht hervor, dass das Alter vernachlässigt werden kann, da eine starke Korrelation zwischen dem Alter und der Körpergröße der Schüler besteht. Außerdem hat die Entfernung den größten Einfluss auf die Geschwindigkeit. Das Geschlecht spielt dagegen nur eine untergeordnete Rolle. Andere nicht messbare individuelle Merkmale haben einen größeren Einfluss als die Körpergröße. Die Ergebnisse des anderen Experiments mit verschiedenen Gruppenzusammensetzungen hinsichtlich des Geschlechts in Deutschland zeigen, dass es keine Korrelation zwischen dem Geschlecht benachbarter Fußgänger hinsichtlich minimalen Abständen und Reaktionszeiten gibt. Zudem bleibt das Modell konsistent, wenn zusätzliche Faktoren bei der Analyse der menschlichen Faktoren berücksichtigt werden.



---

# Contents

<b>Acknowledgments</b>	<b>I</b>
<b>Abstract</b>	<b>II</b>
<b>Zusammenfassung</b>	<b>IV</b>
<b>List of Figures</b>	<b>VII</b>
<b>List of Tables</b>	<b>X</b>
<b>List of Listings</b>	<b>XI</b>
<b>List of Notations</b>	<b>XII</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 State of Research . . . . .	2
1.3 Objectives and Approach . . . . .	7
1.4 Thesis Outline . . . . .	10
<b>2 Experiments</b>	<b>11</b>
2.1 Single-File Experiment at GBS in Germany . . . . .	12
2.1.1 Experimental Setup . . . . .	12
2.1.2 Data Processing . . . . .	13
2.2 Single-File Experiment at MEH in Germany . . . . .	14
2.2.1 Data Preparation . . . . .	15
2.2.2 Experimental Setup . . . . .	16
2.3 Single-File Experiment at AAU in Palestine . . . . .	19
2.3.1 Experimental Setup . . . . .	19
2.4 Data Selection . . . . .	20

<b>3</b>	<b>Methodology</b>	<b>24</b>
3.1	Measurement Methods . . . . .	24
3.1.1	Headway, Voronoi Distance, Density and Individual Speed . .	24
3.1.2	Stopping Phase . . . . .	25
3.2	Statistical Methods Related to Single-File Experiments . . . . .	27
3.2.1	Individuals in Simple Linear Regression Models . . . . .	27
3.2.2	Human Factors in Multiple Linear Regression Models . . . . .	29
3.2.3	Model Comparison with Mixed Models . . . . .	30
3.2.4	Tukey HSD Test and Gender Group Compositions . . . . .	32
<b>4</b>	<b>Data Analysis and Results</b>	<b>34</b>
4.1	Gender Compositions in Fundamental Diagrams . . . . .	34
4.1.1	Gender Composition in Germany . . . . .	34
4.1.2	Gender Compositions and Cultures . . . . .	40
4.2	Individual Speed-Distance Functions . . . . .	43
4.2.1	Differences by Gender and Age . . . . .	49
4.2.2	Differences by Gender of Neighboring Pedestrians . . . . .	53
4.3	Speed-Distance Relation, Human Factors, and Individuality . . . . .	57
4.3.1	Human Factors and Individuality . . . . .	57
4.3.2	Human Factors Extension . . . . .	63
4.4	Speed-Distance Relation with Fixed and Random Factors . . . . .	68
4.5	Gender Composition Related to Stop-And-Go Waves . . . . .	69
4.5.1	Velocity Distribution for Gender Compositon . . . . .	69
4.5.2	Stop-And-Go Waves in Time-Space Diagrams . . . . .	71
4.5.3	Phase Transformation between Stop-And-Go States . . . . .	73
<b>5</b>	<b>Discussion and Outlook</b>	<b>76</b>
5.1	Conclusions . . . . .	76
5.2	Outlook . . . . .	79
	<b>Appendix A</b>	<b>89</b>

---

# List of Figures

1.1	Trend of the fundamental diagram of recent single-file studies. . . . .	3
1.2	Possible course of velocity distributions for different densities. . . . .	6
1.3	Fundamental diagrams for groups homogeneous in only one human factor vs. individual fundamental diagrams. . . . .	8
2.1	Single-file experiment at GBS in Wuppertal, Germany. . . . .	12
2.2	Trajectories in PeTrack. . . . .	15
2.3	Single-file experiment conducted at the Mitsubishi Electric Halle (MEH) in Düsseldorf, Germany. . . . .	16
2.4	Single-file experiment conducted at the Arab American University (AAU) in Palestine. . . . .	19
2.5	Illustration of the steady-state in a specific run. . . . .	22
3.1	Walking position over time. . . . .	26
3.2	Representation of data in hierarchical levels. . . . .	31
4.1	Fundamental diagram of density vs. velocity for different groups with binned data. . . . .	35
4.2	Fundamental diagram of density vs. velocity for different groups with binned data that represent only the mean values of the velocity. . . .	36
4.3	Boxplots for the velocity for different group compositions in Germany. . . .	39
4.4	Relation of density and velocity for the experiments in Germany and Palestine by binned data of the mean values of the velocity. . . . .	41
4.5	Similarity between different group compositions regarding gender in Germany and Palestine. . . . .	43
4.6	Boxplots for the velocity for different group compositions in Germany and Palestine in a specific interval. . . . .	44
4.7	Headway vs. individual speed diagrams for different main IDs. . . . .	45
4.8	Comparison of the regression line at GBS and MEH. . . . .	47

4.9	The assumption of homoscedasticity. . . . .	48
4.10	Minimum distances for younger, older, male and female students from the experiment at GBS. . . . .	49
4.11	Distribution of $\hat{\beta}_1$ . . . . .	50
4.12	Diagrams of headway vs. individual speed with a regression line. . . .	51
4.13	Distribution of the correlation coefficient between headway vs. indi- vidual speed. . . . .	52
4.14	Boxplots based on a Tukey test for $(d_V)_{i,min}$ and for $\hat{\beta}_{1_i}$ for different group compositions. . . . .	56
4.15	Distribution of the regression coefficients $\hat{\beta}_{3_i}$ . . . . .	61
4.16	ANOVA: Effects on the individual speed. . . . .	62
4.17	Correlation plot for different factors. . . . .	64
4.18	Effects on the speed based on the ANOVA table for low and high density. . . . .	67
4.19	Velocity distributions for different global densities: Female and gender random order. . . . .	70
4.20	Velocity distributions for all gender group compositions (female, male, gender random, and gender alternating) for the highest density. . . .	71
4.21	Time-space diagrams: Female . . . . .	72
4.22	Relative frequency of individuals in the stop. . . . .	73
A.1	Fundamental diagrams: Velocity-density and flow-density based on recent studies. . . . .	89
A.2	Distance vs. speed relationship for several main IDs at GBS. . . . .	90
A.3	Distance vs. speed relationship for several main IDs at GBS. . . . .	91
A.4	Distance vs. speed relationship for several main IDs at GBS. . . . .	92
A.5	Distance vs. speed relationship for several main IDs at GBS. . . . .	93
A.6	Distance vs. speed relationship for several main IDs at MEH. . . . .	94
A.7	Distance vs. speed relationship for several main IDs at MEH. . . . .	95
A.8	Distance vs. speed relationship for several main IDs at MEH. . . . .	96
A.9	Velocity distributions for different global densities: Male and gender alternating. . . . .	97
A.10	Female velocity distributions in detail. . . . .	98
A.11	Time-space diagrams: Male . . . . .	99
A.12	Time-space diagrams: Random order . . . . .	100
A.13	Time-space diagrams: Gender alternating order . . . . .	101

A.14	The mean duration of a person's stay in the stop-and-go state. . . . .	102
A.15	The number of pedestrians in a stop-and-go state over time. . . . .	103

---

# List of Tables

2.1	A detailed overview of the number of different runs for each group composition. . . . .	17
2.2	A detailed overview of the mean values and the standard errors for age, height, weight, and shoulder width for the different group compositions.	18
2.3	A detailed overview of the experimental runs for the different groups of students at GBS in Germany. . . . .	21
4.1	Mean values and standard deviation for the velocity in seven different density intervals for different group compositions in Germany and Palestine. . . . .	37
4.2	Comparison of the differences in the experimental design. . . . .	40
4.3	An overview of the mean values and standard error for $d_{min_i}$ and $\hat{\beta}_{1_i}$ for the different group compositions. . . . .	53
4.4	Scatter of the residuals for model III. . . . .	61
4.5	Mean values and standard deviations for the stop frequency. . . . .	74
A.1	Detailed overview of experimental runs for different gender groups at MEH in Germany. . . . .	104

---

# Listings

4.1	Output table: Simple linear regression for all individuals at GBS in Germany. . . . .	46
4.2	Results of the Tukey HSD test in R for the values $d_{min_i}$ and $\hat{\beta}_{1_i}$ at the MEH in Germany. The last column presents the p-value. . . . .	54
4.3	Table of the Analysis of Variance for modell III at GBS. . . . .	62
4.4	Table of the Analysis of Variance for modell II at MEH. . . . .	66
4.5	Table of the Analysis of Variance for modell II at MEH for low density.	66
4.6	Table of the Analysis of Variance for modell II at MEH for high density.	67

---

# List of Notations

The next list describes abbreviations and notations that are used in this thesis across all chapters:

$\beta$	Regression coefficient
$\Delta t$	Difference between two time points
$\delta t$	Individual-specific time steps
$\epsilon$	Random experimental error
$\lambda$	Function which describes $\hat{x}$ , the distance on the central line from the coordinate origin
$\overline{\delta t}$	Mean value of $\delta t$
$\rho$	Density
$\rho_i(t) = \frac{1}{d_{V_i}}$	Density based on the Voronoi tessellation
$\rho_i(t) = \frac{1}{h_i}$	Density based on the headway
$\rho_{gl}$	Global density
$\hat{\beta}_1$	Stimulus response mechanism (reaction time and the ability to accelerate and brake)
$\hat{\beta}$	Estimated value for the regression coefficient
$\hat{v}$	Estimated speed
<i>Alloence</i>	All other unknown individual influences, e.g. motivation, attention, or excitement
$d_{min}$	Minimum headway for each individual $i$



$d_{V_i}(t)$	Voronoi distance
$h$	Headway
$i$	Individual
$j$	Number of the independent variables
$J(\rho)$	Capacity of a system: Flow in dependence of the density
$J(v)$	Capacity of a system: Flow in dependence of the velocity
$l_{cl}$	Length of the central line
$l_m$	Length of the measurement area
$l_{sl}$	Length of the straight sections
$N$	Number of persons
$n$	Number of all observations for all individuals
$n_i$	Number of observations for each individual $i$
$r_0$	The radius from the middle of the oval to the central line
$t$	Time
$v(\rho)$	Velocity in dependence of the density
$v_i(t)$	Individual speed
$w$	Width of the oval path
$x_i(t)$	Position of one individual $i$ at time $t$
AAU	Arab American University
AIC	Akaike's Information Criterion
ANOVA	Analysis of Variance
f	Female
GBS	Gymnasium Bayreuther Straße

$I_{max}$	Time at which the average speed is reached for the last time
$I_{min}$	Time at which the average speed is reached for the first time
JPSreport	Tool for the analysis of pedestrian trajectories
m	Male
MEH	Mitsubishi Electric Halle
$r_{i,\tau}$	Autocorrelation function (acf)

# Chapter 1

## Introduction

### 1.1 Motivation

In pedestrian dynamics, density, velocity, and flow describe pedestrian streams quantitatively. A fundamental diagram is the relationship between these terms obtained from empirical measurements. These diagrams illustrate the various states of traffic in the field of pedestrian dynamics as well as car traffic, including free-flow, bounded traffic, maximum flow, and congested traffic [1–3]. The maximum possible flow is of significant importance for safety in a crowd. Furthermore, level of service (LOS) concepts are employed to assess pedestrian comfort at different densities [4–6]. Fundamental diagrams vary [1, 4, 6–9] and are used to describe the capacity of various pedestrian facilities such as corridors [10–14], crossings [3, 15], stairs [7, 16] as well as stream types like single-file experiments [17–19]. Safety engineering and civil engineering applications include planning public spaces such as train stations or organizing events. The objective is to prevent congestion and crowding when large crowds of people are present. Fundamental diagrams need further research to achieve more accurate congestion predictions and a deeper understanding of the congested area. Of particular interest is when congestion occurs, when stop-and-go waves [20–23] or lane formation [11, 24] appear, and under what conditions traffic comes to a complete standstill. To further investigate the significance of the influences on speed, it is necessary to look at the fundamental diagrams not only at a macro- or microscopic level [16, 17, 19, 20, 25–32], but also to consider the relevance of individual factors. For this purpose, this thesis aims to develop individual fundamental diagrams of pedestrian dynamics. In addition, by analyzing the data in greater detail, this method should assist in a more comprehensive understanding of the factors that

affect speed. This work focuses on the simplest type of experiment conducted under laboratory conditions, namely the movement of pedestrians in a single-file experiment (see Figure 2.1), which can be observed in queuing systems. This approach minimizes the influence of additional factors, such as those resulting from uni- and bi-directional flow in straight corridors or by multi-dimensional pedestrian streams in crossings. Concerning single-file experiments, several variations in the experimental setup already affect the course of the fundamental diagrams. It is challenging to analyze and quantify the effects of multiple factors, and it needs to be evident which factors are relevant to the fundamental diagram in single-file experiments.

## 1.2 State of Research

The following section reviews the current state of empirical single-file research that has been conducted and analyzed in recent years [17–20, 22, 25, 26, 28–33]. In general, the trend of the resulting fundamental diagrams is similar.

Starting with the fundamental diagrams found in both pedestrian and vehicle textbooks [1, 4–6] they typically show two regions separated by the maximum flow. The region below the capacity is the free flow regime, while the area above the capacity is the congested regime. Recent single-file studies and stream types like uni-, bi-, and multidirectional streams suggest a more complex structure with three or four regimes: free flow, bounded flow, congested flow, and stop-and-go congested. Depending on the duration of the experiments, the different densities that are considered, and the measurement methods used, the number of regimes in the diagram may vary. Importantly, single-file experiments represent the simplest experimental setup and simultaneously extreme cases where participants cannot pass each other, which can result in pedestrian collisions. In addition, it is essential to consider the stationary area where the speeds do not change over time. Therefore, the focus is on the pedestrians already in motion and less on those starting to walk or standing still. Figure 1.1 shows the macroscopic average trend of the fundamental diagram observed in recent single-file studies [17, 19, 25, 34–36] (see Figure A.1 in Appendix 5.2). The figure on the left shows the flow-density relationship, and the right shows the speed-density relationship. Two kinks can be seen in the figure in **a**) and **b**), indicating the presence of three different regimes. First, pedestrians walk behind each other at their desired speed. When the density of pedestrians exceeds a certain threshold  $\rho > \rho_c$ , the individual pedestrians slow down. The change in

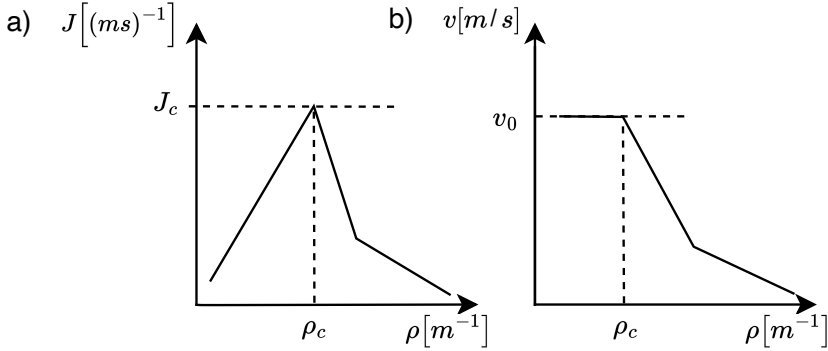


Figure 1.1: The trend of the fundamental diagram of recent single-file studies is shown. In **a)**, the trend of the flow-density relation is shown, and in **b)** of the velocity-density relation.

speed may be influenced by the presence of slower individuals at close distances. Second, as the density increases and the distance between individuals decreases, the group influence on speed increases. The second change in speed seems to occur in high-density areas for similar reasons. Some individuals appear to stop earlier and maintain a greater distance from those in front of them, while others continue to move closer.

Upon closer examination, variations in the overall pattern and magnitude can be observed. These differences can be attributed to various factors related to the pedestrians themselves, such as age, gender, height, cultural background, and human movement characteristics such as step length and frequency [17, 22, 25, 26, 28, 30, 32, 35, 37–45], as well as external factors such as visibility limitations [33], different height adjustments due to smoke [46], temperature [47], sidewalk quality [47, 48], background music [31, 34, 49] and the presence of luggage or trolleys [27]. Variations in experimental setups and measurement methods for speed and density can also contribute to discrepancies in the fundamental diagrams. It is important to note that the measurement areas used in these studies may vary, and the conversion of two-dimensional trajectories to one dimension is only sometimes clearly explained. This list is only exemplary and claims to be incomplete.

Due to the variety of factors that influence pedestrian behavior, and in part because the measurement methods or experimental scenarios in single-file studies vary, it can be challenging to compare the results of different experiments. For example, one study may focus on low-density scenarios only, while another may include data from

high-density situations. In addition, even when considering a homogeneous group concerning one factor, such as age or gender, differences in fundamental diagrams may still be observed. These differences highlight the complexity of pedestrian behavior and the importance of considering multiple factors when analyzing the fundamental diagrams. Furthermore, there are differences between the diagrams even if the same group is considered in terms of the same human factor, like age or gender.

When examining single-file studies considering age as a human factor, apparent differences will be explained below. Cao [17] has conducted a study in which the speeds of three groups are compared: a younger group, an older group, and a mixed group. At lower densities up to  $\rho = 1.5 \text{ m}^{-1}$ , the younger group has higher speeds than the mixed or older group. However, as the densities approached the stopping density, the speed of the mixed group was lower than the speed of the younger group. No data are available for the older group. Moreover, the fundamental diagram for the mixed group has a more complex structure because the diagram that illustrates the relationship between headway and speed in the mixed group shows three regimes: the free, weakly constrained, and strongly constrained regimes.

In contrast, only two regimes are observed in the younger and older groups. After scaling, only minor differences existed between the younger and older groups. However, in general, the diagrams for these three groups could not be transformed into the diagrams for the other groups by scaling the variables to dimensionless quantities, e.g., by dividing the velocity with the free velocity of different pedestrian groups, such as those of varying ages. The research of Ren et al. [25] supports these findings. The elderly group moves slower than the young adult group, and there are no significant differences between the mixed group and the elderly group. In addition, when a group of elderly Chinese is compared with French students [2] in a distance vs. speed diagram, three regimes emerge. Ren also found that pedestrian behavior is influenced by factors such as age, group diversity, and familiarity. Another study by Zhang et al. [50] compared two groups of middle-aged individuals based on income level. Although the fundamental diagrams for these groups are different, they showed similar trends. The group with higher income and more female participants tended to be less active, preferring to keep their distance or keep pace with others. This group was more homogeneous, resulting in more congestion and stop-and-go patterns. On the other hand, the lower-income group with a balanced gender ratio was more active, resulting in a higher flow rate. Regarding age, Subaih et al. [29] compared experiments in Palestine and China with groups of different genders at high densities. They found that older Chinese pedestrians walked at a similar speed to young

Palestinians, while younger Chinese walked faster than the Palestinians. Another study by Ziemer [22] examined age differences by analyzing fifth- and eleventh-grade students in single-file school experiments. Despite significant differences in height between these groups, age did not affect the fundamental diagram. In addition, the fact that the age groups were heterogeneous did not affect the diagrams.

These six studies [2, 17, 22, 25, 29, 50], suggest that age may or may not affect fundamental diagrams, depending on the specific characteristics of the group, e.g., the age. The discussion highlights that factors such as culture, income, gender, and group composition may play a role in addition to age. Ren [25] emphasizes the importance of considering age and diversity within a group regarding gender or culture. Zhang’s [50] research raises the question of whether the observed differences are influenced by income level or gender ratio within the group. In Subaih et al. [29], it remains to be seen whether age, culture, or the mix of genders in the group has the most significant impact on the fundamental diagram.

Studies that have focused on the impact of gender as a human factor in single-file pedestrian dynamics have shown various results, too. Subaih et al. [35] have demonstrated that at densities higher than  $1.0 \text{ m}^{-1}$ , homogeneous gender composition tended to result in higher speeds compared to groups with alternating gender order. However, when considering data from different cultures and age groups, it becomes clear that additional factors must be considered. Another study [37], using data from experiments conducted by Subaih et al. and introduced in [35], found that the headway between pedestrians to the front and the back is crucial. This suggests that gender-based arrangements influence the distances between individuals and should be considered in models of speed-density relationships. While these results highlight the critical role of gender, Paetzke et al. [51] concluded that gender may not have a significant effect based on multiple linear regression analysis of experiments with heterogeneous group compositions.

The discussion above presents conflicting perspectives on how human factors influence the fundamental diagrams of pedestrian dynamics. Although conducted in a controlled laboratory setting, a methodological challenge arises from the potential heterogeneity within groups, even when they appear homogeneous in one aspect. While speed and density measurements are taken for individuals in these studies mentioned above, comparing fundamental diagrams relies on cumulative data from all individuals within each group.

Regarding the stop-and-go congestion regime in fundamental diagrams, a common phenomenon in pedestrian dynamics is stop-and-go waves, where groups of pedestrians

switch between a jammed and a moving phase. In single-file experiments, these waves are visible at a global density of around  $\rho_{gl} = 2 \pm 0.15 \text{ m}^{-1}$  [20, 22, 25, 49, 52]. Recent research has focused on understanding the mechanisms behind stop-and-go waves. This means these studies have analyzed the state of motion and the stopped state. The stopped state is defined as a value for  $v_i(t)$  less than a value within the interval  $[0.05, 0.15] \text{ m/s}$  [20, 22, 25, 49, 52]. In [22], for example, a velocity is selected for which it is assumed that a person is in a stopped state. The value is not equal to 0 because even if a pedestrian stops, the head or body continues to move. It is proposed that the value be chosen as 0.1 m/s or 0.2 m/s, based on the free speed of 1.34 m/s multiplied by the ratio of the limit speed in vehicle traffic (0.08 m/s or 0.15 m/s) [22]. For a first visualization of the stopped state, different velocities  $v_i(t)$  values for each pedestrian in a specific density are presented in space-time diagrams in various colors. They showed that the frequency of stop-and-go waves increases with increasing density. This relationship is also analyzed in velocity distributions based on the number and height of peaks during stop-and-go waves. In general, two peaks appear around a global density of  $2 \text{ m}^{-1}$ , the height of the first peak increasing with increasing density and the second peak shifting to the left at higher densities. This phenomenon is illustrated in Figure 1.2.

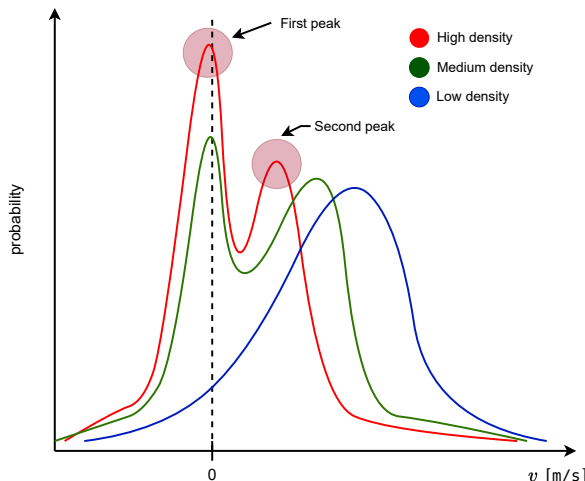


Figure 1.2: The possible course of velocity distributions for different densities. The distributions show how the development of peaks may vary from one peak at low density to two at high density. The peaks may continue approaching each other over time, with the first peak becoming more pronounced.



The five previously mentioned studies provide analysis methods for stop-and-go waves, such as the proportion of pedestrians in the stop, the duration of a person's stay in the stop, the duration of stopping and going, limit speeds for the stop state, or peak analysis in velocity distributions. However, it is not clear whether there are differences in stop-and-go waves at high densities for homogeneous and heterogeneous group compositions in terms of gender.

### 1.3 Objectives and Approach

This thesis analyzes human factors in speed-density relationships based on different statistical methods. Therefore, three different single-file experiments are considered. Two experiments were conducted in Germany. One was conducted at the school Gymnasium Bayreuther Straße (GBS), in Wuppertal, in 2014 [53] and the second at the Mitsubishi Electric Halle (MEH), in Düsseldorf, in 2021 [54]. Another single-file experiment was conducted at the Arab American University (AAU) in Palestine in 2018 [32, 55].

The relationship between speed-density ratio and gender composition in different group compositions, such as homogeneous and heterogeneous groups, e.g., female, male, gender random order, and gender alternating, is analyzed. This is done for the single-file experiment at MEH in Germany. Therefore, a visual comparison is made, and then individual density intervals are analyzed for differences in speed using the Tukey HSD test. Later, this comparison between homogeneous and heterogeneous groups is also considered between two different cultures – Germany and Palestine – as these results in Germany are compared with existing results of a single-file experiment in Palestine, where different group compositions regarding gender are also analyzed in fundamental diagrams. The reason for this is that in a recent study, in the experiments in Palestine, the speed of the homogeneous groups was higher than that of the heterogeneous groups in a specific density range. The cultural comparison will verify this.

Then, this thesis introduces a new method to quantify the impact of individual characteristics of the pedestrian on the fundamental diagram. This is done using regression analysis, starting with simple linear regression and ending with mixed models. In addition, the focus is on how the regression analysis methodology changes the evaluation. The inconsistent picture of how human factors affect the fundamental diagram of pedestrian dynamics is considered. Even if recent studies are conducted

## 1. Introduction

---

under well-controlled laboratory conditions, the main methodological issue is that even if a group is homogeneous for one factor, it may be heterogeneous concerning other factors. This method difference is illustrated in Figure 1.3. At the top, the

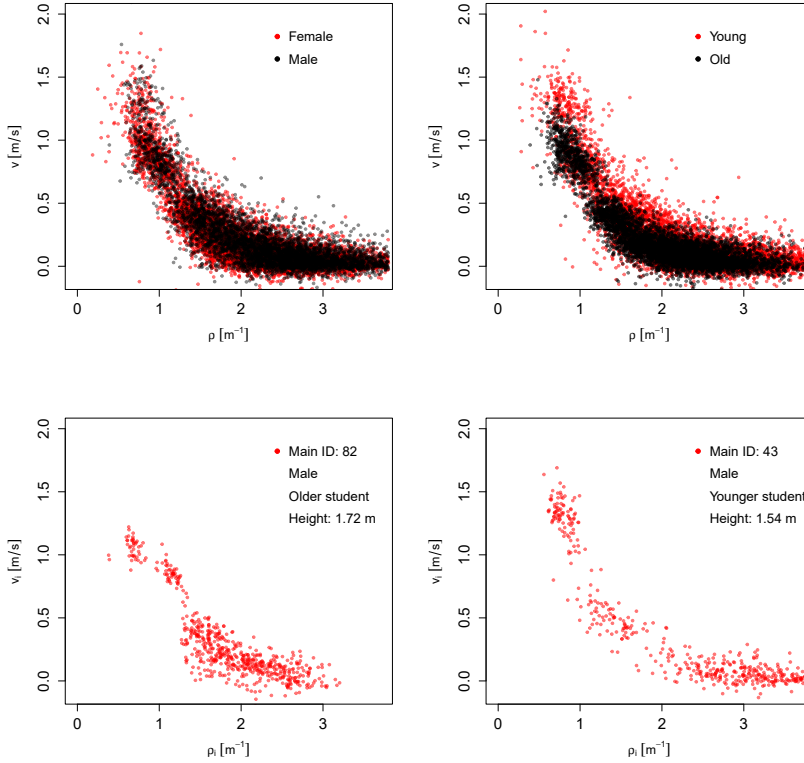


Figure 1.3: At the top are fundamental diagrams for homogeneous groups on one human factor and heterogeneous groups on other human factors. The left side shows the difference between the homogeneous groups in terms of gender, and the right side shows the difference between the homogeneous groups in terms of age. These groups differ in other human factors, such as height or weight. In comparison, the figures below illustrate two fundamental diagrams for two specific persons with main ID 82 and 43 with their human factors such as height, age, and gender.

fundamental diagrams illustrate homogeneous groups concerning one human factor. The diagram on the left represents the human factor of gender, and the diagram on the right represents age. Other human characteristics, such as height, gender, weight, or age, are heterogeneous. In comparison, below are two fundamental diagrams for

individual persons identified by a given main ID. These are only diagrams for a given gender, age, height, and so on. The figure on the left shows a graph for a male pedestrian in the group of older students with the main ID 82 and a height of 1.72 meters. In contrast, the figure on the right shows a graph for a male pedestrian in the younger student group with the main ID 43 and a height of 1.54 meters.

Using simple linear regression, the minimum distance, the reaction time, and the scatter of the data are analyzed to investigate whether individual speed-distance functions show that gender, age, or the gender of neighboring pedestrians affect speed at high densities. This is done for the two different single-file experiments in Germany.

The multiple linear regression analysis is conducted to investigate the influence of distance and individual factors on speed. In the case of the single-file experiment at GBS, human factors such as height, age, and gender are considered. A multiple linear regression analysis is then performed with additional human factors not previously considered, such as the weight or exact height and gender of the previous pedestrian, by considering the single-file experiment at MEH. This tests whether the multiple regression model developed from the GBS single-file experiment can be improved.

Furthermore, as the last part of this thesis, stop-and-go waves are considered and analyzed. In this context, the velocity distribution for different densities for homogeneous and heterogeneous group compositions in terms of gender is compared to the number of peaks that appeared. In addition, it is investigated if the different group compositions are reflected in the stop-and-go waves, e.g., the duration of stopping or going and the number of pedestrians in a stop state are considered.

In summary, the following hypotheses will be analyzed in this thesis for different single-file experiments:

1. The speed-density relation depends on the gender composition of the group of pedestrians.
2. Individual speed-distance functions depend on age and gender.
3. At high densities, the distance between individuals depends on the gender of the neighboring pedestrians.
4. The distance and individual human factors affect the speed.

5. The inclusion of additional human factors that were not previously included, such as the weight, exact height, and gender of the previous pedestrian, improves the multiple linear regression model developed based on the single-file experiment at GBS.
6. Stop-and-go waves are affected by different gender compositions.

## 1.4 Thesis Outline

This thesis is structured as follows. In chapter 2, Experiments and Methodology, a comprehensive overview of the various analyzed single-file experiments is presented. The measurement methods used to collect the data are explained in chapter 3 in the context of the considered single-file experiments. Moving on to chapter 4, namely Data Analysis, the main research questions addressed in this thesis are investigated. Single-file experiments with homogeneous and heterogeneous group compositions concerning gender will be considered to determine if there are significant behavioral differences between male, female, gender-alternated, and gender-random-order pedestrian groups. In addition, individual fundamental diagrams will be introduced, and the influence of individual characteristics will be quantified using regression analysis. Furthermore, stop-and-go waves will be analyzed. Finally, in chapter 5, the conclusions of this work will be discussed in detail. Moreover, this chapter also provides a preview of some exciting extensions that could be explored based on the results of this study.

## Chapter 2

# Experiments

This chapter presents three different single-file laboratory experiments whose data are relevant for data analysis and their comparison in chapter 4. The first experiment is the single-file experiment [53] conducted at the school Gymnasium Bayreuther Straße (GBS) in Wuppertal, Germany, in 2014. The second experiment is a single-file experiment performed within an experimental series [54] of the projects CroMa and CrowdDNA at the Mitsubishi Electric Halle (MEH) in Düsseldorf, Germany, in 2021. Lastly, a series of controlled single-file experiments at the Arab American University (AAU) in Palestine is described [56]. The GBS, Germany, and AAU, Palestine data were available before this research was conducted. For this reason, this dissertation only briefly describes these two experiments. However, the corresponding studies [22, 32, 56] are referenced for those who are interested in a more detailed description. In contrast, the single-file experiment within the CroMa and CrowdDNA experiments was explicitly planned and performed as part of this study to investigate homogeneous and heterogeneous group compositions in terms of gender in fundamental diagrams.

At the beginning of all three experiments, each participant was given specific instructions on walking behind the person in front of them in the oval geometry. This was done to avoid overtaking or rushing. Furthermore, in the experiments under consideration, the density is expressed in units of  $\text{m}^{-1}$ , as it is only a line density. This will become clear in the following subsection 2.1.2 by Equation 2.2. To provide an overview of these experiments, the chapter describes the setup, procedures, data collection, and data processing in detail.

## 2.1 Single-File Experiment at GBS in Germany

### 2.1.1 Experimental Setup

The experimental setup consists of an oval path with a central line, precisely measuring a total length of  $l_{cl} = 16.62$  m (see Figure 2.1 **a**). The specific dimensions of the experiment are clearly illustrated in Figure 2.1, **(a)**. An overhead view of the experiment with the students wearing colored caps on their heads is shown in Figure 2.1, **(b)**. The oval path itself has a width of  $w = 0.8$  m, and each straight

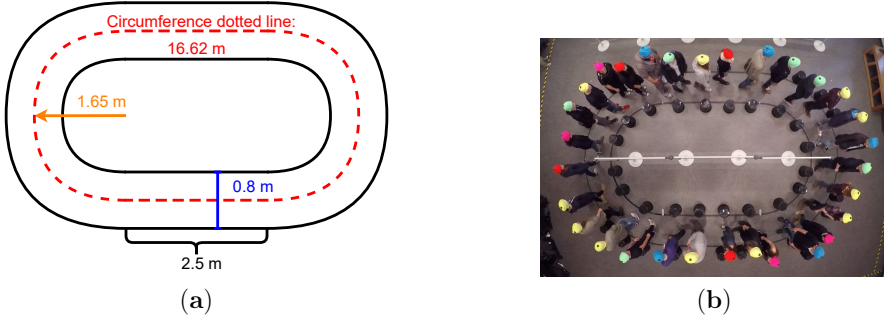


Figure 2.1: Single-file experiment conducted at Gymnasium Bayreuther Straße (GBS) in Wuppertal, Germany. The oval path with the corresponding lengths is shown on the left-hand side **(a)**. An overhead view of the experiment with the students wearing colored caps on their heads is shown on the right **(b)**. The colors indicate five different intervals of height.

section has a length of  $l_{sl} = 2.5$  m.  $r_0$  is the radius from the middle of the oval to the central line and is 1.65 m.

In total, 118 students participated in the experiment, and 46 % were male. The participants in this study are students from fifth and 11th grade, and they represent a wide range of age groups. The younger students are typically between 11 and 12, while the older ones are about 17 or 18 years old. Unfortunately, the exact age of the students is not available in the experiment. Therefore, the average age in the respective grades is assumed. During the experiment, students wear colored caps that serve a dual purpose. Firstly, the caps are used to extract the students' trajectories as they move along the oval setup. This data is crucial for analyzing their movement patterns. Second, the colored caps indicate the different height intervals of the participants [57, 58]. By categorizing the students into five different height intervals, it is possible to extract trajectories with a reduced error resulting from

the perspective distortion [59] and to analyze the effect of body height on speed. The average height of younger students is  $1.48 \text{ m} \pm 0.04 \text{ m}$ , while older students have an average height of  $1.76 \text{ m} \pm 0.07 \text{ m}$ . These height differences may affect how individuals move through the spatial structure and interact with others. Furthermore, each pedestrian is given a main ID. This unique identifier helps to identify and track a particular student across different runs. Therefore, their behavior can be analyzed in various runs and densities. The identification process and gender assignment are done manually by watching videos of the experiments. The videos were compared, and each participant was checked to see which videos they appeared in, resulting in a specific main ID.

In all, 31 runs are performed with varying global densities  $\rho_{gl} \in [0.32, 3.20] \text{ m}^{-1}$  which are calculated by

$$\rho_{gl} = N/l_m , \quad (2.1)$$

the number of persons  $N$  in one run in the measurement area divided by the length of the measurement area, which is set at  $l_m = 15.62 \text{ m}$ . The table provided in Table 2.3 gives an overview of the global densities for each single run in the penultimate column. For more detailed information on the experiment and its methodology, see [22]. This work provides a deeper insight into the experimental design, data collection methods, and analysis techniques used in this study.

### 2.1.2 Data Processing

To simplify the system and ease the analysis process, the participants' two-dimensional trajectories are transformed into a one-dimensional representation. In the first step, this transformation involved mapping the participants' trajectories onto the central line of the oval path. The central line was transferred to a straight line in the second step. This is called a rectification process and follows the methodology outlined by Ziemer [60]. This method is represented in the following Equation 2.2:

$$\lambda : \mathbb{R}^2 \rightarrow \mathbb{R}, \begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \hat{x} \text{ with}$$

$$\hat{x} = \begin{cases} x & , 0 \leq x \leq l_{sl}, y < r_0 \\ l_{sl} + r_0 \cdot \arccos\left(\frac{r_0 - y}{\sqrt{(x - l_{sl})^2 + (y - r_0)^2}}\right) & , x > l_{sl} \\ 2l_{sl} - x + r_0 \cdot \pi & , 0 \leq x \leq l_{sl}, y \geq r_0 \\ 2l_{sl} + r_0 \cdot \pi + r_0 \cdot \arccos\left(\frac{r_0 - y}{\sqrt{(x)^2 + (y - r_0)^2}}\right) & , x < 0 \end{cases} \quad (2.2)$$

The first case describes the rectification process of the lower straight line. The second describes the right-hand curve. The third case represents the upper straight section. Finally, the left-hand curve is defined by the fourth case (see Figure 2.1 a)). However, only the function  $\lambda$  is shown, which describes  $\hat{x}$ , the distance on the central line from the coordinate origin. The function for the transformation, which represents  $\hat{y}$ , is neglected because this analysis does not consider the directional distance. Thus, only the change in movement direction is considered [60]. For the single-file experiments at MEH the two-dimensional trajectories are transformed into a one-dimensional representation in a similar way to the above.

## 2.2 Single-File Experiment at MEH in Germany

This section presents a single-file experiment at MEH as part of the idea to conduct single-file experiments in different countries, where the experimental design is standardized in terms of geometry, number of people, and group composition (homogeneous and heterogeneous groups in terms of gender). This way, the experiments can be compared, and cultural influences on the fundamental diagram can be analyzed. In this context, the geometric structure, in particular the dimensions of the oval, the number of people, and the group's composition are considered essential factors. Another significant aspect of the experiment in Germany is the variation of densities. By varying the number of people from  $N = 4$  to  $N = 40$ , different densities should be included, from low to high. This allows a wide range of situations to be simulated and the effects of different densities on people's behavior to be studied. It is important to emphasize that conducting these experiments in other countries raises several challenges. These include coordinating local teams, sourcing materials and participants, and technical aspects. To minimize these, the design of experiments is crucial [61]. Among other things, they simplify communication in science, document the exact procedure of the respective experiments, and determine how they are carried out. By analyzing the cultural influences in different group compositions in the most straightforward experimental setup, it may be better to



## 2. Experiments

understand the general course of the fundamental diagram. A detailed description of the data preparation and the experimental setup follows.

### 2.2.1 Data Preparation

After the experiment had been carried out and the corresponding data were available, the trajectories were therefore corrected and created using the PeTrack software. PeTrack was used similarly in the other two single-file experiments. However, the description is only given in connection with section 2.2, as I only utilized PeTrack for this experiment. PeTrack automatically tracked the participants. The persons are tracked and recognized in each sequence. At each step, it is checked whether the position of the newly detected person is similar to that of the tracked person or whether it may be necessary to improve the position by repositioning it. More details about PeTrack can be found in [58]. Figure 2.2 shows an example of the trajectories of the run with  $N = 16$  female pedestrians. The trajectories are shown in red over the entire 2-minute period.

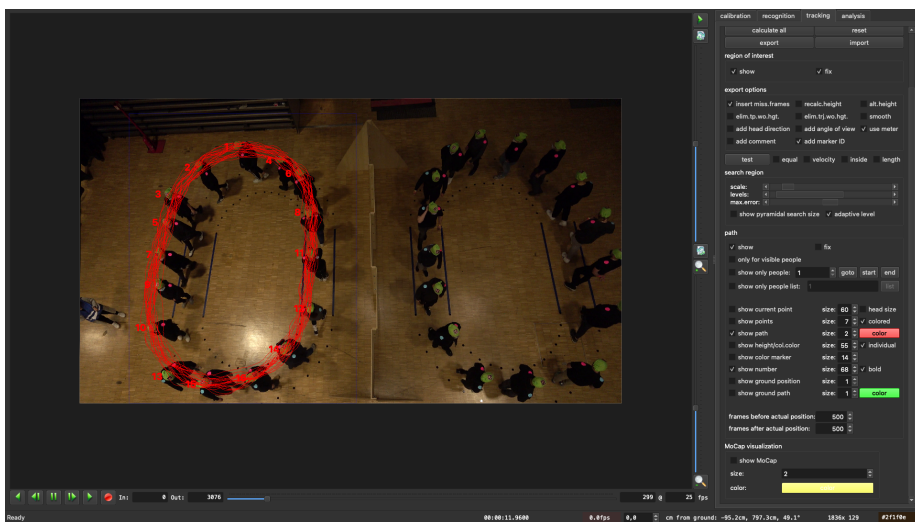


Figure 2.2: An example of the trajectories on the left of a run with a total number of  $N = 16$  female participants in PeTrack.

### 2.2.2 Experimental Setup

Concerning the experimental setup, the measurements of the oval path in the experiment are the total length of the central line  $l_{cl} = 14.97$  m and the width of  $w = 0.8$  m. The straight sections are  $l_{sl} = 2.3$  m long, while the central radius is 1.65 m. In the sketch on the left in Figure 2.3, the two measurement areas used for the data analysis are highlighted in the background. It is important to note that only the straight sections marked by the measurement area are considered for the analysis. This aims to compare the data from the same sections of the oval with the single-file experiments in Palestine. There, too, only the data of the straight areas are used. This experiment has no visual protection in the middle of the oval. There

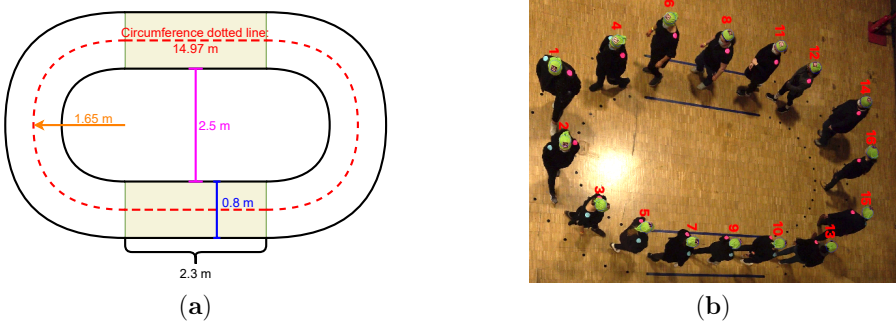


Figure 2.3: A single-file experiment conducted at the Mitsubishi Electric Halle (MEH) in Düsseldorf, Germany, in 2021. Figure (a) shows the oval path with the lengths of the central line, the width of the oval path, the middle radius, the length of the straight sections, and two measurement areas highlighted in the background. Figure (b) shows the experiment from above. The students are wearing green caps with personal ID codes on the top.

is only a barrier in the middle between the two ovals, as seen in Figure 2.2.

The experiment considered four different group compositions to study pedestrian behavior in a controlled single-file environment. The first two group compositions that were carried out were homogeneous in terms of gender, with one group consisting only of male participants (m) and the other entirely of female participants (f). This allowed observing and studying how individuals of the same gender interacted and moved within the oval space. The third group composition was heterogeneous, with male and female participants arranged in alternating order of m, f, m, f, etc. around the oval. This arrangement was designed to simulate a more diverse crowd, with individuals of different genders distributed throughout the space. The

## 2. Experiments

fourth group composition also featured a heterogeneous mix of male and female pedestrians, but this time in a random structure – for example, m, m, f, m, f, f, etc. This random distribution was intended to imitate a realistic scenario in which people are not organized according to particular criteria. To further explore the effect of group composition, ten experimental runs were conducted at different global densities for both homogeneous group compositions. The global density, represented by the number of persons situated in the oval  $N$ , was varied across runs:  $N = 4, N = 8, N = 16, N = 20, N = 24, N = 27, N = 32, N = 36$ , and  $N = 40$ . In contrast, all pedestrians participated in at least one run at each density for the experiments with heterogeneous group compositions. The densities of the heterogeneous group experiments matched those used in the homogeneous group experiments. However, in this case, 25 different experimental runs were performed instead of the ten runs conducted for the homogeneous groups. The number of different runs for each group composition can be seen in Table 2.1. In [54], section A.5. Experimental Configurations Oval Experiments, Table 15 shows the names of the experiments as well as varied parameters in detail. Global densities  $\rho_{gl} = N/l$  for these cases are  $\rho_{gl} \in [0.27, 2.67] \text{ m}^{-1}$ .

Table 2.1: The number of different runs for each group composition is detailed in the columns, female, male, alternating gender, and gender random order, for different numbers of pedestrians  $N$ .

N	$\rho_{gl}$	Female e.g., f, f, f, etc.	Male e.g., m, m, m, etc.	Gender alternating e.g., f, m, f, m, etc.	Gender random order e.g., f, f, m, f, etc.
4	0.27	1	1	10	10
8	0.53	1	2	5	5
16	1.07	1	1	3	3
20	1.34	2	2	2	2
24	1.60	1	1	2	2
27	1.80	0	1	0	0
32	2.14	1	1 ( $N=31$ )	1	1
36	2.40	1	1 ( $N=35$ )	1	1
40	2.67	1	0	1	1

The duration of each run varied between two and three minutes. Two minutes were chosen for runs with  $N < 32$  because, up to this density, the pedestrians had moved a long distance in the oval in the time considered. Participants can only move very slowly at higher densities, and longer distances are considered only in a longer time interval. In total, 80 different pedestrians took part in the experiment, with

## 2. Experiments

---

an equal number of male and female participants. This balanced ratio ensured that gender differences did not bias any observed behaviors.

The green caps worn by all participants with unique personal ID codes on the top served a dual purpose during the experiment. Firstly, they were crucial for data collection and analysis, as the codes were used to extract the trajectories of different participants in different experimental scenarios. Second, the codes assign personal information to each participant, including gender, age, shoulder width, weight, and height [57, 59]. Access to this personal information was essential to understanding how these factors could affect speed and the fundamental diagram. In comparison to the GBS experiment, the ID codes on the caps now also make it easier to assign individuals to different experimental runs.

The values in Table 2.2 provide a comprehensive overview of the mean values and the standard errors of the human physical characteristics of the participants in each group: female, male, gender alternating, and gender random order. Namely, they are age, height, weight, and shoulder width. The average age of the participants was between 26 years and 28 years. Overall, the age is relatively constant. There are some differences in these characteristics between the different groups, particularly in height, weight, and shoulder width. Their average heights ranged between 1.70 m and 1.83 m, and their average weights were between 76.26 kg and 92.24 kg. Lastly, their average shoulder widths were between 0.43 m and 0.49 m. It can be seen that there are expected differences between the male and female groups, such as the fact that the male participants are, on average, taller, heavier, and have wider shoulders.

Table 2.2: A detailed overview of the mean values and the standard errors of the human physical characteristics age, height, weight, and shoulder width of the participants in the four different group compositions: female, male, gender alternating, and gender random order.

	Female	Male	Gender Alternating	Gender Random
$\overline{age} \pm \sigma$ in years	$27.12 \pm 8.11$	$26.42 \pm 4.92$	$27.74 \pm 6.03$	$25.97 \pm 5.14$
$\overline{height} \pm \sigma$ in m	$1.70 \pm 0.08$	$1.83 \pm 0.07$	$1.75 \pm 0.09$	$1.77 \pm 0.11$
$\overline{weight} \pm \sigma$ in kg	$76.26 \pm 21.16$	$88.74 \pm 26.08$	$92.24 \pm 20.61$	$80.95 \pm 20.88$
$\overline{shoulder\ width} \pm \sigma$ in m	$0.43 \pm 0.03$	$0.49 \pm 0.03$	$0.45 \pm 0.04$	$0.46 \pm 0.05$

## 2.3 Single-File Experiment at AAU in Palestine

In the following, a brief overview of the single-file experiments conducted by Subaih et al. [35, 56] at AAU is presented. For more detailed information, please refer to [35].

### 2.3.1 Experimental Setup

In Subaih’s single-file study, an oval path with a total length of the central line of  $l_{cl} = 17.30$  m and a width of  $w = 0.6$  m, marked on the ground, was measured. The straight sections were  $l_{sl} = 3.15$  m long, and the radius up to the central line,  $r_0$  is 1.75 m (see Figure 2.4). In the middle of the oval is a boundary. Only the data

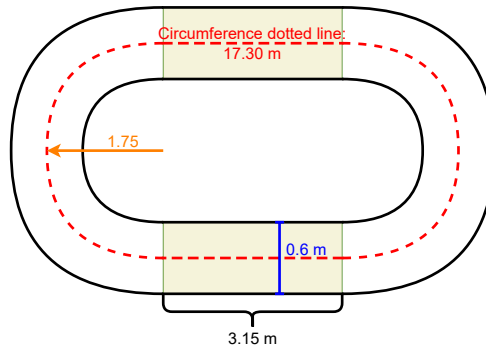


Figure 2.4: A single-file experiment conducted at the Arab American University (AAU) in Palestine, in 2018. The figure shows the oval path with the lengths of the central line, the width of the oval path, the middle radius, the length of the straight sections, and two measurement areas highlighted in the background.

from Subaih’s research, including velocity and density, has been used in this study. This means that the data was used exactly as Subaih created them and that no recalculation of the speed was performed. The background to this is that the same data used in [32] should be compared with the single-file experiments at MEH in Germany in chapter 4.

In the study conducted in Palestine, a total of 47 pedestrians participated in the experiment. Among them, there were 26 female students and 21 male students. The participants’ height ranged from 1.52 m to 1.84 m. The average height of the men was 1.75 m, while the average height of the women was 1.61 m. The age of the participants ranged from 18 to 23 years. The experiment included groups

with different compositions in terms of gender to investigate how group composition —homogeneous and heterogeneous groups —specifically gender diversity, affects speed. In the homogeneous groups, which included males and females, there were  $N = 14$  and  $N = 20$  participants in the oval. On the other hand, in the heterogeneous group composition where the gender alternated, there were also  $N = 24$  and  $N = 30$  pedestrians. In [32] the different groups are called UM (male), UF (female), and UX (gender alternated). Global densities in Palestine were  $0.81 \text{ m}^{-1}$ ,  $1.16 \text{ m}^{-1}$ ,  $1.38 \text{ m}^{-1}$  and  $1.73 \text{ m}^{-1}$ . The experimental scenarios of gender-homogeneous group composition were carried out first.

### 2.4 Data Selection

This section is about selecting the interval of the steady-state and the autocorrelation coefficients to ensure independent speed values for different experimental runs for the single-file experiments at GBS and MEH.

During the data analysis of these single-file experiments, only the manually selected steady-state is taken into account, chosen by looking at the average speed of the given run. The time at which the speed is for the first time higher than 90% of the average speed is the beginning of the steady-state interval (denoted as  $I_{min}$ ) and the time at which the speed is for the last time higher than 90% of the average speed is the end of the steady-state interval (denoted as  $I_{max}$ ). For the experiments at GBS, the interval for which the speed is stable for each experimental run can be found in Table 2.3 in columns four and five. For the experiments at MEH, the values can be found in Appendix 5.2, in Table A.1 (see columns four and five). As an example, the straight red lines in Figure 2.5 indicate the boundaries  $I_{min}$  and  $I_{max}$  of the steady-state in a specific run.

In the next step, the chosen time intervals  $[I_{min}, I_{max}]$  of the steady-states are considered in detail. The aim is to ensure independent measurements in the data analysis. One way to check the correlation of the data is to calculate autocorrelation coefficients for the different steady-state intervals. These coefficients indicate how strongly the observations are correlated with each other at different points in time. A high autocorrelation coefficient (e.g. 1) indicates a high degree of dependence in the data, while a low coefficient indicates a lower degree of dependence. If the data are too strongly correlated, the use of some statistical methods, such as regression analysis or ANOVA, will be severely limited. It may also lead to incorrect conclusions

## 2. Experiments

Table 2.3: A detailed overview of the experimental runs at GBS in Germany for the different groups of students from fifth grade, 11th grade, and both fifth and 11th grade and their general characteristics. Column by column, from left to right, the following information is given: The number of the run, the duration of a run in seconds, the interval of the steady-state with the beginning  $I_{min}$  and the end  $I_{max}$  in seconds, the total number of persons in a run, the number of male and female pedestrians in a run, the global density in  $m^{-1}$  and the mean values for individual specific time-steps and their standard deviation in seconds.

	Run	Duration	$I_{min}$	$I_{max}$	$N$	m	f	$\rho_{gl}$	$\overline{\delta t} \pm \sigma$
5th grade (GBS)	1101	93.75	7	77.60	16	8	8	1.02	$2.01 \pm 0.38$
	1102	123.12	0	123.12	50	28	22	3.20	$1.06 \pm 0.82$
	1103	167	2	167	40	20	20	2.56	$2.02 \pm 0.73$
	1104	138.52	0	138.52	32	20	12	2.05	$2.50 \pm 0.51$
	1105	156.16	12	105	10	6	4	0.64	$1.72 \pm 0.49$
	1106	119.08	0.4	119.08	24	13	11	1.54	$2.48 \pm 0.46$
	2101	77.64	6	77.64	24	13	11	1.54	$1.99 \pm 0.39$
	2102	123.12	6	123.12	5	0	5	0.32	$2.00 \pm 0.54$
	2103	137.28	10	137.28	11	4	7	0.70	$1.99 \pm 0.73$
	2104	155.40	0	155.40	34	16	18	2.18	$2.72 \pm 0.69$
	2105	118.64	12	118.64	16	8	8	1.02	$2.04 \pm 0.37$
	2106	75.16	0.8	75.16	28	18	10	1.79	$1.10 \pm 0.45$
	2107	44.60	4	44.60	27	18	9	1.73	$2.44 \pm 0.71$
	2108	89.36	8	89.36	14	5	9	0.90	$1.09 \pm 0.23$
11th grade (GBS)	1201	134.80	0	100	39	15	24	2.50	$1.27 \pm 1.06$
	1202	78.76	6	69.04	5	1	4	0.32	$0.99 \pm 0.39$
	1203	138	0	108	33	14	19	2.11	$1.10 \pm 0.61$
	1204	107.72	12	100	12	5	7	0.77	$1.48 \pm 0.41$
	1205	114.36	2	96	23	9	14	1.47	$1.14 \pm 0.43$
	1206	91.60	12	81.20	15	5	10	0.96	$1.83 \pm 0.26$
	1207	97.84	4	76	21	5	16	1.34	$1.55 \pm 0.40$
	2201	103	8	99.04	5	3	2	0.32	$1.39 \pm 0.32$
	2202	109.40	0	72	39	17	22	2.50	$1.31 \pm 0.63$
	2203	118.20	8.80	111.44	9	3	6	0.58	$1.31 \pm 0.18$
	2204	121.76	0	96	29	13	16	1.86	$1.17 \pm 0.41$
	2205	133.88	6	122.32	15	6	9	0.96	$1.75 \pm 0.27$
5th + 11th grade (GBS)	1301	142.84	0	104	42	20	22	2.69	$0.81 \pm 0.17$
	1302	140.92	0	104	44	25	19	2.82	$1.02 \pm 0.73$
	1303	91.08	4	89.44	5	2	3	0.32	$1.13 \pm 0.13$
	1304	134.12	0	104	33	17	16	2.11	$2.12 \pm 1.12$
	1305	76.96	4	70.68	11	5	6	0.70	$2.36 \pm 0.67$

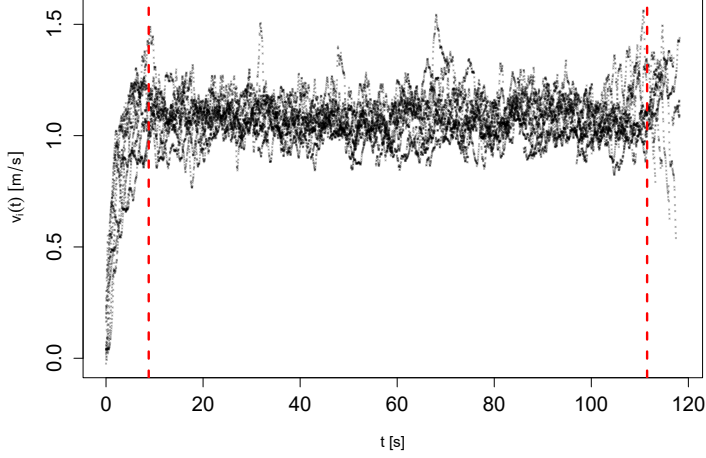


Figure 2.5: Illustration of the steady-state in a specific run, where the vertical red lines represent  $I_{min}$  and  $I_{max}$ , the beginning and end of the interval.

or over- or under-estimation of specific effects. To ensure that the statistical methods can be applied correctly, it is therefore crucial to carefully analyze and, if necessary, modify the time interval of the steady-state.

For this, the degree of dependence between individual speed observations  $v_i(t)$  in the case of a time lag is examined. To ensure that the data for each person  $i$  are statistically independent, the speed values are only taken into account when the autocorrelation function (acf)  $r_{i,\tau} < 0.3$  is achieved for the first time. The values of  $r_{i,\tau}$  are given by

$$r_{i,\tau} = \frac{\sum_{t=\tau+1}^{n_i-\tau} (v_{i,t-\tau} - \bar{v}_i)(v_{i,t} - \bar{v}_i)}{\sum_{t=1}^n (v_{i,t} - \bar{v}_i)^2}, \quad (2.3)$$

where  $\tau$  is the time lag,  $n_i$  the number of observations of individual  $i$  and  $i = 1, \dots, N$  the number of persons. This decision is applied to the length of the steady-state intervals for each individual. As a result,  $N$  individual-specific  $\delta_t$  time steps are calculated for each experimental run. Only the mean values of the individual-specific time steps,  $\bar{\delta}_t$ , are used to analyze the various runs. Please refer to Table 2.3 where the last column shows these values and the standard deviation. On average, the time gap between each measurement is approximately 1.64 seconds, based on all values for  $\bar{\delta}_t$  for the single-file experiment at GBS.



## 2. Experiments

---

After this, the data set can be used for further data analysis, including the regression and analysis of the different fundamental diagrams. To obtain more data points for each participant, data from a given individual is combined from experimental runs of various densities. The number of observations  $k = 1, \dots, n_i$  differs for each person  $i$  due to the combination of different runs and individual  $\delta_t$  time steps. This variation in the number of individual observations is a point of complexity in the analysis and a difference from other studies.

The values of  $\bar{\delta}_t$  for the single-file experiment at MEH in Germany can be found in Appendix 5.2, in Table A.1, too. On average, the time gap between each measurement is approximately 1.38 seconds to ensure independence between successive measurements, such as those for speed. Based on the average time gap values of 1.64 seconds for the single-file experiment at GBS and 1.38 seconds for the single-file experiment at MEH, it could also be assumed that the average time gap for a single-file experiment could be around 1.5 seconds. It can also be concluded that there is no relationship between  $\bar{\delta}_t$  and the associated global density  $\rho_{gl}$  (see Table 2.3 and Table A.1).

# Chapter 3

## Methodology

### 3.1 Measurement Methods

This section describes calculating various parameters, such as individual velocity, density, Voronoi tessellation, and headway. In this scenario, the function  $x_i(t)$  represents the location of person  $i$  in movement direction at a given time  $t$ , where  $t$  falls within the time interval  $[t_{min}, t_{max}]$  s. It should be noted that the pedestrian  $i + 1$  is moving in front of the person  $i$  while the person  $i - 1$  is following behind. To analyze the one-dimensional trajectories obtained by tracking the head in video recordings, the software JPSreport [62] is used.

#### 3.1.1 Headway, Voronoi Distance, Density and Individual Speed

In the single-file experiment at GBS, the headway  $h_i(t)$  of student  $i$  at time  $t$  is calculated by

$$h_i(t) = x_{i+1}(t) - x_i(t) . \quad (3.1)$$

This formula describes the distance between the centers of the heads  $x_i(t)$  and  $x_{i+1}(t)$  of person  $i$  and pedestrian  $i + 1$  at time  $t$ . Here, the density of person  $i$  is calculated by  $\rho_i(t) = \frac{1}{h_i(t)}$ . In the single-file experiment at MEH by contrast, the Voronoi distance  $d_{V_i}(t)$  of pedestrian  $i$  at time  $t$  is calculated by

$$d_{V_i}(t) = \frac{1}{2} \cdot (x_{i+1}(t) - x_{i-1}(t)) , \quad (3.2)$$

which is half of the distance between the centers of the heads  $x_{i+1}(t)$  and  $x_{i-1}(t)$ . The density of person  $i$  is  $\rho_i(t) = \frac{1}{d_{V_i}}$  in this case. For the experiment at MEH, the Voronoi distance (see Equation 3.2) is used in place of the headway formula (see Equation 3.1), for the experiment at GBS. The reason for this is, that the experiment will later be compared with the experiment at AAU in Palestine, in which the Voronoi distance is also used. As the GBS experiment had already been analyzed previously, this led to a change in the calculation of the distance.

The individual speed is calculated using the software JPSreport in all experiments with Equation 3.3. It applies:

$$v_i(t) = \frac{x_i(t + \frac{\Delta t}{2}) - x_i(t - \frac{\Delta t}{2})}{\Delta t} . \quad (3.3)$$

It is important to note that the default value  $\Delta t$ , which represents the time interval between two time points, is carefully chosen to minimize the effect of oscillations in the original trajectories. These oscillations are typically caused by motion in discrete steps, resulting in a microscopic periodic speed pattern in one dimension. By selecting an appropriate  $\Delta t$ , these oscillations are effectively smoothed out, allowing a more accurate analysis of the autocorrelation of speed without the need to account for these small fluctuations. This ensures that the focus remains on the overall trends and patterns in speed variations over time. Thus considering [63],  $\Delta t = 0.8$  s has been selected. The assumption of a fixed value is considered valid in this context. In addition, the data considers both the intended direction of travel and negative pedestrian velocities. This is important because, as illustrated in Figure 3.1, there is a tendency for noticeable oscillations to occur at lower speeds. Even when a pedestrian is standing still, small movements, such as shifting the weight from one leg to the other, can cause the head to sway, resulting in observable changes in the trajectory, including movement in the negative  $x$  direction. This phenomenon is exemplified by the interval  $x \in [9.5, 10.5]$  m in Figure 3.1, where fluctuations in position and velocity can be observed even when the pedestrians appear to be stationary. Consequently, instances of negative velocities are not uncommon in pedestrian motion.

#### 3.1.2 Stopping Phase

The following formulas and assumptions are used to analyze the stopping and going phases. In general, only the data in steady-state are considered. Based on the studies of Ziemer, Eilhardt, and Li et al. [22, 49, 52], the value 0.1 m/s is chosen

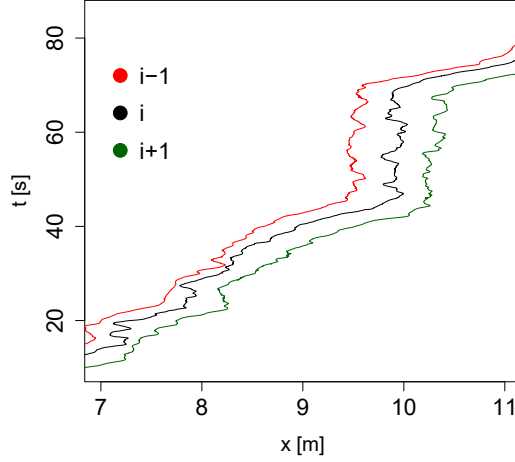


Figure 3.1: Walking position over time for three individuals, showing movements in the positive and negative  $x$ -direction.

as the speed limit assumption to distinguish between stopping and walking. This value is explained in more detail in section 1.2. So if  $v_i(t) < 0.1$  m/s a stop is defined for individual  $i$  at time  $t$  in  $[I_{min}, I_{max}]$ . First, to count the number of pedestrians stopping at a given time,  $v_{i_{stop}}(t)$  is introduced.

$$v_{i_{stop}}(t) = \begin{cases} 1, & v_i(t) < 0.1, \\ 0, & \text{else.} \end{cases} \quad (3.4)$$

Second, the number of individuals for which  $v_i(t) < 0.1$  is valid at time  $t$  is counted, see Equation 3.5.

$$ped_{stop}(t) = \sum_{i=1}^N v_{i_{stop}}(t), \quad (3.5)$$

where  $N$  is the number of all individuals. A further criterion this thesis considers for the stopping and moving phase is the average number of times a person stops. The formula Equation 3.6 is used for this criterion. This normalized formula is used for

this purpose so that the experiments can be compared with each other.

$$time_{stop_N}(i) = \frac{\sum_{t=\max(I_{min})}^{\min(I_{max})} v_{t_{stop}}(i)}{fps}, \quad (3.6)$$

where  $\max(I_{min})$  and  $\min(I_{max})$  are the values of the maximum of the minimum and the minimum of the maximum of the steady-state interval of the group compositions with the same  $N$ ,  $fps = 25$  frames per second, and  $v_{t_{stop}}(i)$  is defined by:

$$v_{t_{stop}}(i) = \begin{cases} 1, & v_t(i) < 0.1, \\ 0, & \text{else.} \end{cases} \quad (3.7)$$

To obtain the average time of a person within a given density at a stop, the individual values of  $time_{stop_N}(i)$  are summed and divided by  $N$  the number of pedestrians in the corresponding density.

## 3.2 Statistical Methods Related to Single-File Experiments

This section explains the regression models, the simple linear regression model, the multiple linear regression model and the mixed model, and the Tukey HSD test concerning the single-file experiments. The corresponding sections in chapter 4 describe the methods used for data analysis and verifying various assumptions. The regression model is also discussed regarding how well it predicts or explains the dependent variable. Furthermore, this study uses the statistical software R to analyze the data.

### 3.2.1 Individuals in Simple Linear Regression Models

In general, a simple linear regression model is given by

$$y = \beta_0 + \beta_1 \cdot w + \epsilon. \quad (3.8)$$

The model describes the relationship between  $y$ , a dependent variable, and  $w$ , one independent variable.  $\beta_0$  and  $\beta_1$  are unknown regression coefficients, and  $\epsilon$  is an error term that should ideally be as small as possible. The reason for assuming a linear relationship in the fundamental diagram is explained in detail in section 4.2.

### 3. Methodology

---

It is applied to the following two hypotheses that will be analyzed in this thesis. The Individual speed-distance functions depend on age and gender, and at high densities, the distance between individuals depends on the gender of the neighboring pedestrians. So the dependent variable  $y$  is represented by the speed  $v$  and the independent variable  $w$  is defined by the headway  $h$  or the Voronoi distance  $d_V$ . Depending on the number of individual observations  $n_i$ , the following simple linear regression model for one individual  $i$  results:

$$v_{k_i} = \beta_{0_i} + \beta_{1_i} \cdot h_{k_i} + \epsilon_{k_i} , \quad (3.9)$$

or

$$v_{k_i} = \beta_{0_i} + \beta_{1_i} \cdot (d_V)_{k_i} + \epsilon_{k_i} , \quad (3.10)$$

where  $k = 1, \dots, n_i$ . Assuming that  $i = 20$  and that there are 48 individual observations for this person, the following formula results from Equation 3.9 as an example:

$$\begin{pmatrix} v_{120} \\ \vdots \\ v_{4820} \end{pmatrix} = \beta_{0_{20}} + \beta_{1_{20}} \cdot \begin{pmatrix} h_{120} \\ \vdots \\ h_{4820} \end{pmatrix} + \begin{pmatrix} \epsilon_{120} \\ \vdots \\ \epsilon_{4820} \end{pmatrix}$$

In the context of this study, the values  $\beta_{0_i}$  and  $\beta_{1_i}$  (in the example:  $\beta_{0_{20}}$  and  $\beta_{1_{20}}$ ) need to be estimated for a good fit. This based on the least squares method for each participant  $i$ , e.g., for each main ID. They represent the intercept and the slope of the regression line. This results in the following general equation for the regression line with the best values for  $\beta_{0_i}$  and  $\beta_{1_i}$  for each individual:

$$\hat{y}_{k_i} = \hat{\beta}_{0_i} + \hat{\beta}_{1_i} \cdot w_{k_i} . \quad (3.11)$$

Regarding the values that  $\hat{\beta}_{1_i}$  can take, three cases are distinguished:

$$\hat{\beta}_{1_i} \begin{cases} > 0, \text{positive correlation between } w \text{ and } y \\ < 0, \text{negative correlation between } w \text{ and } y \\ = 0, \text{no correlation between } w \text{ and } y, \end{cases}$$

whereby the third case  $\hat{\beta}_{1_i} = 0$  will not exist with real data.

Furthermore, the number of different values of  $\hat{\beta}_{0_i}$  and  $\hat{\beta}_{1_i}$  corresponds to the total number of people participating in an experiment. In the case of the single-file

experiment at MEH there are 80 participants. This means that there are 80 different values for both  $\hat{\beta}_{0_i}$  and  $\hat{\beta}_{1_i}$ , corresponding to each participant.  $\epsilon_{k_i}$  is calculated by subtracting the estimated value of  $\hat{y}_{k_i}$  from the true value of  $y_{k_i}$ :

$$\epsilon_{k_i} = y_{k_i} - \hat{y}_{k_i}.$$

### 3.2.2 Human Factors in Multiple Linear Regression Models

In contrast to simple linear regression, multiple linear regression is a statistical technique used to analyze the relationship between multiple independent variables and a single dependent variable. This method predicts the dependent variable based on the linear combination of two or more independent variables. The main goal of multiple linear regression is to determine the strength and direction of the relationships between the independent and dependent variables. As mentioned in subsection 3.2.1, the dependent variable is still the same, so  $y_m$  represents the speed. The independent variables  $w_1, \dots, w_j$  are defined by different factors such as headway, age, or height. The analysis results will provide information on how each independent variable contributes to predicting the dependent variable and which factors may affect the speed at which pedestrians walk.

In general, a multiple linear regression model is given by

$$y_m = \beta_0 + \beta_1 \cdot w_1 + \dots + \beta_j \cdot w_j + \epsilon_m \text{ with } m = 1, \dots, n, \quad (3.12)$$

and  $n$  is the number of all observations over all pedestrians  $N$ , and  $j$  is the number of the independent variables. The model is now presented in detail concerning the given context, considering a total number of individuals  $N$ . This is done using matrix notation:

$$v = M \cdot \beta + \epsilon \text{ with} \quad (3.13)$$

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \quad M = \begin{pmatrix} 1 & w_{1,1} & w_{1,2} & \dots & w_{1,N+3} \\ 1 & w_{2,1} & \dots & \dots & w_{2,N+3} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & w_{n,1} & \dots & \dots & w_{n,N+3} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_4 \\ \vdots \\ \beta_{N+3} \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

The regression coefficients  $\beta_4$  to  $\beta_{N+3}$  can also be represented as the following sum:

$$\sum_{i=4}^{N+3} \beta_{4_i}.$$

$\beta_{4_i}$  is a regression coefficient that refers to an independent variable —introduced as *alloence* (all other unknown individual influences) in the equations in chapter 4 in more detail— that depends on  $N$ , the number of individuals, and therefore occurs  $N$ -time. As the fourth regression coefficient in the later equations,  $i$  starts at  $i = 4$ . Fulfilling certain assumptions is necessary to apply the model to the data and interpret the regression analysis's results [64]. One of the most essential assumptions for successful regression analysis is linearity. The dependent variable  $y$  and the independent variable  $w$  must have a linear relationship. This means that the dependent variable can be represented as a linear function of the independent variables. Another critical assumption is homoscedasticity. The residuals, the difference between the observations and the predictions, must have a constant variance. If the variance of the residuals is not constant, this is called heteroscedasticity. This can affect the validity of the regression analysis. For meaningful regression analysis, normality is also a necessary assumption. The error components should be normally distributed. This allows statistical tests to be performed correctly. There should also be no multicollinearity between the independent variables. Multicollinearity occurs when two or more independent variables have a high correlation. If two independent variables are too highly correlated, then only one of them should be used in the regression model. This is why the independent variables, such as age, gender, or height, are investigated for their correlation. Finally, there should be no autocorrelation in the components of the error.

#### 3.2.3 Model Comparison with Mixed Models

The concept of mixed models that include both fixed and random effects is considered in this section. These models are particularly relevant to the study of pedestrian dynamics as they allow the analysis of individual and fixed characteristics of people, as discussed in [65–67]. Mixed models are valuable when dealing with correlated data, such as repeated measures and multilevel data, with multiple individual or group observations. In previous sections, simple and multiple regression analyses have only included fixed effects. This approach treats all data points as independent without considering any hierarchical structure. However, data points



are often not truly independent, as they may be produced by the same individual or grouped based on some other characteristic. The data are considered hierarchical in these situations, and statistical models should consider this structure. Random effects will be incorporated into the model to capture the hierarchical structure of the data and address this issue. By incorporating random effects, the model can better account for dependencies within the data and provide more accurate results when analyzing complex data sets in pedestrian dynamic research. Figure 3.2 illustrates how data is structured in hierarchical levels. Here, the sample space  $\Omega$  is the single-file experiment. It is then further structured into a group composition such as a homogeneous or heterogeneous group, a specific density, and a pedestrian  $i$  with a fixed set of variables. This allows the comparison of person  $i$  and person  $i + 1$  in the same run and analyze their differences, or pedestrian  $i$  in different group compositions. The difference between fixed and random effects is explained as

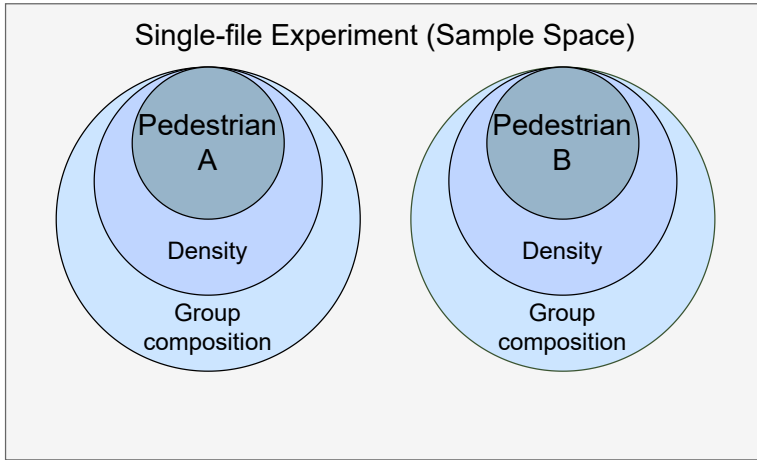


Figure 3.2: Representation of how data related to a single-file experiment is structured hierarchically.

follows:

- Fixed effects are observable variables, such as headway, height, age, or gender. They are a fixed set of variable levels.
- Random effects are used in statistical modeling to account for variability at a higher level where data points are grouped. Unlike fixed effects, random effects are categorical and represent a sample of infinite possible levels. These levels

may include individual characteristics that are not considered as fixed factors, such as personal attributes such as attention or motivation. In the context of this study, random effects are incorporated into the variable *alloence*.

In general, a mixed model is given by Equation 3.14:

$$y = X \cdot \beta + Z \cdot u + \epsilon \text{ with} \quad (3.14)$$

$y$  is an unknown vector of observations,  $X$  is the known matrix for the fixed effects,  $\beta$  is an unknown vector of fixed effects,  $Z$  is the known matrix for the random effects,  $u$  is an unknown vector of random effects and  $\epsilon$  is an unknown vector of random errors. Parameters are estimated by optimizing the objective function in a mixed model. The analysis uses the lme4 package [66] in R. In this study, the mixed model is used to analyze whether the model explains more of the variance than the regression model with only fixed effects and is, therefore, preferred.

#### 3.2.4 Tukey HSD Test and Gender Group Compositions

This explanation of the Tukey HSD (honestly significant difference) test is based on [68–70]. The test is used in chapter 4, Data Analysis.

The Tukey method is a statistical procedure developed by John Tukey to perform multiple comparisons in a single step. This test aims to determine which means are significantly different from each other. This thesis shows the difference between mean speed values in density intervals for various gender group compositions. The key concept behind the Tukey test is to calculate the statistically significant difference between all pairs of means based on a studentized range distribution ( $q$ ), called the  $q$  distribution. Performing the Tukey HSD test in R requires using the TukeyHSD function. This function allows the comparison of means across multiple groups and the identification of statistically significant differences. It can be used to determine whether observed variations between group means are due to chance or represent actual differences in group mean values by calculating the honest significant difference. It is important to note that the assumptions of the Tukey test, such as the data's normality and the variances' homogeneity, should be met to obtain accurate results. The next step is to explain the null hypothesis  $H_0$  and the alternative hypothesis  $H_1$ .

- $H_0$ : All means of the tested groups are equal ( $\mu_i = \mu_j$ )
- $H_1$ : At least one mean is different ( $\mu_i \neq \mu_j$ )

where  $i$  and  $j$  represent two different groups, e.g., male and female or male and gender alternating order. The test is based on the formula for the sample t-test statistic in Equation 3.15.

$$q_s = \frac{|\overline{X_A} - \overline{X_B}|}{SE}, \quad (3.15)$$

where  $\overline{X_A}$  and  $\overline{X_B}$  are the two means being compared, and  $SE$  is the standard error for the sum of the mean values.  $\overline{X_A}$  could be the average speed, e.g., for female pedestrians in a given density interval, and  $\overline{X_B}$  could be the average speed for, e.g., male pedestrians. The value  $q_s$  is compared to the value  $q_\alpha$  to decide whether to accept or reject the null hypothesis. Here, the significance level is  $\alpha = 0.05$ , and the critical value for  $q_{0.05}$  is chosen from a table of studentized range distributions. The null hypothesis is rejected if  $q_s > q_\alpha$ . Then, the means are different at the level of  $\alpha : 0 \leq \alpha \leq 1$ .

## Chapter 4

# Data Analysis and Results

This chapter presents a detailed insight into different statistical methods of analysis for single-file experiments and the results of the hypotheses, listed in chapter 1, section 1.3, which are of great importance for the relationships in fundamental diagrams. Once the data had been prepared and specific criteria as mentioned in chapter 2 and chapter 3 had been defined to ensure that only information relevant to the research questions were included in the study, different hypotheses could be tested. By systematically analyzing the data, interesting patterns and trends could be identified.

### 4.1 Gender Compositions in Fundamental Diagrams

This section compares homogeneous and heterogeneous gender group compositions in different density intervals and examines their influence on speed. First, a single-file experiment at the MEH in Germany is considered. Further, this experiment is compared with a single-file experiment from Palestine at the AAU. Results for homogeneous and heterogeneous gender group compositions in Palestine are already available for this comparison. Overall, in this section, the data are investigated concerning the initial hypothesis that the gender composition of a group of participants influences the speed-density relationship.

#### 4.1.1 Gender Composition in Germany

Concerning the hypotheses mentioned before, an analysis is conducted across different density intervals in Germany to determine whether there are consistent speed differences between homogeneous and heterogeneous group compositions based

#### 4. Data Analysis and Results

on gender. First, a visual comparison is made. This visual representation allows a first estimation of how homogeneous and heterogeneous groups with varying gender compositions may influence speed within different density intervals. Figure 4.1 illustrates a density vs. velocity fundamental diagram for groups categorized as

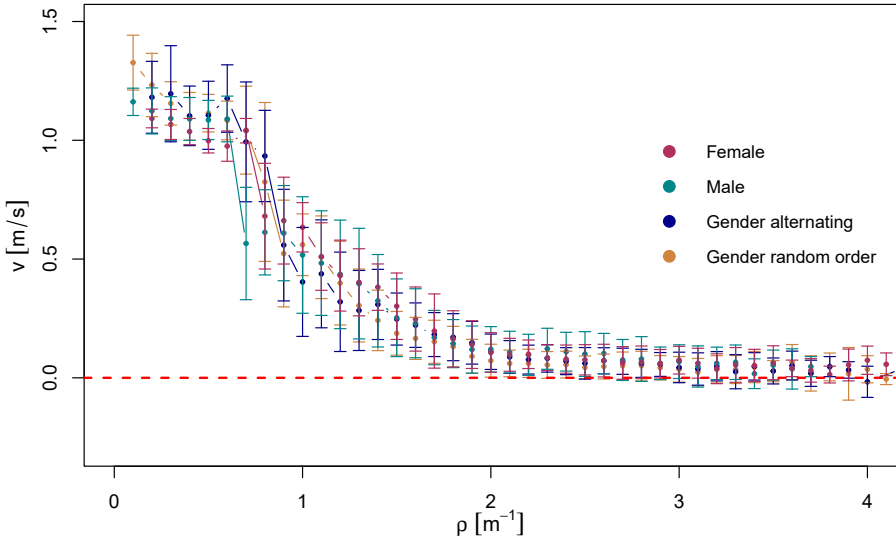


Figure 4.1: Fundamental diagram of density vs. velocity for the groups female, male, gender alternating, and gender random order at MEH in Germany with binned data within 0.1 density intervals [71].

the fundamental diagram, it is possible to see if there are any patterns or differences in speed based on the gender composition of the group. In addition, the data presented in Figure 4.1 provides a basis for a more detailed statistical analysis. The data is binned to allow more precise observation of trends and potential variances. For example, by binning the data according to density intervals for different gender compositions, a deeper understanding of the implications of group dynamics on speed can be gained. This approach allows a comprehensive investigation of the relationship between speed, density, and gender composition within groups of participants. The data suggest that the correspondence of the fundamental diagrams depends on the specific density intervals considered. For densities up to about  $1.15 \text{ m}^{-1}$ , there appears to be no systematic variation in inequality and equality between different group compositions. However, as densities increase beyond  $1.15 \text{ m}^{-1}$  but remain

#### 4. Data Analysis and Results

below  $2.0 \text{ m}^{-1}$ , it becomes apparent that homogeneous group compositions have higher velocities compared to heterogeneous groups. Interestingly, for densities above  $2 \text{ m}^{-1}$ , the individual curves show a notable similarity in their course. For a more thorough analysis, the mean velocity values for each group composition are compared across seven small density intervals:  $[0.15, 0.25]$ ,  $[0.55, 0.65]$ ,  $[0.85, 0.95]$ ,  $[1.05, 1.15]$ ,  $[1.25, 1.35]$ ,  $[2.05, 2.15]$ , and  $[3.05, 3.15]$ . This selection of intervals is designed to examine the differences between homogeneous and heterogeneous groups in more detail. The results of this analysis are presented in the following section.

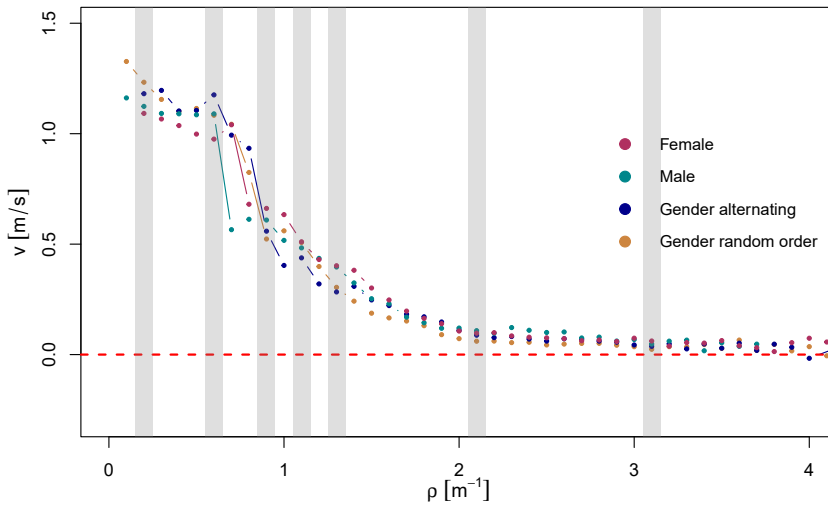


Figure 4.2: Fundamental diagram of density vs. velocity for the groups female, male, gender alternating and gender random order at MEH in Germany with binned data representing only the mean values of velocity. In addition, the single intervals  $[0.15, 0.25]$ ,  $[0.55, 0.65]$ ,  $[0.85, 0.95]$ ,  $[1.05, 1.15]$ ,  $[1.25, 1.35]$ ,  $[2.05, 2.15]$ , and  $[3.05, 3.15]$  are highlighted in grey [71].

are visually highlighted in grey within the fundamental diagram shown in Figure 4.2. When analyzing the mean velocity values within these selected density intervals, areas where velocity differences between homogeneous and heterogeneous groups are most pronounced can be identified. First, the Kolmogorov–Smirnov test is used to analyze the velocity distribution across all group compositions, which revealed significant differences between the distributions in each interval. A Tukey HSD test is then performed to determine whether two means for pairwise comparisons of all group compositions are significantly different. Table 4.1 provides detailed information on the mean speed values and standard deviations for each group

#### 4. Data Analysis and Results

composition in German and Palestinian settings for all seven intervals. The results of the experiments conducted by Subhai et al. in Palestine will be discussed further in the following subsection of this comparative analysis in subsection 4.1.2. In addition, Table 4.1 shows which groups are equal in the pairwise comparisons of all the group compositions. Therefore, the same colors indicate equality between the corresponding groups in an interval. For each interval, comparisons between groups where the

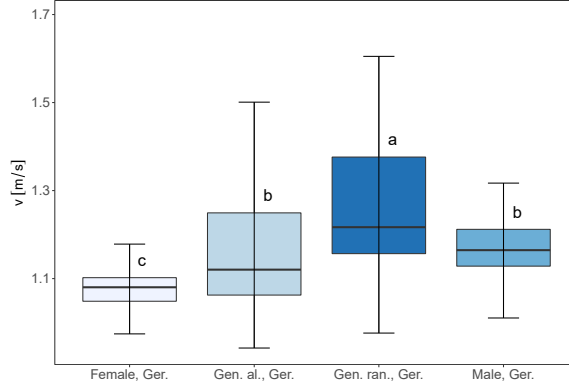
Table 4.1: Means and standard deviations for velocity ( $\bar{v} \pm \sigma$ ) in seven different density intervals for different group compositions in Germany and Palestine. The same colors in an interval indicate equality between the corresponding groups [71].

	[0.15, 0.25]	[0.55, 0.65]	[0.85, 0.95]	[1.05, 1.15]	[1.25, 1.35]	[2.05, 2.15]	[3.05, 3.15]
Female, Germany	1.07 $\pm$ 0.04	1.10 $\pm$ 0.05	0.62 $\pm$ 0.14	0.58 $\pm$ 0.12	0.41 $\pm$ 0.14	0.10 $\pm$ 0.07	0.06 $\pm$ 0.06
Male, Germany	1.17 $\pm$ 0.09	1.11 $\pm$ 0.10	0.61 $\pm$ 0.19	0.48 $\pm$ 0.23	0.44 $\pm$ 0.21	0.11 $\pm$ 0.09	0.06 $\pm$ 0.08
Gender alternating, Germany	1.17 $\pm$ 0.16	1.14 $\pm$ 0.15	0.74 $\pm$ 0.24	0.43 $\pm$ 0.22	0.30 $\pm$ 0.19	0.10 $\pm$ 0.07	0.05 $\pm$ 0.06
Gender random order, Germany	1.26 $\pm$ 0.14	1.11 $\pm$ 0.07	0.54 $\pm$ 0.30	0.55 $\pm$ 0.15	0.35 $\pm$ 0.16	0.07 $\pm$ 0.07	0.04 $\pm$ 0.06
Female, Palestine			1.14 $\pm$ 0.01	0.80 $\pm$ 0.11	0.68 $\pm$ 0.07		
Male, Palestine			0.94 $\pm$ 0.20	0.81 $\pm$ 0.13	0.74 $\pm$ 0.13		
Gender alternating, Palestine			1.00 $\pm$ 0.18	0.70 $\pm$ 0.14	0.50 $\pm$ 0.18	0.11 $\pm$ 0.05	

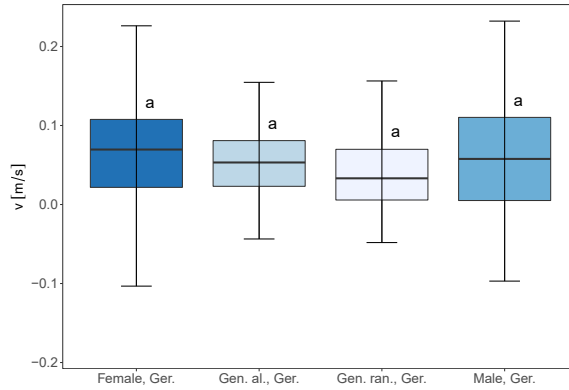
$p$ -value is larger than 0.05 — indicating that their means are equal — are highlighted in the same color. Within the first four intervals, ranging from 0.15 m<sup>-1</sup> to 1.15 m<sup>-1</sup>, both the group with the highest change in speed and the composition of the groups that are considered equivalent by the Tukey test vary from one interval to the next. Initially, there is significant equality between the two means between the male and gender alternating groups. In the second interval, equality is observed between the male, gender alternating, and gender random order groups, followed by equality between the female, male, and gender random order groups; and finally, in the fourth interval, equality is found between the female and gender random order and male and gender alternating groups. As a result, there does not appear to be a consistent pattern of equality or inequality between the different group compositions in the low-density regime. Within the range of densities between 1.25 m<sup>-1</sup> and 1.45 m<sup>-1</sup>, it

is observed that the mean values of the homogeneous group compositions either equal or are above those of the heterogeneous group compositions. This suggests a degree of unity and consistency within groups with similar characteristics or properties. There is a trend in the data in which groups that are more homogeneous in their composition tend to have higher mean values than groups that are more diverse. Moving into the density range between  $1.45 \text{ m}^{-1}$  and approximately  $2.15 \text{ m}^{-1}$ , an interesting pattern appears. The two means are significantly equal for all group compositions except the gender random order group. However, it is worth noting that the speed of the gender random order group stands out as being the lowest of all the groups in this interval. At a density of  $2.15 \text{ m}^{-1}$  and above, there is a shift in which equality is observed across all of the group compositions. This suggests that mean differences between group compositions decrease as density increases beyond a certain threshold. The boxplots in Figure 4.3 illustrate the results of the Tukey test in addition to Table 4.1. The figure on the top, denoted as Figure 4.3a), illustrates the first interval  $[0.15, 0.25]$ , where equality is only observed between the groups of male and the groups of alternating gender. The final interval  $[3.05, 3.15]$ , where all group compositions show equality, is shown in the figure at the bottom, labeled as Figure 4.3b). The boxplots provide a snapshot of central tendencies and variability within each group and highlight any disparities or similarities among different compositions. Letters above the boxes in the figures visually represent this distinction between equality and inequality of group composition. The identical letters indicate equality between particular group compositions. In addition, the boxplots shown in both figures provide a comprehensive overview of the data distribution within each group composition. These boxplots show key statistical measures such as minimum and maximum values, the median—represented by the line inside the box—and lower and upper quantiles that define the box’s boundaries. The intensity of the boxes’ colors corresponds to the median value within each group composition. Darker colors indicate that the median is higher. The progression from unequal distributions in the first interval to complete equality in the last interval provides a compelling view of the evolution of group dynamics over different densities. Furthermore, trends and patterns that may not be immediately apparent from the raw data alone can be identified by examining these patterns across intervals.





(a)



(b)

Figure 4.3: Boxplots of the velocity for different group compositions at MEH in Germany. Equal letters above the boxes indicate equality of the mean velocity values between the corresponding groups. The first density interval  $[0.15, 0.25] \text{ m}^{-1}$  is represented by (a) and the last density interval  $[3.05, 3.15] \text{ m}^{-1}$  is represented by (b) [71].

### 4.1.2 Gender Compositions and Cultures

Further, the data from the experiments at the AAU are compared with those at the MEH. However, there are some differences between the experimental setup at AAU in Palestine and the setup at MEH in Germany. For example, the oval size in Palestine is larger than the camera area in the single-file experiments in Germany, which requires an adjustment of the values in the oval. In Germany, there is also no boundary in the middle of the oval, potentially affecting people's behavior. Between the two ovals, there is only a barrier in the middle. These two experiments are compared in Table 4.2 to show their similarities and differences. It is necessary to consider these differences or similarities when interpreting the results. Compared to the studies conducted at MEH in Germany, the participants

Table 4.2: Comparison of the differences in the experimental design for the single-file experiments at MEH in Germany and AAU in Palestine regarding the total number of pedestrians  $N$ , gender ratio, densities in  $\text{m}^{-1}$ , age in years, height in m, and the length of the central line in m.

	MEH, Germany	AAU, Palestine
Total $N$	80	47
Ratio m/f	50% male	45% male
$\rho_{gl}$ in $\text{m}^{-1}$	[0.27, 2.67]	[0.81, 1.73]
$\overline{age}$ in years	[26, 28]	[18, 23]
$\overline{height}$ in m	[1.70, 1.83]	[1.52, 1.84]
$l_{cl}$ in m	14.97	17.30

in Palestine were younger and of shorter stature. In Palestine, the central line and the straight sections are longer in the oval, and the width is smaller. For both experiments, the experimental scenarios of gender-homogeneous group composition were carried out first. In addition, the experiments in Germany and Palestine used the same measurement methods.

First of all, the binned data are plotted in a fundamental diagram up to a density of  $2.5 \text{ m}^{-1}$  (see Figure 4.4). The upper graph shows the data for Germany, while the lower graph shows the data for Palestine.

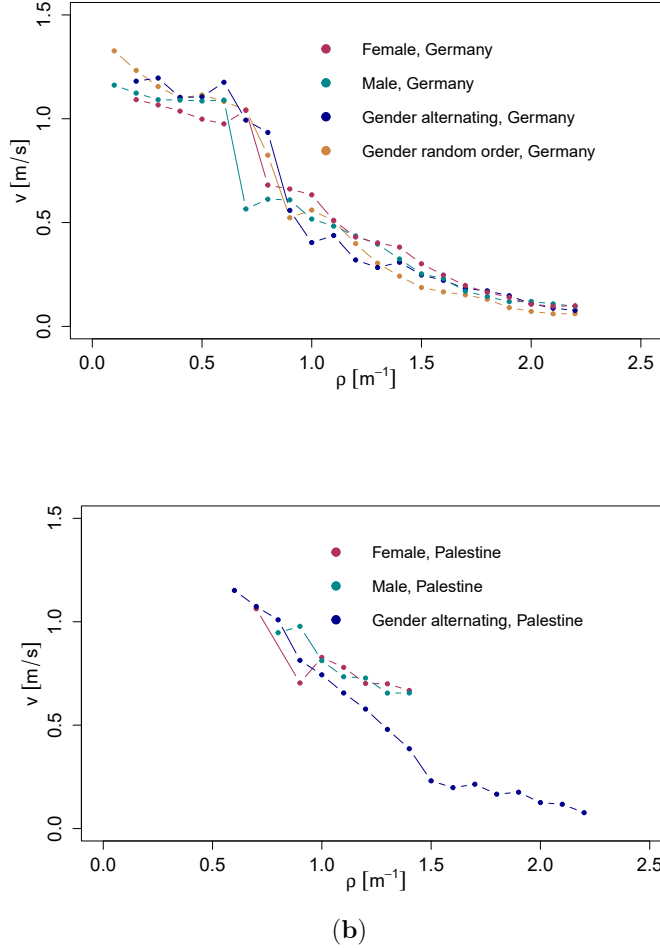


Figure 4.4: Relation between density and velocity for the experiments in (a) at MEH in Germany and (b) AAU in Palestine by binned data of mean values of velocity for different group compositions [71].

Looking at the data from both countries in more detail, it is evident that patterns are emerging regarding the groups' composition and corresponding speeds. A more consistent relationship between density and speed across different group compositions appears in Germany. In Palestine, however, there is more variation in the effect of group composition on speed at different densities. A visual analysis of Figure 4.4 shows that in the density range  $[0.6, 1.5]$  m<sup>-1</sup>, the mean speed of the experiments

#### 4. Data Analysis and Results

---

conducted in Palestine is higher than that of the experiments conducted in Germany. At densities above  $1.5 \text{ m}^{-1}$ , the heterogeneous group in Palestine begins to approach the average speed of the groups in Germany.

Focusing only on the Palestinian data, the heterogeneous group composition in Palestine shows higher speeds at densities below  $0.8 \text{ m}^{-1}$ . As density increases, the homogeneous group composition of males and females shows higher mean speeds than the heterogeneous group composition. The differences observed between Germany and Palestine could be attributed to various factors such as cultural norms, social dynamics within the groups, or even the environmental conditions during the experiments, e.g., in Palestine, only students took part in the experiment, and the professor was present here. In addition, it could be that the participants in Palestine are younger, and some are smaller than in Germany, as mentioned in Table 4.2.

Tukey's HSD test is again applied to statistically analyze the German and Palestinian experimental data. The respective means and standard deviations of velocity are shown in Table 4.1 for the experiments conducted in both countries. It is important to note that for the experiments conducted in Palestine, data are available only for density intervals ranging from  $[0.85, 0.95] \text{ m}^{-1}$  to  $[2.05, 2.15] \text{ m}^{-1}$ . When the results are analyzed, it is found that up to a density of about  $1.0 \text{ m}^{-1}$ , the means are statistically the same for all group compositions in Palestine, as well as for the gender alternating group in Germany, and the group compositions of female, male, and gender random order in Germany. Similarly, in the interval  $[2.05, 2.15] \text{ m}^{-1}$ , there are no significant differences between the means of different group compositions within Germany or across all group compositions in Palestine. When the density exceeds  $1.25 \text{ m}^{-1}$ , it is observed that the composition of the heterogeneous groups in Palestine approaches the composition of the groups in Germany. Only within the interval  $[1.25, 1.35] \text{ m}^{-1}$  do homogeneous group compositions show higher mean speed values than heterogeneous groups, a trend observed in both Germany and Palestine. In other density intervals, no significant differences are found in Palestine, while Germany shows no systematic variations. The previously described progression within each density interval is also shown in Figure 4.5. Each point represents the equality between the groups named above the points. The arrows show the change from one interval to the next. From the density interval  $[1.05, 1.15] \text{ m}^{-1}$  to the interval  $[1.25, 1.35] \text{ m}^{-1}$ , the speed in Palestine is higher than in Germany. In the density interval  $[2.05, 2.15] \text{ m}^{-1}$ , however, the values are the same for all available groups in Germany and Palestine.

Figure 4.6 visually presents the results of the Tukey test, focusing specifically

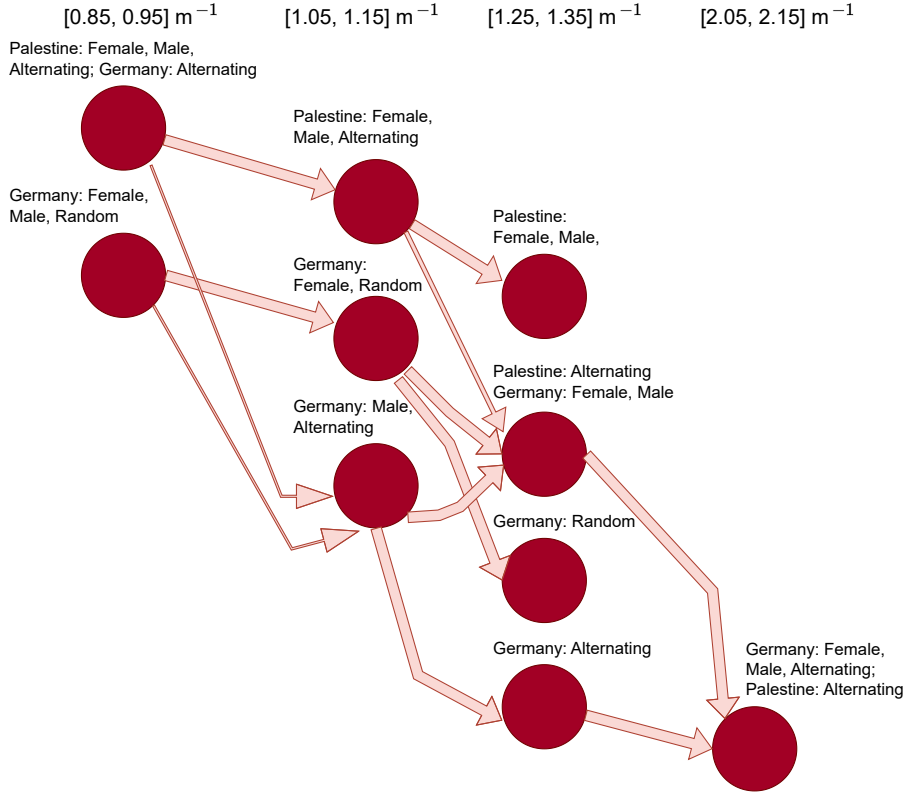


Figure 4.5: The chart illustrates the similarity between the different group compositions regarding gender at MEH in Germany and AAU in Palestine in four different density intervals. Each point represents the equality between the groups named above the points. The arrows show how to move from one interval to the next.

on the comparison between Germany and Palestine within the density interval of  $[1.25, 1.35] \text{ m}^{-1}$ . The letters above the boxplots are the same for the female and male groups in Palestine (a), the female and male groups in Germany, and gender alternating in Palestine (b). For the gender alternating and gender random order groups in Germany, inequality is only in (c) and (d).

## 4.2 Individual Speed-Distance Functions

After the fundamental diagrams have been examined for differences between homogeneous and heterogeneous groups regarding gender, the focus now shifts

#### 4. Data Analysis and Results

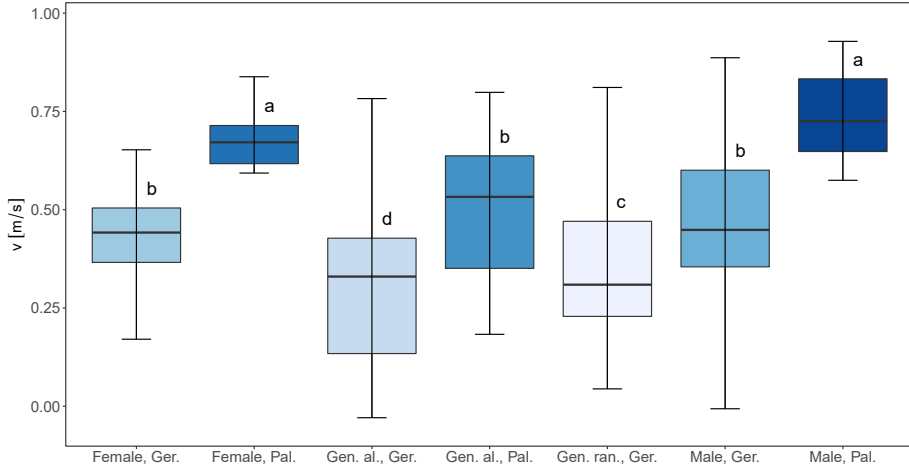


Figure 4.6: Boxplots of velocity for different group compositions at MEH in Germany and AAU Palestine in the interval  $[1.25, 1.35] \text{ m}^{-1}$ . The equal letters above the boxes indicate equality for the mean values of velocity between the corresponding group compositions [71].

to analyzing individuals within the fundamental diagram. Therefore, this section addresses the research question concerning the potential influence of gender and age on individual speed-distance functions at high densities.

First, another limitation of the data set is presented. The visual representations in Figure 4.7 demonstrate the relationship between headway and individual speed for a particular main ID of an individual of four pedestrians. The data presented in the diagrams show specific patterns or regimes. However, whether these regimes can be categorized as two or three separate regimes needs to be clarified. The complexity of pedestrian interactions and the need for a detailed approach to analyzing pedestrian behavior is highlighted by identifying different regimes within the data. Nevertheless, the free-flow branch starts clearly at a headway of about  $h \approx 1.5 \text{ m}$ . The beginning of the area selected for the free speed is supported by previous research by Ziemer [22] and Cao [17], who also focused on studying younger age groups. Their findings are consistent with the observations in this study and provide additional support for limitations in the following analysis. This work considers the branch where the speed is affected by neighboring pedestrians. Accordingly, the effect on the free speed is not analyzed. So, in particular, only the cases in which the headway for each individual  $i$  is less than 1.5 meters are relevant. To illustrate that the linear progression for

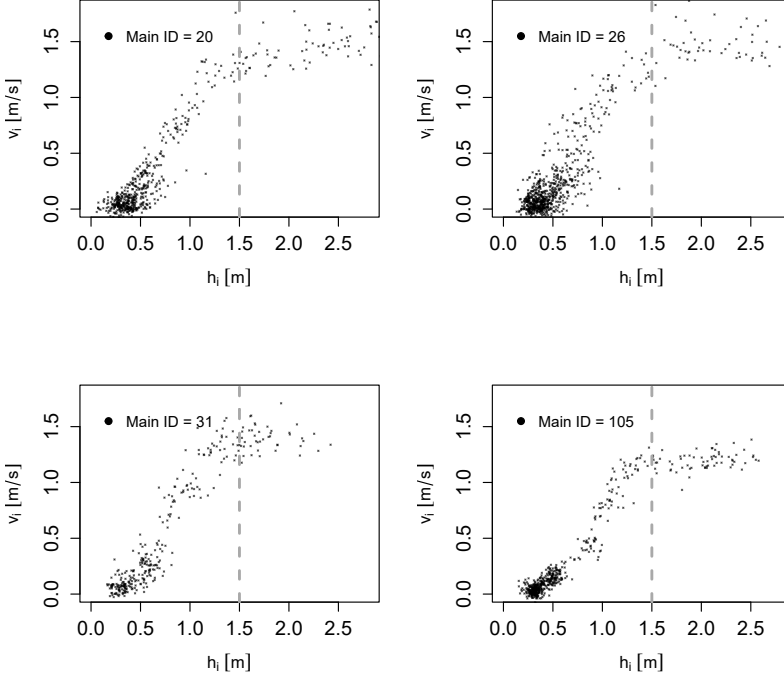


Figure 4.7: Headway vs. individual speed diagrams for four different main IDs to illustrate the different linear sections and to show that the free flow area starts at  $h_i \approx 1.5$  m.

$h < 1.5$  m does not only occur for selected individuals, for further pedestrians in the experiment at GBS and also for the single-file experiment at MEH, further diagrams for the course can be found in Appendix 5.2, in sections A.2 and A.3.

It is now possible to apply a simple linear regression analysis. To do this, several assumptions are checked for validation. These assumptions are checked for simple linear regression for all individuals in both single-file experiments. If they are valid, the analysis focuses on the regression analysis of each individual  $i$  in detail. First, there is no need to check for hidden relationships between variables because only one independent and one dependent variable are considered. The second step checked whether the dependent variable follows a normal distribution. This was found to be the case. As explained before, only the values for  $h < 1.5$  m are considered. Therefore, a straight line could describe the distribution of data points for the

#### 4. Data Analysis and Results

---

relationship between the independent variable *distance* and the dependent variable *speed*. For the two single-file experiments at GBS and MEH, the p-value is close to zero. This indicates that the model is a good fit for the data.

The next Listing 4.1 illustrates the output table of the simple linear regression analysis for all individuals for the single-file experiment at GBS in Germany. The first row presents the model equation. Then, the model residuals are summarized. The coefficients show the estimated values in the first column (Estimate). In this case, the intercept value is  $-0.27$ , and the estimated effect of headway on speed is  $0.96$ . The second column shows the standard error of the estimates (Std. Error). Then, the t-statistic and p-value indicate whether a relationship exists between the dependent and independent variables. This is the case because  $p\text{-value} < 0.05$ . The following rows are the diagnostics of the model. In this case, the p-value is essential to decide whether the model fits the data well. This is also the case, given that the p-value is less than  $0.05$ .

```
1 Call:
2 lm(formula = indVelocity ~ Headway, data =
   DataEveryXFramesNEWHeadway15)
3
4 Residuals:
5      Min       1Q   Median       3Q      Max
6 -1.17411 -0.09998 -0.01280  0.08186  1.31766
7
8 Coefficients:
9              Estimate Std. Error t value Pr(>|t|)
10 (Intercept) -0.269732   0.001677  -160.8   <2e-16 ***
11 Headway      0.960507   0.002904   330.8   <2e-16 ***
12 ---
13 Signif. codes:  0 "****" 0.001 "***" 0.01 "**" 0.05 "." 0.1 " " 1
14
15 Residual standard error: 0.165 on 49686 degrees of freedom
16 Multiple R-squared:  0.6877, Adjusted R-squared:  0.6877
17 F-statistic: 1.094e+05 on 1 and 49686 DF, p-value: < 2.2e-16
```

Listing 4.1: Output table: Simple linear regression for all individuals at GBS in Germany.

Based on the formulas in subsection 3.2.1 and on the values in the column Estimate in the output table, the following formula with estimated values results for



all individuals in both single-file experiments:

$$\text{GBS: } \hat{y}_m = -0.27 + 0.96 \cdot h_m , \quad (4.1)$$

$$\text{MEH: } \hat{y}_m = -0.18 + 0.66 \cdot (d_V)_m , \quad (4.2)$$

with  $m = 1, \dots, n$  and  $n$  is the number of all observations over all pedestrians  $N$ . These two equations are illustrated in Figure 4.8. When all individuals are considered,

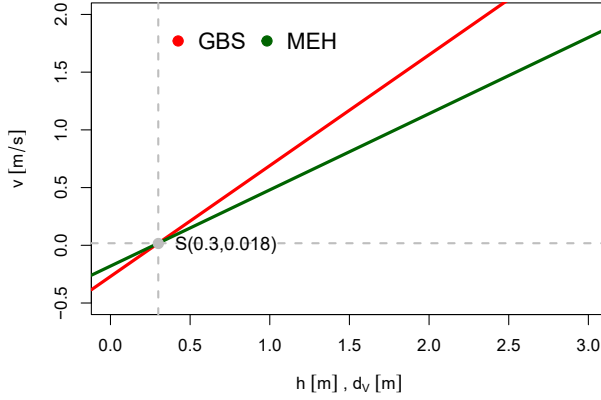


Figure 4.8: Comparison between the regression lines based on Equation 4.1 at GBS and Equation 4.2 at MEH for all individuals.

it can be observed that in the comparison between the experiments at GBS and MEH, from a speed of  $0.018 \text{ m}^{-1}$  and a headway of  $0.3 \text{ m}$ , the individuals are faster in the experiment at GBS than in the experiment at MEH. Below this threshold, the opposite is true. The individuals in the experiment at MEH seem to achieve even higher speeds at shorter distances.

Considering each individual and the given context, the individual estimation equation for the Equation 3.11 is as follows for the distance  $h_k$ :

$$\hat{v}_{k_i} = \hat{\beta}_{0_i} + \hat{\beta}_{1_i} \cdot h_{k_i} , \quad (4.3)$$

whereby  $\hat{\beta}_{0_i}$  and  $\hat{\beta}_{1_i}$  are the estimated values that minimize fitting error. The

#### 4. Data Analysis and Results

Equation 3.11 is the following in terms of the Voronoi distance ( $d_V$ ) for each individual:

$$\hat{v}_{k_i} = \hat{\beta}_{0_i} + \hat{\beta}_{1_i} \cdot (d_V)_{k_i} . \quad (4.4)$$

By transforming the formula for  $\hat{v}_{k_i} = 0$ , the minimum distance for each individual is obtained:

$$d_{min_i} = -\frac{\hat{\beta}_{0_i}}{\hat{\beta}_{1_i}} . \quad (4.5)$$

Finally, the fourth assumption, the goodness of fit of the simple linear regression model, i.e., homoscedasticity, was tested. For this purpose, the general model was considered again. This model takes into account all individuals. The assumption of homoscedasticity is illustrated in the following Figure 4.9. An example of this is the single-file experiment at GBS. In the first step, the red lines, which represent

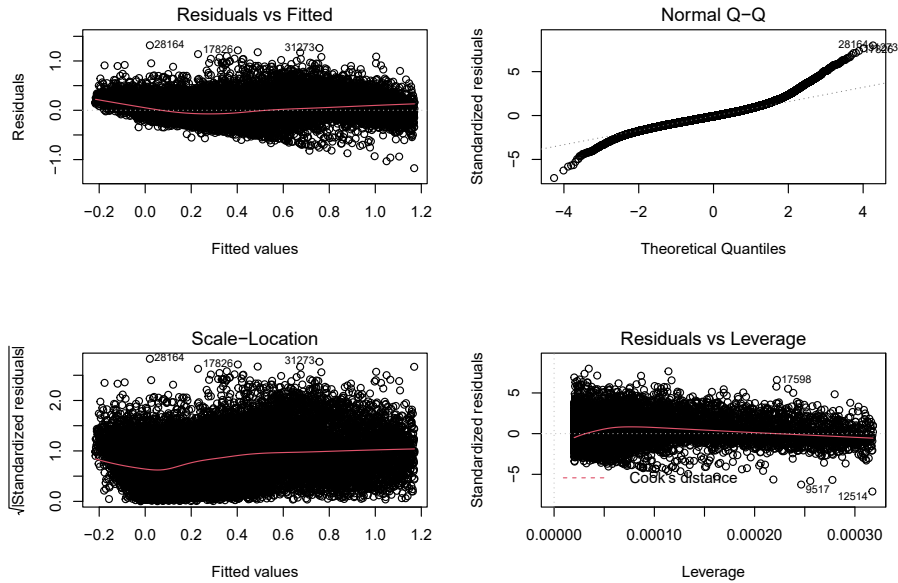


Figure 4.9: The assumption of homoscedasticity for the single-file experiment at GBS in Germany.

the residuals, are considered. In principle, they should be horizontal and centered around the zero point. This means that there are no outliers or biases in the data. Most of the time, this is the case, so linear regression is valid. Secondly, the Normal

Q-Q plot in the top right-hand corner is essential. It shows the real model residuals and the theoretical residuals from a perfect model. In the Normal Q-Q plot in the top right, we can see that the real residuals from our model form an almost perfectly one-to-one line with the theoretical residuals from a perfect model. A perfect model is one in which there is a nearly perfect one-to-one line between the theoretical and real residuals. So, based on these residuals, the model satisfies the assumption of homoscedasticity.

The following two subsections present in more detail the results of the simple linear regression analysis for the single-file experiments at the GBS and MEH based on each individual and Equation 4.3, Equation 4.4, and Equation 4.5.

##### 4.2.1 Differences by Gender and Age

The relationship between minimum distances and age and between minimum distances and gender of the students from the experiment at GBS is shown in Figure 4.10. On the left, it can be seen that older students tend to have larger minimum

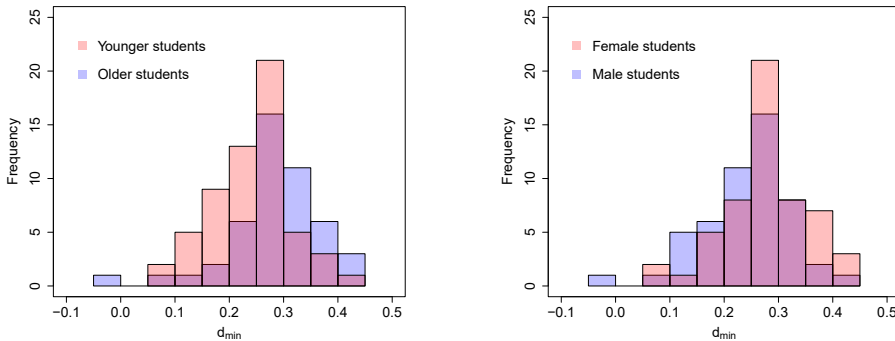


Figure 4.10: The left figure shows the minimum distances for younger and older students, and the right one shows minimum distances for male and female students from the experiment at GBS at  $\hat{v}_{k_i} = 0$ .

distances when their speed is equal to zero,  $\hat{v}_{k_i} = 0$ . The mean values for the minimum distances are  $\mu_{old} = 0.28 \pm 0.01$  for the older students and  $\mu_{young} = 0.24 \pm 0.01$  for the younger students, with the standard errors indicating the variability of these measurements. Furthermore, on the right, a comparison between male and female students shows that female students generally have slightly greater minimum distances than male students. The mean minimum distance for male students is

#### 4. Data Analysis and Results

$\mu_{male} = 0.25 \pm 0.01$ , while it is  $\mu_{female} = 0.28 \pm 0.01$  for female students. Despite this difference, it is important to note that the difference between genders is less pronounced.

In the context of this study, the slope  $\hat{\beta}_{1_i}$  of the regression line plays a crucial role in understanding the stimulus-response mechanism among students. This slope can be interpreted as the reaction time required for acceleration and braking in various situations. A comparison can be made between younger and older students

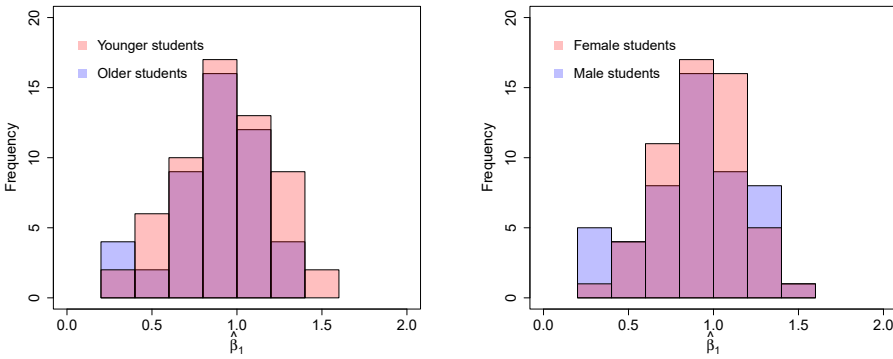


Figure 4.11: The left-hand figure shows the distribution of  $\hat{\beta}_1$ , reaction time and acceleration/braking ability, for younger and older students, and the right-hand figure the distribution for male and female students.

and between male and female students (see Figure 4.11). These comparisons show only minimal differences. The left-hand figure shows that the mean values and the standard errors for the older students are  $\mu_{old} = 0.87 \pm 0.04$ , while for the younger students, they are  $\mu_{young} = 0.94 \pm 0.04$ . Similarly, the right-hand one shows that the mean is  $\mu_{male} = 0.89 \pm 0.04$  for male students and  $\mu_{female} = 0.93 \pm 0.03$  for female students. These results are an indication that there is a slight variation between the groups. Figure 4.11 also shows that the values for  $\hat{\beta}_1$  are all larger than 0. This indicates a positive correlation between the dependent variable  $v$  and the distance.

A t-test was performed to validate these conclusions further. First, normal Q-Q plots were used to confirm that the distributions were normally distributed. The t-test was then performed with the null hypothesis  $H_0$ , which showed no significant difference between the groups. If the null hypothesis were to be accepted in all cases, this means that the distributions for younger and older students, as well as for female and male students, are equal in terms of minimum distances or  $\hat{\beta}_1$ . The

#### 4. Data Analysis and Results

null hypothesis is only rejected when comparing younger and older students at the minimum distance, as indicated by a p-value of less than 0.05. This suggests that there is no similarity between these two groups concerning age. To conclude, while there may be some slight variations in certain comparisons between different groups of students, overall, there appears to be a high degree of consistency across age and gender groups when comparing the values of minimum distances or  $\hat{\beta}_1$ . Overall, while the initial comparisons may have shown minimal differences between younger and older students or between male and female students, there is still much to be investigated to understand the complexity of the factors that impact speed.

The following analysis investigates the scatter observed around the regression line in more detail. When the entire group is divided into subcategories based on age and gender, e.g., younger and older students and females and males, different patterns are seen in the headway vs. speed plots. It is evident that within each subgroup, there are variations in the range of scattering. There are headway vs. individual speed diagrams with low and high scattering. Looking more closely at

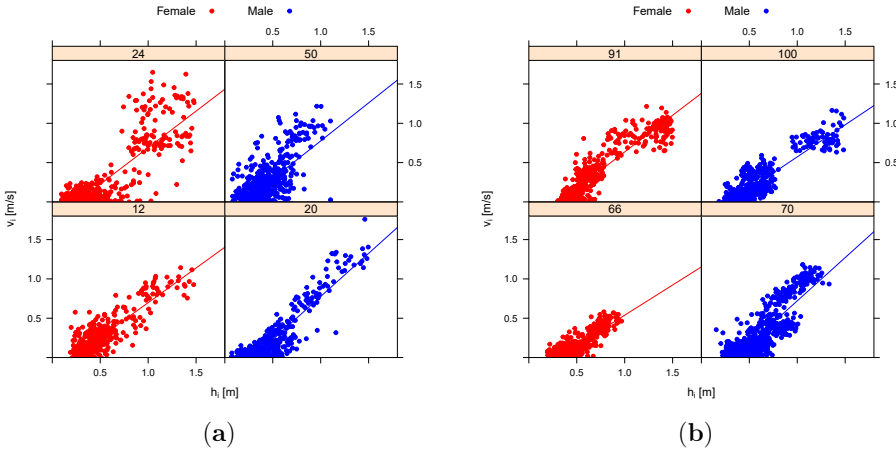


Figure 4.12: Diagrams of headway vs. individual speed with a regression line for representative younger students in the fifth grade in (a) and for older students in the eleventh grade in (b). These two groups are divided into male and female main IDs. The numbers in the orange box represent the different main IDs. High and low scatterings exist for each gender and age group [51].

Figure 4.12, it can be seen that for both the younger and the older students, four main representative IDs have been chosen to illustrate this phenomenon. Each point on the graph represents the measured values  $l = 1, \dots, n_i$  for one individual  $i$ , with

#### 4. Data Analysis and Results

the regression line plotted according to Equation 3.11 in different densities. The left side of the figure (a) focuses on the younger group of pedestrians, while the right (b) highlights the older group of students. In addition, males are indicated by blue dots and females by red dots. To investigate whether there is a difference in the scatter around the regression line based on age and gender, a study of the correlation between headway and individual speed is conducted (see Figure 4.13). Figure 4.12, indicate that there may be a higher scattering for younger students compared to older students. This hypothesis is supported by the histograms in Figure 4.13. Looking

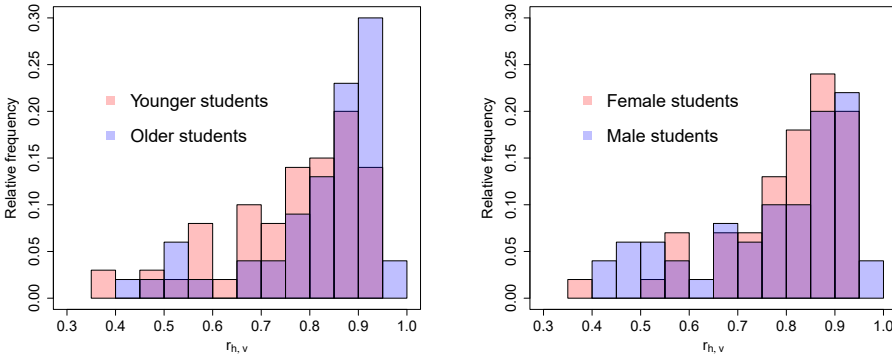


Figure 4.13: Distribution of the correlation coefficient  $r_{h,v}$  between headway and individual speed for younger and older students and female and male students.

more closely, the mean correlation coefficient  $\bar{r}_{h,v}$  for the younger groups is 0.77 and slightly higher for the older groups with 0.82. Upon dividing the data by gender, no significant difference was observed in the correlation coefficients between male and female students within both age groups. This is because the group correlation coefficient is low. Specifically, the correlation coefficient is 0.76 for young male students, while for young female students, it is 0.77. In contrast, the correlation coefficient for older students is 0.80 for males and 0.84 for females. Interestingly, when looking at all male and female students separately, the scatter appears slightly larger for male students than for female students (see Figure 4.13.)

In general, the linear regression diagrams plotting headway versus individual speed for different individuals suggest that for  $h < 1.5$  m, there is virtually no difference between gender and age. However, there may be a more significant variation among younger students. This assumption is supported by the histograms in Figure 4.13, which show the correlation coefficients of the individuals in the different groups, such

as the younger and older students. In addition, there is no significant difference between male and female students regarding scatter patterns. On closer investigation, a comparison of the minimum distances at  $\hat{v} = 0$  shows differences between younger and older students, with older students showing greater distances. These differences are less pronounced when male and female students are compared. Furthermore, an analysis of  $\hat{\beta}_1$  – reaction time and ability to accelerate and brake – shows virtually no differences between younger and older students and male and female students.

### 4.2.2 Differences by Gender of Neighboring Pedestrians

The following analysis focuses on the simple linear regression for the single-file experiment at MEH. The study mainly investigates the minimum distances  $d_{min_i}$  when the speed is equal to zero  $\hat{v}_{k_i} = 0$ , and the reaction time  $\hat{\beta}_{1_i}$  required for accelerating and braking. Four groups are considered: female, male, alternating genders, and gender random order. This allows comparison between two homogeneous groups (female and male), two heterogeneous groups (alternating genders and gender random order), and homogeneous and heterogeneous groups. The mean values and standard deviations for  $d_{min_i}$  and  $\hat{\beta}_{1_i}$  for the four different group compositions in Germany are presented in Table 4.3.

Table 4.3: The columns give an overview of the mean values and standard error for  $d_{min_i}$  and  $\hat{\beta}_{1_i}$  for the different group compositions: female, male, gender alternating, and gender random.

	Female	Male	Gender Alternating	Gender Random
$\overline{d_{min_i}} \pm \sigma$	$0.31 \pm 0.07$	$0.34 \pm 0.08$	$0.34 \pm 0.08$	$0.36 \pm 0.08$
$\overline{\hat{\beta}_{1_i}} \pm \sigma$	$0.96 \pm 0.22$	$0.95 \pm 0.23$	$0.94 \pm 0.18$	$0.90 \pm 0.21$

Interestingly, the values shown in the table are consistent across the different group compositions. The value is only slightly lower for females. The mean values and the standard errors for the minimum distances are  $\mu_{Female} = 0.31 \pm 0.07$  for the female group and  $\mu_{Male} = 0.34 \pm 0.08$  for the male group. Furthermore, for the heterogeneous groups, the mean values and the standard errors are  $\mu_{Gender\ alternating} = 0.34 \pm 0.08$  for the gender alternating group and  $\mu_{Gender\ random} = 0.36 \pm 0.08$  for the gender random order group. In comparison, the mean values and the standard errors for  $\hat{\beta}_{1_i}$  for the different groups are  $\mu_{Female} = 0.96 \pm 0.22$  for the female group, while it is  $\mu_{Male} = 0.95 \pm 0.23$  for the male group. In contrast, the values for the heterogeneous groups are  $\mu_{Gender\ alternating} = 0.94 \pm 0.18$  for the gender alternating

#### 4. Data Analysis and Results

---

group and  $\mu_{Gender\ random} = 0.90 \pm 0.21$  for the gender random order group. While the values may appear similar at first view, closer examination may show slight differences that require further investigation. It could be said that  $d_{min_i}$  is lower in the homogeneous groups and higher in the heterogeneous groups and that the values for  $\hat{\beta}_{1_i}$  in the groups are precisely the opposite, e.g., higher in the homogeneous groups and lower in the heterogeneous groups. A shorter distance seems to result in a longer reaction time, whereas a longer distance results in a slower one. Therefore, It could be assumed that at a shorter distance, the participants pay more attention to the person in front of them and react faster. However, it is also important to note that the differences are small and, therefore, less pronounced.

In addition to the data presented in Table 4.3, an analysis comparing the different values for  $d_{min_i}$  and  $\hat{\beta}_{1_i}$  was performed using a statistical method known as the Tukey HSD test as described in subsection 3.2.4. This test determined whether there were significant differences between the means of different group compositions. Each group composition was directly compared in pairs with the other compositions to ensure a more comprehensive analysis. The Tukey test takes into account that the sample sizes are approximately equal. If the  $p$ -value obtained from the Tukey test is greater than 0.05, there is no significant difference between the means of the group compositions being compared. This indicates equal groups. This study showed that for all four group compositions: female, male, gender alternating, and gender random order in Germany, the  $p$ -values for both values  $d_{min_i}$  and  $\hat{\beta}_{1_i}$  were consistently greater than 0.05. The results of the Tukey HSD test in R for  $d_{min_i}$  and  $\hat{\beta}_{1_i}$  are shown below in Listing 4.2. The last column, namely p adj, presents the p-value. For example, in the first row, the gender-alternating group is compared with the gender random-order group. The  $p$ -value for this comparison is approximately 0.528. All in all, this clearly shows that all the groups are equal in the paired comparisons. This suggests that across these different group compositions, there is no statistically significant difference in the means of the values  $d_{min_i}$  and  $\hat{\beta}_{1_i}$ . The equality observed in the results indicates that factors such as gender or the structure of different group compositions do not have a significant impact on the values  $d_{min_i}$  and  $\hat{\beta}_{1_i}$  in Germany.

```

1  Tukey multiple comparisons of means
2    95% family-wise confidence level
3 Fit: aov(formula = minHeadway ~ group, data = Test_all)
4 $group
5           diff      lwr      upr    p adj
6 Gen. al. -Female  0.026 -0.024  0.075  0.528
7 Gen. ran.-Female  0.050 -0.002  0.102  0.063

```



#### 4. Data Analysis and Results

```

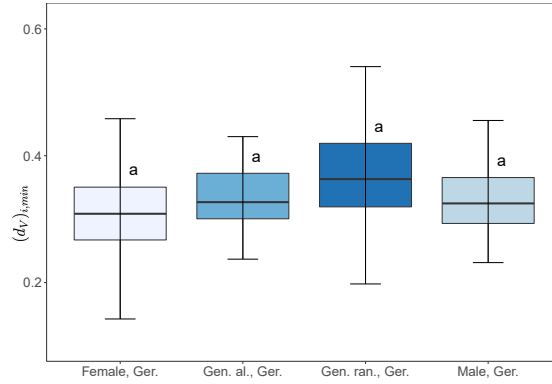
8 Male      -Female      0.022 -0.033 0.077 0.730
9 Gen. ran.-Gen. al.    0.024 -0.026 0.074 0.596
10 Male      -Gen. al.   -0.004 -0.058 0.050 0.997
11 Male      -Gen. ran.  -0.028 -0.084 0.028 0.555
12 Fit: aov(formula = beta1 ~ group, data = Test_all)
13 $group
14
15 Gen. al.   -Female      diff      lwr      upr      p adj
16 Gen. ran. -Female      -0.016 -0.148 0.115 0.988
17 Male      -Female      -0.012 -0.158 0.135 0.997
18 Gen. ran. -Gen. al.    -0.041 -0.175 0.093 0.855
19 Male      -Gen. al.     0.005 -0.138 0.148 0.999
20 Male      -Gen. ran.    0.046 -0.102 0.194 0.851

```

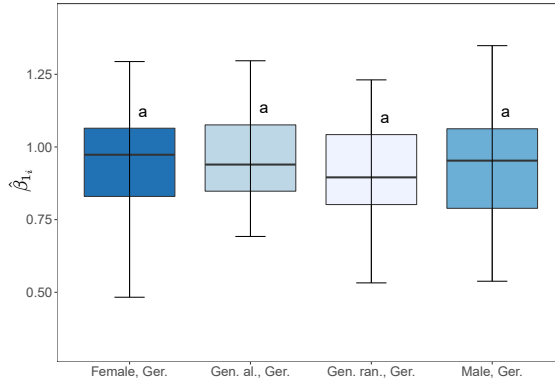
Listing 4.2: Results of the Tukey HSD test in R for the values  $d_{min_i}$  and  $\hat{\beta}_{1_i}$  at the MEH in Germany. The last column presents the p-value.

The boxplots in Figure 4.14 provide a visual representation of the results of the Tukey test for  $d_{min_i}$  in Figure 4.14 (a) and  $\hat{\beta}_{1_i}$  in Figure 4.14 (b) across different group compositions in Germany, including female, male, gender alternating and random gender order.

Presenting these results visually improves understanding of the behavior of the values  $d_{min_i}$  and  $\hat{\beta}_{1_i}$  under a variety of conditions. The letters displayed above each boxplot serve as indicators of equality or inequality between the group compositions. If the same letter appears above more than one box, it means that these groups are statistically equal in terms of the values  $d_{min_i}$  and  $\hat{\beta}_{1_i}$ . Boxplots also provide a comprehensive overview of data distribution within each group composition. Furthermore, the figures show that the four groups are equal in terms of  $d_{min_i}$  and  $\hat{\beta}_{1_i}$ , as indicated by the similarity of the letters at the top of the boxplots. The consistency of the results across the group compositions suggests that there are minimal differences in the values  $d_{min_i}$  and  $\hat{\beta}_{1_i}$  between the groups female, male, gender alternating and gender random order in Germany. This equality implies that factors such as gender or the groups' composition do not significantly affect these variables when analyzed at high densities. It should be noted that while there may be slight differences between group compositions, these differences are not significant when considering the overall distribution of the values  $d_{min_i}$  and  $\hat{\beta}_{1_i}$ . The minor differences may be due to random variation rather than meaningful differences between groups. So, it is confirmed that there are no differences between the group compositions at high density. Consequently, the hypothesis that, at high densities, the distance



(a)



(b)

Figure 4.14: Boxplots based on a Tukey test for  $(d_V)_{i,min}$  in (a) and for  $\hat{\beta}_{1_i}$  in (b) for different group compositions of female, male, gender alternating, and gender random order in Germany. The same letters above the boxes indicate equality for the mean values between the corresponding groups [71].

between individuals depends on the gender of the neighboring pedestrians cannot be confirmed.

### 4.3 Speed-Distance Relation, Human Factors, and Individuality

The following study investigates the effect of independent variables like distance, gender of the previous pedestrian, and various individual factors such as gender, height, weight, shoulder width, and age on the dependent speed variable. By using multiple linear regression analysis and considering a range of independent variables, a robust model for predicting speed can be developed. The speed of each individual is denoted as  $v_m$ , where  $m = 1, \dots, n$  and  $n$  is the number of all observations of all individuals. Thus, the focus of the analysis shifts from all observations of one individual to all observations of all individuals. This is essential to obtain the full range of variability in speed. At the same time, a best-fit model aims to construct a model that minimizes the number of relevant independent variables included while still including all relevant factors that affect speed.

#### 4.3.1 Human Factors and Individuality

Firstly, multiple linear regression analysis is considered in the single-file experiment at GBS. Here, several independent variables are introduced that are expected to significantly impact an individual's speed. These individual characteristics measured include the headway  $h$ , *age*, *height*, and *gender*. In addition, the variable *alloence* is included to consider any other unknown individual effects that may influence speed, such as motivation, attention, or excitement. By including this variable, the aim is to have a model that includes all of the factors that affect speed. In addition, it is important to ensure that the independent variables chosen in the model are not to be highly correlated with each other. This ensures that each variable contributes unique information to the speed prediction and avoids problems such as multicollinearity. Ideally, these variables should also have a clear theoretical or empirical relationship to the speed of individuals to improve the model's prediction. A useful understanding of how these factors interact to influence the speed can be gained by carefully constructing the model and selecting relevant variables.

Concerning the relationship between age and height, it is important to note that

there is a correlation. For younger and older students, height can vary significantly. This is because students' bodies are still growing. The strong correlation between the variables *height* and *age* with ( $r_{height,age} = 0.89$ ) suggests that it may not be necessary to include both of them in a statistical model. Depending on the research question and the data quality, including only one of these variables may be sufficient. In this case, *height* is chosen over *age* because age is measured as binary (0 for younger, 1 for older students) while height is categorized into five levels. In contrast, there is no or weak correlation between the other independent variables in the data. This means ( $r_{w,y} < 0.29$ ). Thus, the assumption of observations' independence is fulfilled for the multiple linear regression analysis. Normality and linearity are also assumed to hold. The assumption of homoscedasticity will be checked at a later stage. In the following multiple linear models, all variables are considered without units. The first full model, based on the general multiple linear regression model in Equation 3.12 and including all measured individual characteristics, is as follows:

$$\text{Model I}_{\text{GBS}} : v_m = \beta_0 + \beta_1 \cdot h_m + \beta_2 \cdot \text{gender}_m + \beta_3 \cdot \text{height}_m + \epsilon_m . \quad (4.6)$$

This model aims to explain the relationship between the dependent variable and all relevant independent variables headway *h*, *gender*, and *height*. However, it is essential to consider whether including these variables in the model is necessary or whether a simpler model with fewer variables might be sufficient. When exploring the relationship between individual characteristics and the fundamental diagram, this model allows us to analyze which of the two individual characteristics, height or gender, has a more substantial effect on the fundamental diagram. Furthermore, in the context of the analysis, it is important to distinguish between the following two research questions:

- 1 Which individual characteristic, height or gender, affects the fundamental diagram more?
- 2 How strongly do individual characteristics affect the fundamental diagram?

To respond to the second research question, it is necessary to consider additional variables that would potentially improve Model I and affect the dependent variable  $v_m$ . Building on the initial model focused on height, headway, and gender, a more extended model can be developed to consider all other unknown individual effects. One way of doing this is to introduce a new variable called *alloence*. This variable, which represents additional factors, is not explicitly measured in the dataset. By

including both the known variables (headway, height, and gender) and the unknown variable (represented by *alloence*) in the model, it is possible to explore how these hidden influences are related to variations in speed  $v_m$ . The extended model with the new variable *alloence* is shown in the following equation:

$$\begin{aligned} \text{Model II}_{\text{GBS}} : v_m = & \beta_0 + \beta_1 \cdot h_m + \beta_2 \cdot \text{gender}_m + \beta_3 \cdot \text{height}_m \\ & + \sum_{i=4}^{N+3} \beta_{4_i} \cdot \text{alloence}_{m_i} + \epsilon_m , \end{aligned} \quad (4.7)$$

where  $\text{alloence}_{m_i} = 1$  for all  $m$  belonging to individual  $i$  and 0 for all other  $m$ .  $\beta_{4_i}$  is an individual coefficient across all measurement points for each student.

The next part of the analysis considers which variables from Equation 4.6 and Equation 4.7 should be included in the model. This is done to create the most optimal model with the fewest variables. One way of doing this is to evaluate the model using Akaike's Information Criterion (AIC) [72]. This involves systematically investigating whether removing independent variables in Equation 4.7 can help to improve the model. By carefully considering the impact of each variable on the overall performance of the model, more accurate and efficient models can be created that are designed to extract the most relevant information from the data without unnecessary complexity. Therefore, using AIC in model evaluation provides essential information about which variables are most influential in predicting outcomes. A lower AIC value is an indication of a better model fit. The process continues until further reduction is no longer an advantage. It is important to note that AIC does not provide an absolute criterion for selecting a best-fit model. Applying the AIC procedure to Model II, it is evident that *gender* can be removed from the model without reducing its performance. However, the evaluation of Model I using the same approach shows that gender should remain in the model because of its significant influence on the result. This comparison indicates that the model can only be reduced to Model III by considering unmeasured individual characteristics. Therefore, the variable *alloence*, which describes the unknown individual effects, significantly contributes to the model. This result also shows that Model I only includes some relevant individual characteristics that describe the influences on the speed.

$$\begin{aligned} \text{Model III}_{\text{GBS}} : v_m = & \beta_0 + \beta_1 \cdot h_m + \beta_2 \cdot \text{height}_m \\ & + \sum_{i=3}^{N+2} \beta_{3_i} \cdot \text{alloence}_{m_i} + \epsilon_m . \end{aligned} \quad (4.8)$$

The Akaike's information criterion results are consistent with the conclusions drawn from the analysis of minimum distance, reaction time, and the correlation between headway and individual speed discussed in subsection 4.2.1 above. The difference between younger and older students is more pronounced than that between male and female students, where the difference is minimal. In Model II, where gender is found to have a minimal effect when other individual characteristics are considered, it is reasonable to exclude it from the model to reduce its complexity. However, in Model I, where other personal factors are not considered, gender cannot be ignored because of its low influence. It must be considered for a comprehensive analysis.

Continuing with Model III, the estimated regression coefficients based on the multiple linear regression analysis are shown in Equation 4.9:

$$\hat{v}_m = 0.23 + 0.98 \cdot h_m - 0.34 \cdot height_m + \sum_{i=3}^{N+2} \beta_{3_i} \cdot alloence_{m_i} . \quad (4.9)$$

By examining the estimated regression coefficients in Equation 4.9, it is possible to gain insight into the effect of each variable specifically on the outcomes investigated. The values  $\hat{\beta}_0 = 0.23$ ,  $\hat{\beta}_1 = 0.98$ , and  $\hat{\beta}_2 = -0.34$  show the importance of the predictor variables in influencing speed. Each predictor variable has a significant effect on speed changes. The headway contributes more to the speed than the height. This is reflected not only in the values but also in the sign of the estimated regression coefficients. The values for  $\hat{\beta}_{3_i}$  are illustrated by the distribution in Figure 4.15. The variable *alloence* can have either a positive or a negative effect on the individual speed. The regression coefficient  $\hat{\beta}_{3_i}$  associated with each individual determines the strength of the effect on individual speed. It is important to note that most effects are statistically significant with  $p < 0.05$ , although some values appear weak. A notable finding from this analysis is that 83.3% of the observed effects are considered statistically significant. In addition, a study was carried out to see if there were any differences in the results or the models when the data from the younger and older students were examined separately. Surprisingly, no significant differences were found between the two groups.

The assumption of homoscedasticity is now checked by using the scatter of the residuals in model III see Equation 4.8. The values are shown in Table 4.4.

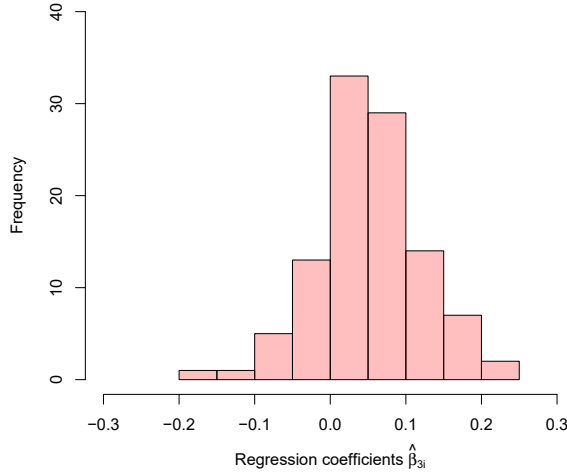


Figure 4.15: Distribution of the regression coefficients  $\hat{\beta}_{3i}$  of all other unknown individual effects for each main ID [51].

Table 4.4: Summary of the scatter of the residuals for model III: the minimum value, the first quartile, the median, the third quartile, and the maximum value.

Min	1Q	Median	3Q	Max
-1.10	-0.09	-0.01	0.08	1.33

Thus, the model is consistent with the assumption of homogeneity as the residuals are approximately centered around zero and have a similar spread on either side.

Next, Model III (see Equation 4.8) is used to analyze the effect of the independent variables, *headway*  $h$ , *height*, and *alloence* on the dependent variable *speed*. Therefore, an ANOVA table is considered. ANOVA, which stands for Analysis of Variance, is a statistical method that can be used to determine if there are significant differences in the mean values of independent variables. This test allows to analyze whether the dependent variable, such as speed, is affected by various independent variables mentioned before. ANOVA aims to explain as much of the variance in the dependent variable as possible using the independent variables. If there is much variance within data groups, there is more chance that a sample selected from the data will have different means. By understanding how the mean of each independent

#### 4. Data Analysis and Results

variable differs from the others, it is possible to understand which are related to the dependent variable. In an ANOVA table, p-values are essential to determine whether there are statistically significant differences between groups.

```

1 Analysis of Variance Table
2 Response: indVelocity
3
4      Df    Sum Sq Mean Sq    F value    Pr(>F)
5 Headway      1 2978.91 2978.91 127538.492 < 2.2e-16 ***
6 Z            1   28.62  28.62   1225.120 < 2.2e-16 ***
7 MainID      105  166.20   1.58    67.769 < 2.2e-16 ***
8 Residuals   49580 1158.04   0.02
9 ---
10 Signif. codes:  0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1

```

Listing 4.3: Table of the Analysis of Variance for modell III at GBS.

Based on the ANOVA (see Listing 4.3), it becomes clear that all variables considered in Model III do affect individual speed. Any effect is significant because  $p < 0.05$ .

Figure 4.16 illustrates the various effects of the independent variables on the individual speed based on the ANOVA table in a pie chart. This is done based on the column named Sum Sq in Listing 4.3. The headway has the most significant effect

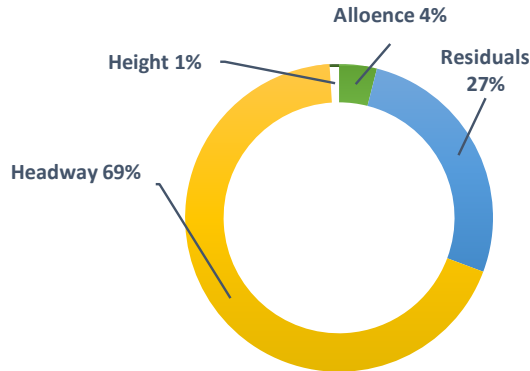


Figure 4.16: Different effects of the independent variables on individual speed based on the ANOVA table in Listing 4.3.

with 69%. This is followed by all other unknown individual effects with 4%. The height has the lowest effect at 1 %. The same analysis for Model II (see Equation 4.7) shows that gender has a more minor effect than height. However, this effect of the variable *height* is so small, around 1%, that it can be neglected. The final model



without height and gender is presented in Equation 4.10.

$$\text{Model IV}_{\text{GBS}} : v_m = \beta_0 + \beta_1 \cdot h_m + \sum_{i=2}^{N+1} \beta_{2i} \cdot \text{alloence}_{m_i} + \epsilon_m . \quad (4.10)$$

The residuals are further analyzed for first-order autocorrelation using the Durbin-Watson test. This test allows to investigate whether significant effects on speed are being ignored. The null hypothesis in this test, denoted as  $H_0 : \rho_1 = 0$ , where  $\rho_1$  is the theoretical autocorrelation coefficient, suggests that the error term is not autocorrelated. If the null hypothesis is rejected, it can also be decided whether the data has positive or negative autocorrelation. When the null hypothesis is rejected, it indicates that the error term does not fulfill the standard assumptions of a multiple linear regression model. In this case, the Durbin-Watson statistic shows a value of 0.7077, leading to the rejection of the null hypothesis. In addition, there is positive autocorrelation due to the low value of  $DW$  and the limits of the critical values of the Durbin-Watson statistic. This finding highlights that headway has a significant effect on individual speed. However, it also implies that other factors besides height and unknown individual effects play a role in determining speed. These additional factors must be addressed when analyzing the speed results. For example, distinctions between acceleration and deceleration phases and consideration of stimulus-response mechanisms in reaction time and acceleration capabilities should be carefully modeled. Furthermore, this finding highlights the importance of modeling locomotion in steps to accurately capture all relevant factors influencing speed outcomes. In the next step, further human factors are included in the model.

#### 4.3.2 Human Factors Extension

In the previous subsection 4.3.1, the multiple linear regression analysis results indicated that headway has the most significant effect on speed. In contrast, other human factors, such as gender, have minimal effects or could be ignored. In addition, subsection 4.3.1 provides further insight into the methodology and structure of the model used in the analysis. The focus now shifts to whether incorporating additional human factors improves model sensitivity. This expanded model should provide a more detailed understanding of how human factors contribute to speed outcomes. Hidden patterns and relationships that may have been missed in previous analyses can be uncovered by incorporating diverse variables. This is done by including additional variables such as the weight, age, exact height, and gender of the pedestrian in front.

#### 4. Data Analysis and Results

Consequently, the new model examines speed as a function of Voronoi distance, height, gender, age, weight, and gender of the previous pedestrian. The variable "gender.prev" is used for this purpose, while individual effects such as motivation, attention, or excitement mentioned in the previous section are represented by the variable *alloence*.

It is expected that strong correlations between certain human factors could exist. Therefore, a correlation plot is considered. In Figure 4.17, positive correlations between the variables are displayed in blue, e.g., weight and shoulder width, and negative correlations in red color. The stronger the correlation between the variables,

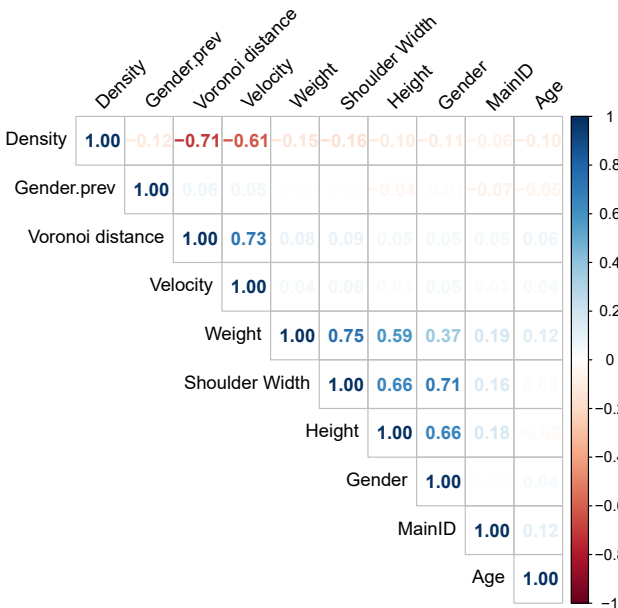


Figure 4.17: Correlogram: Correlation between certain factors, like weight, height, and gender, in a correlation plot. Positive correlations are shown in blue, and negative ones in red. The intensity of the colors increases as the correlation between the variables increases.

the higher the intensity of the colors. The correlation coefficients and corresponding colors are shown in the legend to the right of the correlogram. Based on the measurement of the correlation between the factors considered, it can be concluded that there are apparent dependencies, such as positive and strong correlations between gender and height ( $r = 0.66$ ), gender and shoulder width ( $r = 0.71$ ), or weight and

shoulder width ( $r = 0.75$ ). Due to the correlation, shoulder width is neglected in the following. Furthermore, in this analysis, one model is considered for all individuals. So all four groups, female, male, gender random order, and gender alternating are included in one model. This is because the results in section 4.1 and subsection 4.2.2 showed no significant differences between the groups.

First, the model evaluation using Akaike's Information Criterion (AIC) as explained in subsection 4.3.1 is applied to the following model:

$$\begin{aligned} \text{Model I}_{\text{MEH}} : v_m = & \beta_0 + \beta_1 \cdot (d_V)_m + \beta_2 \cdot \text{gender}_m + \beta_3 \cdot \text{height}_m + \beta_4 \cdot \text{age}_m \\ & + \beta_5 \cdot \text{weight}_m + \beta_6 \cdot \text{gender.prev}_m + \sum_{i=7}^{N+6} \beta_{7_i} \cdot \text{alloence}_{m_i} \\ & + \epsilon_m , \end{aligned} \quad (4.11)$$

where  $m = 1, \dots, n$  and  $n$  is the number of all observations of all individuals.

Step by step, an analysis was conducted to determine which factors to include to create the best possible model with the fewest variables without compromising accuracy. The Akaike Information Criterion (AIC) procedure guided this process. This method determined that the variable weight could be omitted from the model without significantly affecting its prediction performance. Thus, the final model for all individuals and all densities that resulted from this process is the following:

$$\begin{aligned} \text{Model II}_{\text{MEH}} : v_m = & \beta_0 + \beta_1 \cdot (d_V)_m + \beta_2 \cdot \text{gender}_m + \beta_3 \cdot \text{height}_m + \beta_4 \cdot \text{age}_m \\ & + \beta_5 \cdot \text{gender.prev}_m + \sum_{i=6}^{N+5} \beta_{6_i} \cdot \text{alloence}_{m_i} + \epsilon_m . \end{aligned} \quad (4.12)$$

Now model II in Equation 4.12 is applied to investigate the effect of the independent variables Voronoi distance  $d_V$ , gender, height, age, gender.prev, and alloence on the dependent variable speed,  $v_m$ . So again, an ANOVA table is considered. The ANOVA table in Listing 4.4 shows that the  $p$ -values for all variables are less than 0.05. This means that all variables considered in model II (see Equation 4.12) affect the individual speed. Again, the headway affects the velocity most, followed by all other unknown individual effects.

#### 4. Data Analysis and Results

```

1 Analysis of Variance Table
2
3 Response: indVelocity
4
5      Df Sum Sq Mean Sq  F value    Pr(>F)
6 Headway      1 4248.7   4248.7 97637.310 < 2.2e-16 ***
7 GenderVorderIDm1f0 1    5.0     5.0  114.614 < 2.2e-16 ***
8 Age          1    3.0     3.0   68.614 < 2.2e-16 ***
9 HeightM      1   11.9    11.9  274.533 < 2.2e-16 ***
10 GenderAmBfCdivers 1    1.7     1.7   38.875 4.527e-10 ***
11 MainID      71  442.8     6.2  143.334 < 2.2e-16 ***
12 Residuals   203445 8853.0    0.0
13 ---
14 Signif. codes:  0 "****" 0.001 "***" 0.01 "**" 0.05 "." 0.1 " " 1

```

Listing 4.4: Table of the Analysis of Variance for modell II at MEH.

Based on the column Sum Sq, the table clearly shows the variables *gender.prev*, *age* and *gender* have very little influence. Up to now, the models are designed for the whole density. Now, a distinction is made between low and high density. For this purpose,  $N \leq 20$  ( $\rho_{gl} \leq 0.75$ ) was selected for the low density and  $N > 20$  ( $\rho_{gl} > 0.75$ ) for the high density. Considering the results of the previous sections, low density is the region in which there is no systematic difference between equality and inequality between the mean values of the velocities in the different group compositions. For these two density separations, the AIC procedure reduces the model in Equation 4.11 again. The variables *age*, *weight*, *gender*, and *height* are omitted, and the following model results for low and high density:

$$\begin{aligned}
 \text{Model II}_{\text{MEH(low/high)}} : v_m &= \beta_0 + \beta_1 \cdot (d_V)_m + \beta_2 \cdot \text{gender.prev}_m \\
 &+ \sum_{i=3}^{N+2} \beta_{3_i} \cdot \text{alloence}_{m_i} + \epsilon_m .
 \end{aligned} \tag{4.13}$$

These cases provide the following two ANOVA tables in Listing 4.5 and Listing 4.6:

```

1 Analysis of Variance Table
2 Response: indVelocity
3
4      Df Sum Sq Mean Sq  F value    Pr(>F)
5 Headway      1 2422.11  2422.11 40518.391 < 2.2e-16 ***
6 GenderVorderIDm1f0 1   27.16   27.16  454.428 < 2.2e-16 ***
7 MainID      71  293.88    4.14   69.242 < 2.2e-16 ***
8 Residuals   41910 2505.30    0.06
9 Signif. codes:  0 "****" 0.001 "***" 0.01 "**" 0.05 "." 0.1 " " 1

```

Listing 4.5: Table of the Analysis of Variance for modell II at MEH for low density.

#### 4. Data Analysis and Results

```

1 Analysis of Variance Table
2
3 Response: indVelocity
4
5      Df Sum Sq Mean Sq  F value    Pr(>F)
6 Headway      1   364.0    363.96 9955.562 < 2.2e-16 ***
7 GenderVorderIDmif0      1     6.5     6.47  176.843 < 2.2e-16 ***
8 MainID      74   129.3     1.75   47.807 < 2.2e-16 ***
9 Residuals  179620  6566.7     0.04
10 ---
11 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Listing 4.6: Table of the Analysis of Variance for modell II at MEH for high density.

It can be seen here that the gender of the person in front of individual  $i$  has a small effect overall, but at higher densities, it has an even smaller effect on the speed than it does at a low density. In the next step, these results from the Sum Sq column of the ANOVA table are once again presented in pie charts for better visualization of low density in Figure 4.18 (a) and high density in Figure 4.18 (b). The effect of

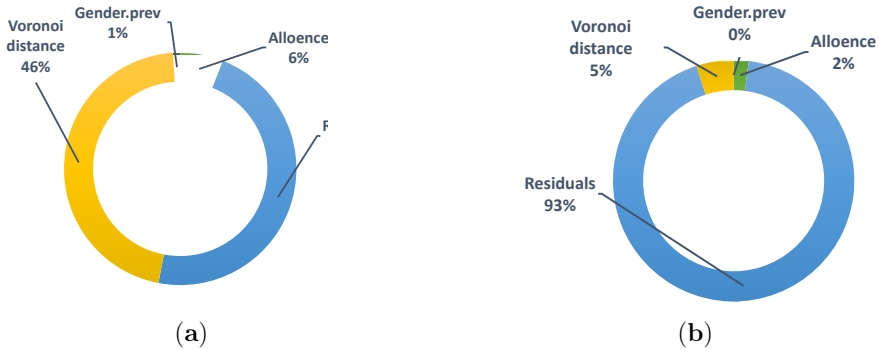


Figure 4.18: Effects on the speed based on the ANOVA table in pie charts for a low density,  $\rho_{gl} \leq 0.75$  in (a) and a high density,  $\rho_{gl} > 0.75$  in (b).

the gender of the pedestrian in front of the observer is minimal, accounting for less than 1%, and can be considered insignificant. In addition, the model is more suitable for low densities than high densities, as speed is more influenced by the Voronoi distance at lower densities. At higher densities, the effect of the Voronoi distance is reduced due to slight variations caused by head movements within millimeters. Regarding the third hypothesis that the inclusion of additional human factors not previously included, such as the weight, exact height, and gender of the pedestrian in front, improves the multiple linear regression model developed in subsection 4.3.1,

in Equation 4.10 it can be seen that this hypothesis is false. When considering all individuals and splitting the model into low and high density, and when including the pedestrian's gender in front, there may be a slight variation, so the model is different. However, as this effect is so small, less than 1%, it can be ignored, and the model remains unchanged compared to Equation 4.10. Only in this case is the hypothesis correct and the model is identical. The formula Equation 4.14 for the experiment at MEH is the same as Equation 4.10 for the experiment at GBS.

$$\text{Model III}_{\text{MEH}} : v_m = \beta_0 + \beta_1 \cdot h_m + \sum_{i=2}^{N+1} \beta_{2i} \cdot \text{alloence}_{m_i} + \epsilon_m . \quad (4.14)$$

## 4.4 Speed-Distance Relation with Fixed and Random Factors

The current analysis uses a mixed model to consider both fixed and random effects as discussed in subsection 3.2.3. The data set is not considered completely independent because the same individual collects it. This approach differs from previous models that have only included fixed effects such as headway, height, or gender and treat all data points as completely independent. Random effects, in contrast, are not directly observable. These could include personal attributes such as attention or motivation, such as those captured in the variable *alloence*, which are not treated as fixed factors. So, the new variation allows for a more comprehensive examination of the data because a hierarchical structure and multiple observations of a given individual are considered, too.

As mentioned, the final model in Equation 4.10 for the experiment at GBS is the same as in Equation 4.14 for the experiment at MEH. Thus, the model in Equation 4.15 is used as a simple model for comparison with a mixed model.

$$\text{Model}_{\text{Simple}} : v_m = \beta_0 + \beta_1 \cdot h_m + \sum_{i=2}^{N+1} \beta_{2i} \cdot \text{alloence}_{m_i} + \epsilon_m . \quad (4.15)$$

The aim is to determine which of the two models leads to improvement and should be used. An essential aspect of this analysis is the transition of the variable *alloence* from a fixed factor in previous models to a random factor in the mixed model. By treating *alloence* as a random factor, its potential correlation with other variables such as height or gender.

To assess this comparison between the two models, a  $\chi^2$  test is performed to evaluate whether the mixed model accounts for more variance than the regression model with only fixed effects and, thus, whether it is more effective than the simple model. If the null hypothesis is rejected, the mixed model is preferred to the simple model. This applies if  $p < 0.05$  applies. In the present analysis,  $H_0$  is rejected. These results suggest that the mixed model that includes the individual effects as a random factor is preferable to the simple model. Consequently, the use of mixed models for multivariate analysis is recommended in this context, as it helps to avoid the potential hiding of the effects of the variables of interest. By incorporating these random factors, the results may be more accurate. However, one reason for less accuracy could also be using the linear model. For some students, the free speed could be included because this could start earlier than at a headway of 1.5 m.

### 4.5 Gender Composition Related to Stop-And-Go Waves

This chapter presents a detailed analysis of the speed of different gender group compositions. The data from the single-file experiment at MEH is employed for this purpose. The primary focus is on the investigation of stop-and-go waves. Initially, velocity distributions are analyzed. Subsequently, time-space diagrams are investigated. Overall, only the global densities  $N \geq 20$  ( $N=20$  ( $\rho_{gl} = 1.33 \text{ m}^{-1}$ ),  $N=24$  ( $\rho_{gl} = 1.6 \text{ m}^{-1}$ ),  $N=32$  ( $\rho_{gl} = 2.13 \text{ m}^{-1}$ ),  $N=36$  ( $\rho_{gl} = 2.4 \text{ m}^{-1}$ ),  $N=40$  ( $\rho_{gl} = 2.67 \text{ m}^{-1}$ )) are used to illustrate the progression from go to the occurrence of stops for each of the various groups. The limiting velocity for the definition of a stop state is  $v_i(t) < 0.1 \text{ m/s}$ , as specified in subsection 3.1.2.

#### 4.5.1 Velocity Distribution for Gender Composition

This section investigates the velocity distributions for the gender group compositions of female, male, gender alternating order, and gender random order for varying densities. For this analysis, only the data from the lower straight section of the oval are considered, precisely the values of  $x$  in the range  $[0, 2.5] \text{ m}$ . Figure 4.19 illustrates the velocity distributions for the two group compositions, female on the left and gender random order on the right. These distributions were calculated for different global densities, namely,  $\rho_{gl} = 2.67 \text{ m}^{-1}$ ,  $\rho_{gl} = 2.40 \text{ m}^{-1}$ ,  $\rho_{gl} = 2.13 \text{ m}^{-1}$ ,

#### 4. Data Analysis and Results

$\rho_{gl} = 1.60 \text{ m}^{-1}$ ,  $\rho_{gl} = 1.33 \text{ m}^{-1}$  and  $\rho_{gl} = 1.33 \text{ m}^{-1}$ .

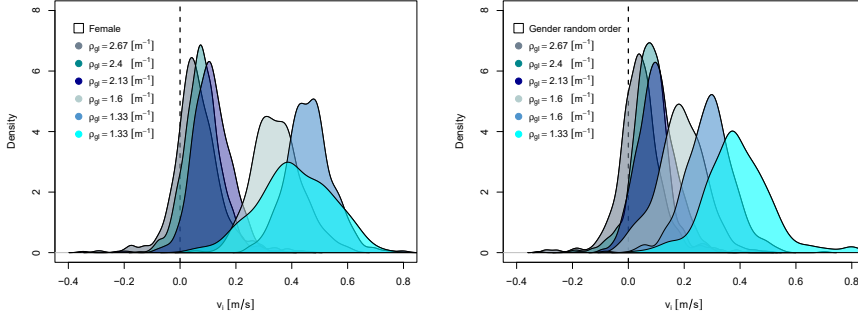


Figure 4.19: Velocity distributions for different global densities  $\rho_{gl} = 2.67 \text{ m}^{-1}$ ,  $\rho_{gl} = 2.40 \text{ m}^{-1}$ ,  $\rho_{gl} = 2.13 \text{ m}^{-1}$ ,  $\rho_{gl} = 1.60 \text{ m}^{-1}$ , and  $\rho_{gl} = 1.33 \text{ m}^{-1}$ . The figure on the left illustrates the female group composition, while the figure on the right illustrates a gender-random order.

For these two groups, lower values for the velocity are observed more frequently in the gender-random order group. In general, for global densities  $\rho_{gl} \geq 2.13 \text{ m}^{-1}$  for the groups female, male, gender-random order, and gender-alternating group, the velocity distributions are approximately equally distributed. The respective illustrations of the velocity distribution for the group compositions male and gender alternating can be found in Appendix 5.2, in Figure A.9 and Figure A.10. Figure 4.20 compares the velocity distribution of the highest densities observed in the given group compositions. The male group exhibits lower speeds, which are more strongly distributed around 0. This may indicate a tendency for males to stop more frequently or slow down. This phenomenon can be seen even though the maximum global density for the male group is  $\rho_{gl} \geq 2.34 \text{ m}^{-1}$  ( $N = 35$ ). It should be noted that this is the first experimental run for this group composition and the other group compositions for  $N = 40$ . However, no group composition exhibits multiple peaks in the velocity distribution. Only one peak is identified at high densities. An exception is when the velocity distribution for individuals is considered instead of the velocity distribution for all persons within a density. As a result, the velocity distribution will have two peaks.



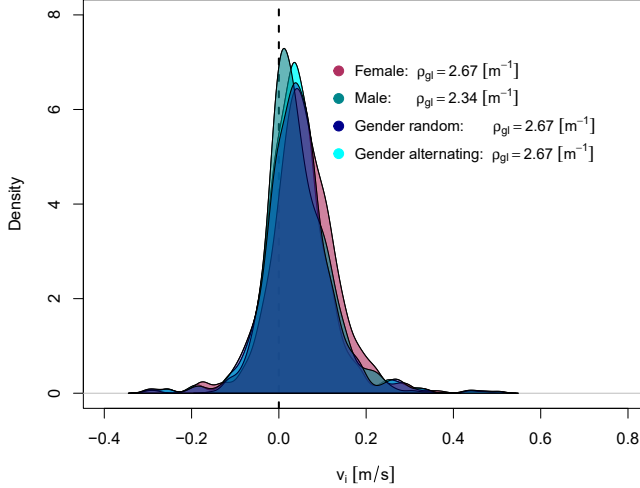


Figure 4.20: Velocity distributions for all four gender group compositions for the highest density.

### 4.5.2 Stop-And-Go Waves in Time-Space Diagrams

In this section, the stop-and-go waves are analyzed within time-space diagrams. This analysis aims to visualize the first occurrence times of the stop-and-go waves and identify the relevant global densities for further analysis.

First, the entire area of the oval, about 15 meters, is considered. In addition, the diagrams consider a time interval of about 60 seconds. Figure 4.21 shows an example of the course of the stop-and-go waves for the female group at a global density of  $\rho_{gl} = 1.33 \text{ m}^{-1}$  ( $N = 20$ ) up to a global density of  $\rho_{gl} = 2.67 \text{ m}^{-1}$  ( $N = 40$ ). The first small stops occur already at a density of  $N = 20$ . However, these first signs are rather sporadic and less pronounced. Only at a global density of  $\rho_{gl} = 2.13 \text{ m}^{-1}$  ( $N = 32$ ) do the stop-and-go changes become significantly more frequent and intense. From this density, there are more frequent and recognizable changes between stop and go. The dynamics of stop-and-go waves are closely linked to the number of pedestrians in a certain experimental run. At lower densities (e.g.,  $N = 20$ ), the individuals are still relatively free to move, which leads to few and short stops. With increasing density (e.g.,  $N = 32$ ), the space per person decreases, leading to increased interaction

#### 4. Data Analysis and Results

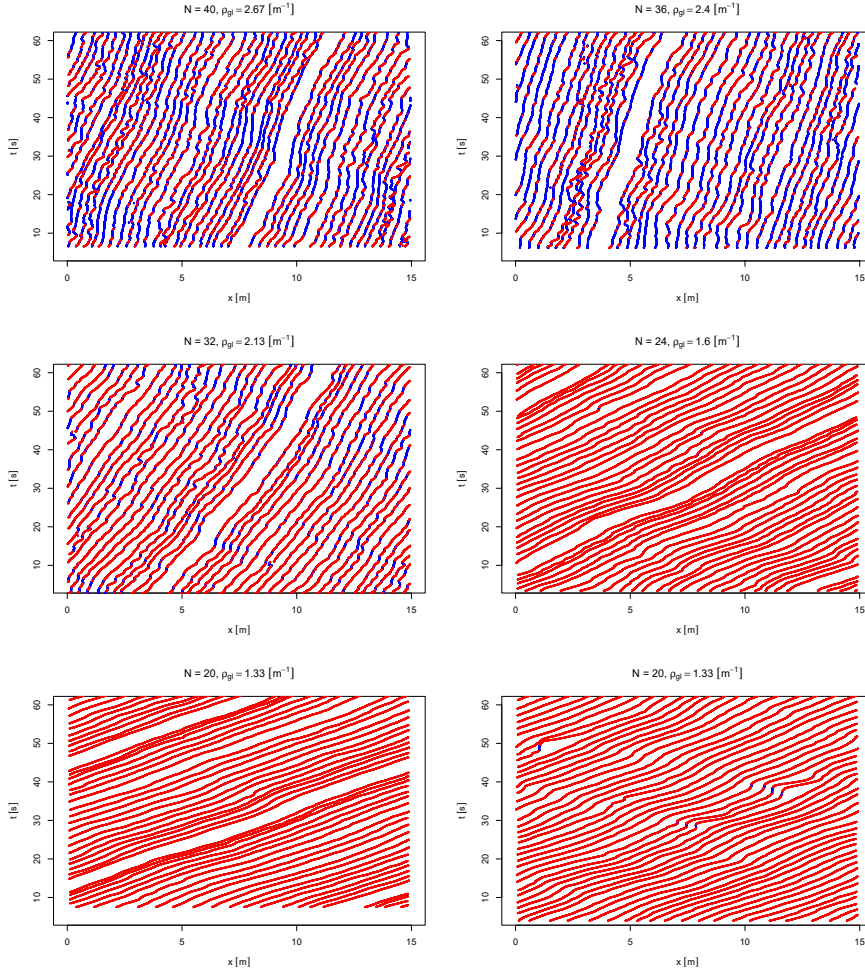


Figure 4.21: Time-space diagrams for the group composition female for different global densities.  $\rho_{gl} = 2.67 \text{ m}^{-1}$  top left,  $\rho_{gl} = 2.40 \text{ m}^{-1}$  top right,  $\rho_{gl} = 2.13 \text{ m}^{-1}$  middle left,  $\rho_{gl} = 1.60 \text{ m}^{-1}$  middle right,  $\rho_{gl} = 1.33 \text{ m}^{-1}$  bottom left and  $\rho_{gl} = 1.33 \text{ m}^{-1}$  bottom right.

and, ultimately, more frequent stops. The time-space diagrams in Appendix 5.2, Figure A.11, Figure A.13, Figure A.12 also illustrate these trends for different groups: male, gender random order, and gender alternating order. It is shown that the occurrence of the stop-and-go waves follows a similar course in all groups. Compared to previous studies [22, 25, 49, 52], the results confirm that stop-and-go waves only occur from a global density of about  $2 \pm 0.15 \text{ m}^{-1}$ . It should be noted that slight deviations from this threshold may occur in specific group dynamics and individual behavior. The observations of the progression of the stop-and-go waves are not only visually understandable but can also be measured quantitatively. This phase transformation will be examined in more detail in the subsequent section, considering the gender composition of the groups.

### 4.5.3 Phase Transformation between Stop-And-Go States

The percentage ratio of pedestrians in stop-and-go states is considered for analysis in the phase transition between stop-and-go states. This is calculated using the formula in Equation 3.5. Figure 4.22 illustrates the relative frequency of individuals

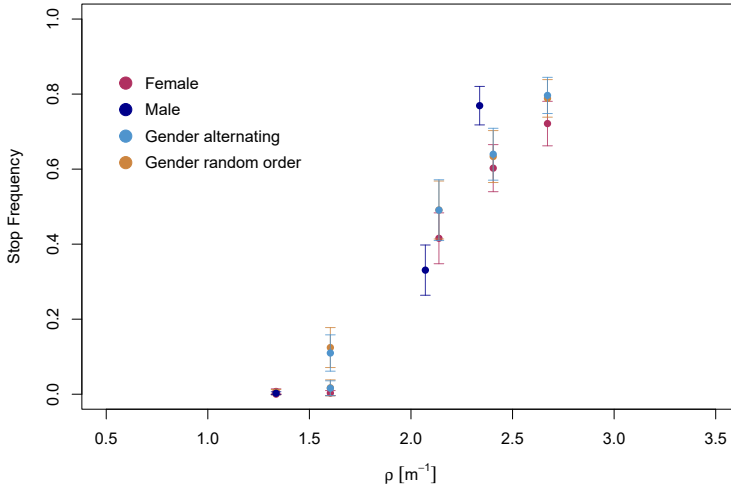


Figure 4.22: The relative frequency of individuals in the stop is illustrated for four group compositions for the global densities from  $\rho_{gl} = 1.33 \text{ m}^{-1}$  to  $\rho_{gl} = 2.67 \text{ m}^{-1}$  by the mean values and the standard deviations.

#### 4. Data Analysis and Results

in the stop for the group compositions female, male, gender alternating and gender random order for varying global densities from  $\rho_{gl} = 1.33 \text{ m}^{-1}$  to  $\rho_{gl} = 2.67 \text{ m}^{-1}$  by the mean values and the standard deviations. These values are presented in Table 4.5.

The global density of  $1.33 \text{ m}^{-1}$  ( $N = 20$ ) indicates that the groups are mainly in the go state. This density marks the beginning of a transition in which the relationship between movement and standstill changes continuously. Although a complete standstill is not shown in the available data, the corresponding Figure 4.22 illustrates the transition between the states for the individual groups. The phase transformation between the stop-and-go states can be classified according to Ehrenfest's criteria [73] and identified with a specific order parameter. In this case, the transition is second-order, as a continuous transition from the go to the stop state can be observed. This indicates no abrupt change between the two states but rather a continuous change. In the study by Eilhardt [52], a similar inhomogeneous state is described in the density range from around  $2.25 \text{ m}^{-1}$  to  $2.6 \text{ m}^{-1}$ . These results support the observations from the current experiment and demonstrate that the transformation from movement to standstill occurs in a comparable density range. In the work of Ziemer [22] in Section 5.3.2, Figure 5.7, the course for the single-file experiment at GBS is the same for the 5th grade.

Table 4.5: Mean values and standard deviations for the stop frequency of different pedestrians that are illustrated in Figure 4.22 for different global densities and group compositions.

$\rho_{gl}$	N	Gender Alternating	Gender random order	Female	Male
2.67	40	$0.80 \pm 0.05$	$0.79 \pm 0.05$	$0.72 \pm 0.06$	/
2.40	36	$0.64 \pm 0.07$	$0.63 \pm 0.07$	$0.60 \pm 0.06$	/
2.34	35	/	/	/	$0.77 \pm 0.05$
2.13	32	$0.49 \pm 0.08$	$0.49 \pm 0.08$	$0.42 \pm 0.07$	/
2.07	31	/	/	/	$0.33 \pm 0.07$
1.60	24	$0.11 \pm 0.05$	$0.12 \pm 0.05$	$0.003 \pm 0.01$	/
1.60	24	$0.02 \pm 0.02$	$0.01 \pm 0.01$	/	/

The mean values and standard deviations presented in Table 4.5 demonstrate that the mean values increase with increasing density. Interestingly, the fluctuation range of the standard deviations remains relatively constant despite increasing density. The mean values are highly similar for the group's gender random order and gender alternating, and these groups have the highest stop frequency in the various density ranges. In comparison, the values for the female group are somewhat lower. Due

to the varying density, it is impossible to directly compare the values for the male group and those for the other groups. However, it is notable that the values for a global density of  $2.34 \text{ m}^{-1}$  ( $N = 35$ ) are comparable with the values for the global density of  $2.67 \text{ m}^{-1}$  ( $N = 40$ ) and are very similar to those of the other groups. This discrepancy may be because the first run for the male group had a global density of  $2.34 \text{ m}^{-1}$  ( $N = 35$ ), while the other groups started their first run at a global density of  $2.67 \text{ m}^{-1}$  ( $N = 40$ ).

The mean duration of a person's stay in the stop-and-go state is calculated based on Equation 3.6. For the gender alternating and gender random order groups, the values are nearly equal for the global densities of  $2.67 \text{ m}^{-1}$  ( $N = 40$ ) and  $2.40 \text{ m}^{-1}$  ( $N = 36$ ). The lower mean duration in the stop state is presented only for the female group composition (see Figure A.14).

Regarding the research question of whether there are differences in stop-and-go waves at high densities for homogeneous and heterogeneous group compositions regarding gender, the velocity distributions at high densities are nearly identical for all group compositions. Considering all individuals in a density, there is only one peak in the distribution. Two peaks appear only when the velocity distributions of specific individuals are examined in detail. Time-space diagrams illustrate that stop states are visible for global densities  $\rho_{gt} > 1.33 \text{ m}^{-1}$  in all group compositions. In addition, the mean values for the number of pedestrians in a stop state increased continuously with increasing density. Individuals in a stop state are more frequent in heterogeneous gender group compositions, with gender random order and gender alternating, than in the homogeneous female group. Furthermore, the mean and standard deviation trend is approximately the same for the heterogeneous group compositions. Regarding the mean duration of a person's stay in the stop-and-go state, the time for the stop state is higher in the heterogeneous groups, which are also almost equal. In conclusion, the heterogeneous groups are practically equal and slightly different from the homogeneous groups.

## Chapter 5

# Discussion and Outlook

This chapter presents the findings and implications of the previous chapter in the following section: conclusions. Then, an outlook on future research is given.

### 5.1 Conclusions

This study addresses two main aspects. First, it compares different group compositions based on gender and other cultures. Second, it employs regression analysis to analyze and quantify the effects of multiple factors on speed. The second point addresses a specific issue: previous research on single-file experiments showed that age can affect fundamental diagrams. However, there is a contradictory relation to how age affects the diagrams. One potential explanation for these discrepancies is that previous experiments are not directly comparable, as even if a group is homogeneous in one factor, it may be heterogeneous in other factors, such as gender. Previous comparisons of fundamental diagrams have utilized only the cumulative data of all individuals in the group. In this new study, individual fundamental diagrams are introduced, considered, and analyzed in one-dimensional single-file experiments. Furthermore, to describe pedestrians on a personal level and for regression analysis, additional factors such as height, weight, gender, and age are considered.

First, the hypothesis that the relationship between speed and density is influenced by the gender composition of the group of pedestrians is investigated. This hypothesis is based on the results of previous single-file experiments conducted by Subaih et al. at AAU[35]. The results of the study above in Palestine indicated that homogeneous groups concerning gender tend to have higher speeds than heterogeneous groups at a given density of approximately  $1.0 \text{ m}^{-1}$  and above. New single-file experiments

were conducted in Germany at MEH with homogeneous and heterogeneous group compositions to verify this result. The comparison of different gender compositions across different density ranges revealed the following patterns: There is no discernible variation in speed between group compositions at low densities. Only at densities around  $1.4 \text{ m}^{-1}$  did homogeneous groups exhibit higher speeds than heterogeneous groups. Between densities of  $1.5 \text{ m}^{-1}$  and  $2.15 \text{ m}^{-1}$ , the differences decreased as the speeds of the homogeneous groups approached those of the heterogeneous gender group and the gender alternating group. The randomly ordered gender group is the slowest in this interval. Finally, at high densities, the mean values of the speeds of all groups are similar. In conclusion, there is a progression from unequal distribution in the initial interval to complete equality in the final interval.

A comparison of the results of the Palestine experiments with those of the German experiments reveals some similarities, although only in some details. No consistent difference exists between the homogeneous and heterogeneous groups for densities below  $0.8 \text{ m}^{-1}$ . However, a discrepancy between the two groups is observed around a density of  $1.4 \text{ m}^{-1}$ , less pronounced in the German experiments. Furthermore, the speed in the density range of  $1 \text{ m}^{-1}$  to  $1.5 \text{ m}^{-1}$  is higher in Palestine than in Germany. For higher densities, data are available only for the heterogeneous group in Palestine, and their mean speed approaches that of the German groups. Therefore, the first hypothesis is confirmed. However, this relevance to the German experiment is only weak upon closer examination. The discrepancy is only discernible in a narrow range of densities and is relatively minor. As density increases, group compositions become increasingly similar. It cannot be ruled out that this phenomenon may be more pronounced in other cultural contexts. For certain density intervals, the speed is higher for all group compositions in Palestine, and in a middle-density interval, it is higher for homogeneous groups than for heterogeneous groups. The data for homogeneous group compositions in high-density intervals are unavailable for Palestine. However, for high-density intervals, the data for heterogeneous groups in Palestine at AAU are comparable to those in Germany at MEH.

Regarding the significance of the effect, it is essential to note that the confirmation of the hypothesis depends on the testing method and data preparation. The size of the binning intervals utilized to group the data also plays a role. Depending on the size of these intervals selected for different test methods, the described systematic variation may not be apparent. This was evident even with intervals as small as 0.2. Other tests, such as the t-test, give similar results to the Tukey HSD test. Nevertheless, more sensitive tests, such as the Kolmogorov–Smirnov test, demonstrate differences

in the speed distributions across most density intervals, leading to the rejection of the hypothesis. In all density intervals, the differences in mean speeds between different group compositions are smaller than the standard deviation.

The second hypothesis states that individual speed-distance functions depend on age and gender, and the third hypothesis, which suggests that at high densities, the distance between individuals depends on the gender of neighboring pedestrians, is investigated in the context of a simple linear regression analysis. This investigation is conducted for single-file experiments carried out at the GBS in Germany for the second hypothesis and at the MEH in Germany for the third hypothesis. The analysis at GBS aimed to determine whether gender and age play a role in the individual speed-distance functions. Factors such as minimum distance, reaction time, and data scatter are analyzed using simple linear regression on individual fundamental diagrams. The results in subsection 4.2.1 showed no tendency regarding scatter patterns. The comparison of older and younger students and female and male students for the reaction time showed no differences between these groups. In contrast, the minimum distances significantly differ between older and younger students. Older students exhibited a more pronounced effect than younger students; generally, age was more pronounced than gender. For the experiment at MEH, and therefore for the third hypothesis, the mean values of  $(d_V)_{i,min}$  and  $\hat{\beta}_{1_i}$  are compared using the Tukey HSD test across different group compositions (female, male, gender alternating, and gender random order). Based on this test, no significant differences are observed. Therefore, it is concluded that there are no significant differences in reaction time and minimal distances between the different group compositions. This means that the third hypothesis that at high densities, the distance between individuals depends on the gender of the neighboring pedestrians cannot be confirmed.

For the fourth hypothesis, a multiple linear regression analysis examines how distance and individual factors such as gender, height, and age affect speed. The speed models at GBS are based on *headway*, *gender*, *height*, and *alloence*. *Alloence* is included to account for unmeasurable individual factors such as motivation or attention. The results showed that *alloence* is crucial and that *gender* can only be excluded from the model if the unknown factor is not ignored. The analysis confirms that *headway* has the most significant effect on speed. Other unknown effects follow, with height having a minimal impact. Therefore, the unknown factors have a more significant influence than either height or gender. This is consistent with the finding that gender differences are smaller than differences between younger and older students. Further examination of the residuals indicates that individual



speed is not determined only by the predictor variables. This suggests that there may be additional effects on speed beyond height and other unknown individual factors. Consequently, for the single-file experiment at MEH that includes additional human factors, the fifth hypothesis is that the inclusion of additional human factors such as weight, exact height, and gender of the previous pedestrian will improve the multiple linear regression model developed for the single-file experiment at GBS in subsection 4.3.1. The analysis provides the same model as mentioned before when excluding factors with minimal impact on speed. Once again, headway is identified as the most influential factor on speed, followed by other unknown individual effects. The pedestrian's gender in front affects the model, but it is less than 1% and is considered insignificant. Therefore, the model is unchanged. Thus, it can be concluded that this hypothesis is incorrect.

Furthermore, the extended regression analysis for the experiment at GBS compares a simple model with only fixed factors to a more complex mixed model that includes individual speed as a function of fixed factors such as headway and height, along with other unknown individual effects as random factors. The analysis indicates that the mixed model is a more suitable approach. For this reason, the individual effects should be treated as random so that the impact of the investigated factors is not obscured.

The final hypothesis of this study is that the gender composition of the individuals involved influences stop-and-go waves. The velocity distributions in high-density intervals are approximately the same for all group compositions. A comparison of the mean values and their standard deviations regarding the number of pedestrians in a stop state revealed that these values increased continuously with increasing density. Furthermore, the heterogeneous groups demonstrated a greater prevalence of gender random order and gender alternating than the homogeneous group of females. This is also true for the mean duration of a person's stay in a stop-and-go state. Therefore, concerning the abovementioned hypothesis, heterogeneous groups are almost equal and exhibit slight differences from homogeneous groups in stop-and-go states.

## 5.2 Outlook

For further research, a more detailed look at the influence of culture may be helpful. For example, when comparing data from Germany and Palestine, speed is more significant in Palestine. The unclear cause of this difference indicates the

need for additional data from different cultures. Due to insufficient data, no further conclusions about the relationship between speed and density in Palestine in the high-density area can be drawn.

Moreover, in the high-density intervals of the experiment at MEH conducted in Germany, an interesting phenomenon can be observed in some cases (e.g., for Female for  $N = 36$ , and Male for  $N = 32$ ) where the number of pedestrians who are in a stopping and going state are compared (see Figure A.15). Despite the predominant tendency towards stops, there are phases in which more go movements occur. This unexpected change between phases may be attributed to individuals who sometimes try to walk in rhythm or step in step to make faster progress. This observation raises the question of whether synchronized movement could influence the ratio of stop-and-go within specific densities and possibly delay the occurrence of the stop-and-go waves. A detailed analysis of this aspect would be of great interest, as it could provide insight into how coordinated movements in dense crowds influence the dynamics of the stop-and-go waves in terms of frequency and intensity of the stop-and-go waves. However, the experiments considered here do not provide an optimal basis for a comprehensive synchronized step analysis. Although the participants tried to coordinate with each other and walk in step at the beginning of the homogeneous series of experiments, these efforts were not systematically continued. Due to the limited data, it is not easy to investigate the effects of synchronized walking precisely on stop-and-go dynamics. Nevertheless, future studies investigating synchronized movements could provide valuable insights into how and when synchronized stepping influences the occurrence of stop-and-go waves. Another interesting approach in this regard could be to compare different cultural backgrounds. In some cultures, walking in step is more firmly established than in others.

## Bibliography

- [1] V. M. Predtechenskii and A. I. Milinskii. *Planning for foot traffic flow in buildings*. eng. New Delhi: Amerind, 1978. ISBN: 9780862493264. URL: <https://books.google.de/books?id=3AZaPwAACAAJ>.
- [2] A. Jelić et al. *Properties of pedestrians walking in line: Fundamental diagrams*. Vol. 85. American Physical Society, Mar. 2012, p. 036111. DOI: 10.1103/PhysRevE.85.036111.
- [3] S. Holl. *Methoden für die Bemessung der Leistungsfähigkeit multidirektional genutzter Fußverkehrsanlagen*. ger. OCLC: 989761703. Jülich: Forschungszentrum Jülich GmbH, Zentralbibliothek, 2016. ISBN: 9783958061910. URL: [https://juser.fz-juelich.de/record/825757/files/IAS\\_Series\\_32.pdf](https://juser.fz-juelich.de/record/825757/files/IAS_Series_32.pdf).
- [4] U. Weidmann. *Transporttechnik der Fussgänger: Transporttechnische Eigenschaften des Fussgängerverkehrs, Literaturauswertung*. de. Tech. rep. ETH Zurich, 1993, 114 p. DOI: 10.3929/ETHZ-B-000242008. URL: <http://hdl.handle.net/20.500.11850/242008>.
- [5] S. Buchmueller and U. Weidmann. *Parameters of pedestrians, pedestrian traffic and walking facilities*. Tech. rep. 2006.
- [6] J. J. Fruin. *Pedestrian planning and design*. Tech. rep. fruin1971pedestrian. 1971.
- [7] S. Burghardt, A. Seyfried, and W. Klingsch. *Performance of stairs – Fundamental diagram and topographical measurements*. en. Vol. 37. 2013, pp. 268–278. DOI: 10.1016/j.trc.2013.05.002. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0968090X13000946>.
- [8] J. Zhang. *Pedestrian fundamental diagrams: comparative analysis of experiments in different geometries*. eng. Schriften des Forschungszentrums Jülich IAS Series 14. Jülich: Forschungszentrum Jülich, 2012. ISBN: 9783893368259.

- URL: <https://juser.fz-juelich.de/record/128157/files/FZJ-2012-01052.pdf>.
- [9] L. D. Vanumu, R. K. Ramachandra, and G. Tiwari. *Fundamental diagrams of pedestrian flow characteristics: A review*. en. Vol. 9. 4. 2017, p. 49. DOI: 10.1007/s12544-017-0264-6. URL: <http://link.springer.com/10.1007/s12544-017-0264-6>.
  - [10] S. Cao et al. *Investigation of difference of fundamental diagrams in pedestrian flow*. en. Vol. 506. 2018, pp. 661–670. DOI: 10.1016/j.physa.2018.04.084. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0378437118305120>.
  - [11] C. Feliciani and K. Nishinari. *Empirical analysis of the lane formation process in bidirectional pedestrian flow*. en. Vol. 94. 3. 2016, p. 032304. DOI: 10.1103/PhysRevE.94.032304. URL: <https://link.aps.org/doi/10.1103/PhysRevE.94.032304>.
  - [12] Y. Hu et al. *Social groups barely change the speed-density relationship in unidirectional pedestrian flow, but affect operational behaviours*. en. Vol. 139. 2021, p. 105259. DOI: 10.1016/j.ssci.2021.105259. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0925753521001041>.
  - [13] X. Ren et al. *The fundamental diagrams of elderly pedestrian flow in straight corridors under different densities*. Vol. 2019. 2. 2019, p. 023403. DOI: 10.1088/1742-5468/aafa7b. URL: <https://iopscience.iop.org/article/10.1088/1742-5468/aafa7b>.
  - [14] C. Jin et al. *Observational characteristics of pedestrian flows under high-density conditions based on controlled experiments*. en. Vol. 109. 2019, pp. 137–154. DOI: 10.1016/j.trc.2019.10.013. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0968090X19306527>.
  - [15] S. Cao et al. *Fundamental diagrams for multidirectional pedestrian flows*. Vol. 2017. 3. 2017, p. 033404. DOI: 10.1088/1742-5468/aa620d. URL: <https://iopscience.iop.org/article/10.1088/1742-5468/aa620d>.
  - [16] R. Ye et al. *Pedestrian single-file movement on stairs under different motivations*. en. Vol. 571. 2021, p. 125849. DOI: 10.1016/j.physa.2021.125849. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0378437121001217>.
  - [17] S. Cao et al. *Pedestrian dynamics in single-file movement of crowd with different age compositions*. en. Vol. 94. 1. 2016, p. 012312. DOI: 10.1103/PhysRevE.94.012312. URL: <https://link.aps.org/doi/10.1103/PhysRevE.94.012312>.

- [18] U. Chattaraj, A. Seyfried, and P. Chakroborty. *Comparison of pedestrian fundamental diagram across cultures*. Vol. 12. 3. 2009, pp. 393–405. DOI: 10.1142/S0219525909002209.
- [19] J. Zhang, A. Tordeux, and A. Seyfried. *Effects of Boundary Conditions on Single-File Pedestrian Flow*. Springer International Publishing, 2014, pp. 462–469. ISBN: 9783319115207. DOI: 10.1007/978-3-319-11520-7\_48.
- [20] A. Seyfried, A. Portz, and Schadschneider A. *Phase coexistence in congested states of pedestrian dynamics*. 2010. DOI: 10.1007/978-3-642-15979-4\_53. arXiv: 1006.3546 [physics.soc-ph].
- [21] A. Portz and A. Seyfried. *Analyzing Stop-and-Go Waves by Experiment and Modeling*. Ed. by Richard D. Peacock, Erica D. Kuligowski, and Jason D. Averill. Boston, MA: Springer US, 2011, pp. 577–586. ISBN: 978-1-4419-9725-8.
- [22] V. Ziemer. *Mikroskopische Fundamentaldiagramme der Fußgängerdyamik empirische Untersuchung von Experimenten eindimensionaler Bewegung sowie quantitative Beschreibung von Stau-Charakteristika*. ger. OCLC: 1199751618. Jülich: Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag, 2020. ISBN: 9783958064706. URL: [https://juser.fz-juelich.de/record/877610/files/IAS\\_Series\\_42.pdf](https://juser.fz-juelich.de/record/877610/files/IAS_Series_42.pdf).
- [23] D. Helbing, A. Johansson, and H. Z. Al-Abideen. *Dynamics of crowd disasters: An empirical study*. Vol. 75. American Physical Society, Apr. 2007, p. 046109. DOI: 10.1103/PhysRevE.75.046109. URL: <https://link.aps.org/doi/10.1103/PhysRevE.75.046109>.
- [24] J. Zhang et al. *Ordering in bidirectional pedestrian flows and its influence on the fundamental diagram*. Vol. 2012. 02. Feb. 2012, P02002. DOI: 10.1088/1742-5468/2012/02/P02002. URL: <https://dx.doi.org/10.1088/1742-5468/2012/02/P02002>.
- [25] X. Ren, J. Zhang, and W. Song. *Contrastive study on the single-file pedestrian movement of the elderly and other age groups*. Vol. 2019. 9. 2019, p. 093402. DOI: 10.1088/1742-5468/ab39da. URL: <https://iopscience.iop.org/article/10.1088/1742-5468/ab39da>.
- [26] S. Cao et al. *The stepping behavior analysis of pedestrians from different age groups via a single-file experiment*. Vol. 2018. 3. 2018, p. 033402. DOI: 10.1088/1742-5468/aab04f. URL: <https://iopscience.iop.org/article/10.1088/1742-5468/aab04f>.

- [27] S. Huang et al. *Experimental study on luggage-laden pedestrian movement in narrow seat aisle*. Vol. 257. 2019, p. 012049. DOI: 10.1088/1755-1315/257/1/012049. URL: <https://iopscience.iop.org/article/10.1088/1755-1315/257/1/012049>.
- [28] J. Zhang et al. *Universal flow-density relation of single-file bicycle, pedestrian and car motion*. en. Vol. 378. 44. 2014, pp. 3274–3277. DOI: 10.1016/j.physleta.2014.09.039. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0375960114009517>.
- [29] R. Subaih et al. *Experimental Investigation on the Alleged Gender-differences in Pedestrian Dynamics: A Study Reveals No Gender Differences in Pedestrian Movement Behavior*. Vol. 8. New York, NY: IEEE, 2020, pp. 33748–33757. DOI: 10.1109/ACCESS.2020.2973917.
- [30] G. Zeng et al. *Experimental and modeling study on relation of pedestrian step length and frequency under different headways*. en. Vol. 500. 2018, pp. 237–248. DOI: 10.1016/j.physa.2018.02.095. URL: <https://linkinghub.elsevier.com/retrieve/pii/S037843711830181X>.
- [31] G. Zeng et al. *Experimental study on the effect of background music on pedestrian movement at high density*. en. Vol. 383. 10. 2019, pp. 1011–1018. DOI: 10.1016/j.physleta.2018.12.019. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0375960118312258>.
- [32] R. Subaih et al. *Gender-based Insights into the Fundamental Diagram of Pedestrian Dynamics*. Aug. 2019, pp. 613–624. ISBN: 978-3-030-28376-6. DOI: 10.1007/978-3-030-28377-3\_51.
- [33] S. Cao et al. *Dynamic analysis of pedestrian movement in single-file experiment under limited visibility*. en. Vol. 69. 2019, pp. 329–342. DOI: 10.1016/j.cnsns.2018.10.007. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1007570418303228>.
- [34] D. Yanagisawa, A. Tomoeda, and K. Nishinari. *Improvement of pedestrian flow by slow rhythm*. en. Vol. 85. 1. 2012, p. 016111. DOI: 10.1103/PhysRevE.85.016111. URL: <https://link.aps.org/doi/10.1103/PhysRevE.85.016111>.
- [35] R. Subaih et al. *Experimental Investigation on the Alleged Gender-Differences in Pedestrian Dynamics: A Study Reveals No Gender Differences in Pedestrian Movement Behavior*. Vol. 8. 2020, pp. 33748–33757. DOI: 10.1109/ACCESS.2020.2973917.

- [36] J. Zhang et al. *Homogeneity and Activeness of Crowd on Aged Pedestrian Dynamics*. en. Vol. 83. The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) / The 6th International Conference on Sustainable Energy Information Technology (SEIT-2016) / Affiliated Workshops. Jan. 2016, pp. 361–368. DOI: 10.1016/j.procs.2016.04.137. URL: <https://www.sciencedirect.com/science/article/pii/S1877050916301673>.
- [37] R. Subaih et al. *Questioning the Anisotropy of Pedestrian Dynamics: An Empirical Analysis with Artificial Neural Networks*. en. Vol. 12. 15. 2022, p. 7563. DOI: 10.3390/app12157563. URL: <https://www.mdpi.com/2076-3417/12/15/7563>.
- [38] A. Seyfried, A. Portz, and A. Schadschneider. “Phase Coexistence in Congested States of Pedestrian Dynamics”. en. In: *Cellular Automata*. Ed. by Stefania Bandini et al. Vol. 6350. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 496–505. DOI: 10.1007/978-3-642-15979-4\_53. URL: [http://link.springer.com/10.1007/978-3-642-15979-4\\_53](http://link.springer.com/10.1007/978-3-642-15979-4_53).
- [39] M. Favaretto et al. *Investigating cultural aspects in the fundamental diagram using convolutional neural networks and virtual agent simulation*. en. Vol. 30. 3-4. 2019. DOI: 10.1002/cav.1899. URL: <https://onlinelibrary.wiley.com/doi/10.1002/cav.1899>.
- [40] C. Dias et al. *Pedestrians’ Microscopic Walking Dynamics in Single-File Movement: The Influence of Gender*. en. Vol. 12. 19. 2022, p. 9714. DOI: 10.3390/app12199714. URL: <https://www.mdpi.com/2076-3417/12/19/9714>.
- [41] J. Wang et al. *Linking pedestrian flow characteristics with stepping locomotion*. en. Vol. 500. 2018, pp. 106–120. DOI: 10.1016/j.physa.2018.02.021. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0378437118300979>.
- [42] Y. Ma et al. *Pedestrian stepping dynamics in single-file movement*. en. Vol. 98. 6. 2018, p. 062311. DOI: 10.1103/PhysRevE.98.062311. URL: <https://link.aps.org/doi/10.1103/PhysRevE.98.062311>.
- [43] W. Song, W. Lv, and Z. Fang. *Experiment and Modeling of Microscopic Movement Characteristic of Pedestrians*. en. Vol. 62. 2013, pp. 56–70. DOI: 10.1016/j.proeng.2013.08.044. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1877705813012265>.

- [44] J. Wang et al. *Step styles of pedestrians at different densities*. Vol. 2018. 2. 2018, p. 023406. DOI: 10.1088/1742-5468/aaac57. URL: <https://iopscience.iop.org/article/10.1088/1742-5468/aaac57>.
- [45] A. Fujita et al. *Traffic flow in a crowd of pedestrians walking at different speeds*. en. Vol. 99. 6. 2019, p. 062307. DOI: 10.1103/PhysRevE.99.062307. URL: <https://link.aps.org/doi/10.1103/PhysRevE.99.062307>.
- [46] J. Ma et al. *Experimental study of single-file pedestrian movement with height constraints*. Vol. 2020. 7. 2020, p. 073409. DOI: 10.1088/1742-5468/ab99c0. URL: <https://iopscience.iop.org/article/10.1088/1742-5468/ab99c0>.
- [47] Y. J. Kim, C. Lee, and J. H. Kim. *Sidewalk Landscape Structure and Thermal Conditions for Child and Adult Pedestrians*. Vol. 15. 1. 2018. DOI: 10.3390/ijerph15010148. URL: <https://www.mdpi.com/1660-4601/15/1/148>.
- [48] H. Maryam et al. *Sidewalk Measurements from Satellite Images: Preliminary Findings*. 2021. arXiv: 2112.06120 [cs.CV].
- [49] M. Li, L. Yang, and P. Zhang. *Single-file movement experiments of male under the influence of rhythm*. Vol. 84. 2023. DOI: <http://doi.org/10.54941/ahfe1003618>.
- [50] J. Zhang et al. *Homogeneity and Activeness of Crowd on Aged Pedestrian Dynamics*. Vol. 83. Dec. 2016, pp. 361–368. DOI: 10.1016/j.procs.2016.04.137.
- [51] S. Paetzke, M. Boltes, and A. Seyfried. *Influence of individual factors on fundamental diagrams of pedestrians*. en. Vol. 595. 2022, p. 127077. DOI: 10.1016/j.physa.2022.127077. URL: <https://www.sciencedirect.com/science/article/pii/S0378437122001248>.
- [52] C. Eilhardt. *Computer simulation of pedestrian dynamics at high densities*. 2014. URL: <https://kups.ub.uni-koeln.de/5848/>.
- [53] FZJ IAS-7. *Data archive of experimental data from studies about pedestrian dynamics, Single-file motion of pupils: School GymBay*. DOI: 10.34735/ped.2014.2.
- [54] A. K. Boomers et al. *Pedestrian Crowd Management Experiments: A Data Guidance Paper*. Vol. 8. May 2023, pp. 1–57. DOI: 10.17815/CD.2023.141. URL: <https://collective-dynamics.eu/index.php/cod/article/view/A141>.



- [55] FZJ IAS-7. *Data archive of experimental data from studies about pedestrian dynamics, Influence of gender in single-file movement*. DOI: 10.34735/ped.2018.5.
- [56] Forschungszentrum Jülich. *Influence of gender in single-file movement*. <http://ped.fz-juelich.de/da/2018singleFile>. DOI: 10.34735/ped.2018.5.
- [57] M. Boltes et al. *Automatic Extraction of Pedestrian Trajectories from Video Recordings*. en. Ed. by Wolfram W. F. Klingsch et al. Berlin, Heidelberg: Springer, 2010, pp. 43–54. ISBN: 9783642045042. DOI: 10.1007/978-3-642-04504-2\_3.
- [58] *PeTrack*. <http://ped.fz-juelich.de/petrack>. Accessed: 2020-11-24.
- [59] M. Boltes and A. Seyfried. *Collecting pedestrian trajectories*. Vol. 100. Special issue: Behaviours in video. 2013, pp. 127–133. DOI: <https://doi.org/10.1016/j.neucom.2012.01.036>.
- [60] V. Ziemer, A. Seyfried, and A. Schadschneider. *Congestion Dynamics in Pedestrian Single-File Motion*. en. Ed. by Victor L. Knoop and Winnie Daamen. Cham: Springer International Publishing, 2016, pp. 89–96. ISBN: 9783319334820. DOI: 10.1007/978-3-319-33482-0\_12.
- [61] *Design of experiments*. <https://methpsy.elearning.psych.tu-dresden.de/mediawiki/index.php/Versuchsplanung>. Accessed: 2024-02-20.
- [62] W. Kemloh et al. *JuPedSim: an open framework for simulating and analyzing the dynamics of pedestrians*. 2015.
- [63] A. Tordeux and A. Schadschneider. *White and relaxed noises in optimal velocity models for pedestrian flow with stop-and-go waves*. Vol. 49. May 2016, p. 185101. DOI: 10.1088/1751-8113/49/18/185101.
- [64] *Assumptions of Multiple Linear Regression*. <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/assumptions-of-multiple-linear-regression/>. Accessed: 2024-06-09.
- [65] J. T. Ormerod. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Vol. 36. 2006. ISBN: 9781498720960. DOI: <https://doi.org/10.1002/sim.7282>.
- [66] D. Bates. *Computational methods for mixed models*. 2014.
- [67] *Mixed Effects Regression*. <https://ladal.edu.au>. Accessed: 2024-03-17.

## BIBLIOGRAPHY

---

- [68] A. Nanda, B. B. Mohapatra, and A. P. K. Mahapatra. *Multiple comparison test by Tukey's honestly significant difference (HSD): Do the confident level control type I error*. Vol. 6(1). 2021. DOI: 10.22271/math.2021.v6.i1a.636.
- [69] JY. Benjamin and HB. John. *Tukey's contributions to multiple comparisons*. Vol. 30(6). 2002.
- [70] J. A. Rafter, M.L. Abell, and J.P. Braselton. *Multiple comparison methods for Means*. Vol. 44(2). 2002.
- [71] S. Paetzke, M. Boltes, and A. Seyfried. *Influence of Gender Composition in Pedestrian Single-File Experiments*. Vol. 13. 9. 2023. DOI: 10.3390/app13095450. URL: <https://www.mdpi.com/2076-3417/13/9/5450>.
- [72] H. Akaike. *A new look at the statistical model identification*. Vol. 19. 6. 1974, pp. 716–723. DOI: 10.1109/TAC.1974.1100705.
- [73] *Klassifikation der Phasenübergänge nach Ehrenfest*. <https://www.spektrum.de/lexikon/physik/klassifikation-der-phasenuebergaenge-nach-ehrenfest/8061>. Accessed: 2024-06-09.

---

# Appendix A

## A.1

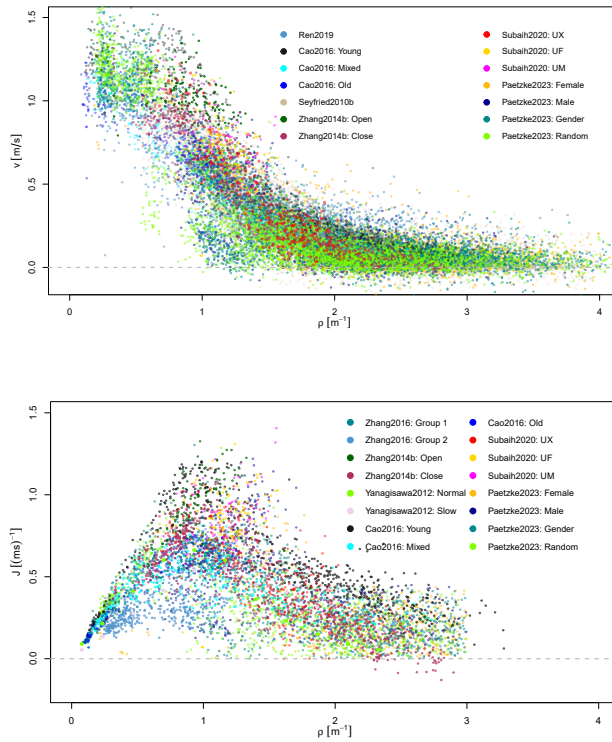


Figure A.1: Two fundamental diagrams based on recent studies [17, 19, 25, 34–36] are presented. The top figure shows the velocity-density relation, and the figure at the bottom is the flow-density relation.

## A.2

The following figures illustrate the distance vs. speed relationship for several main IDs to show the approximately linear course for  $h < 1.5$  m for the data at GBS.

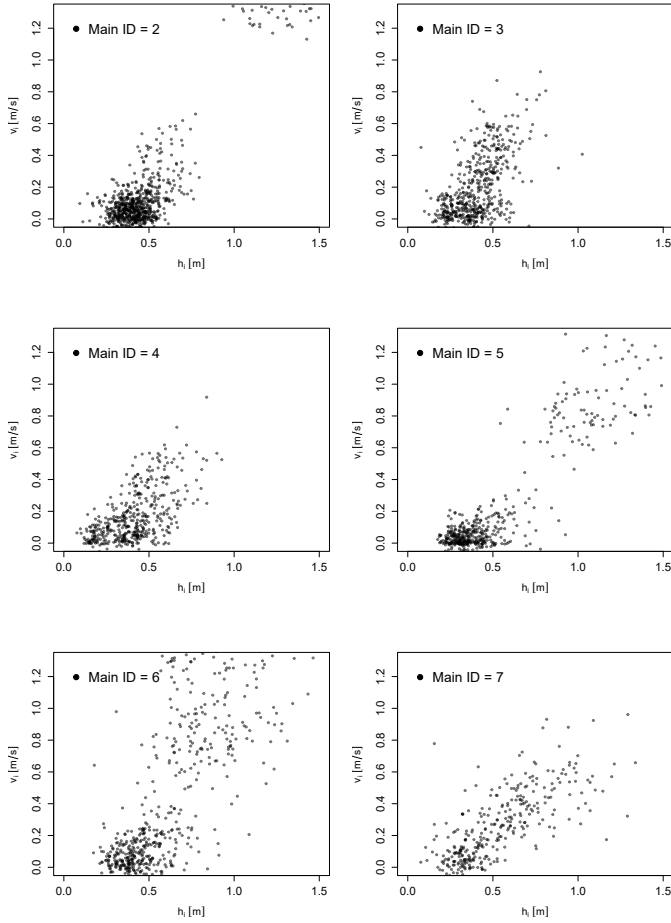


Figure A.2: Distance vs. speed relationship for several main IDs: 2, 3, 4, 5, 6, and 7 at GBS.

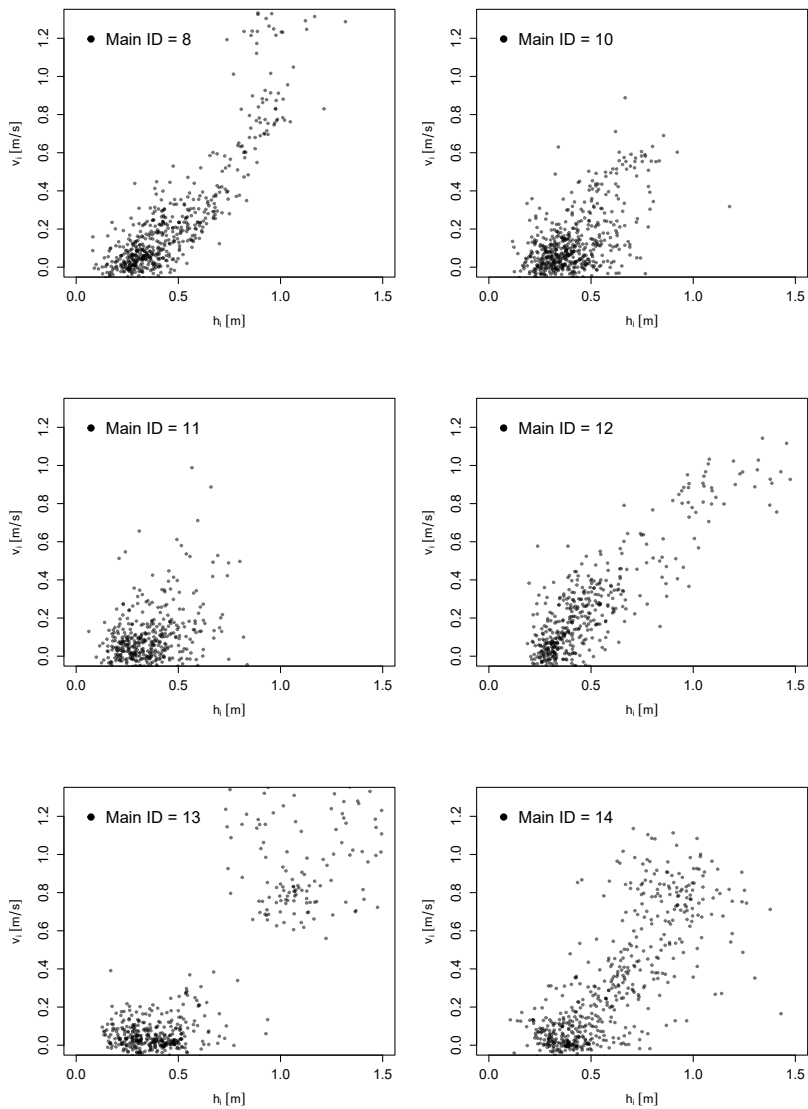


Figure A.3: Distance vs. speed relationship for several main IDs: 8, 10, 11, 12, 13, and 14 at GBS.

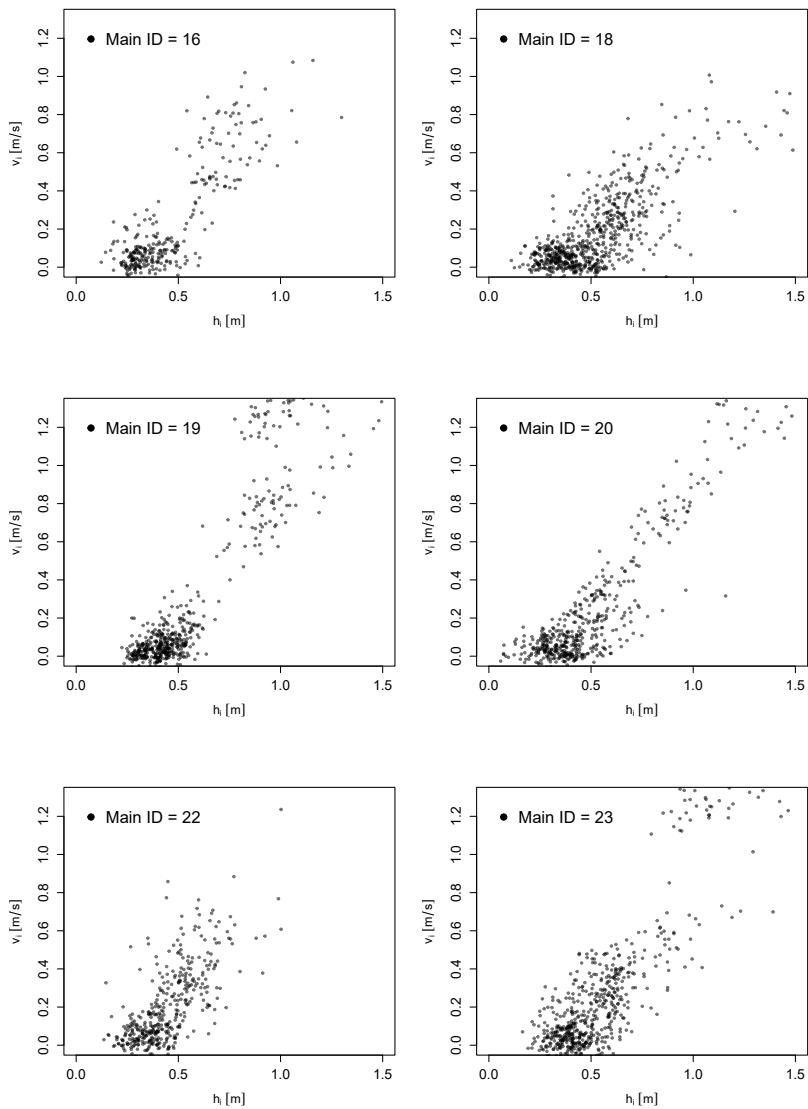


Figure A.4: Distance vs. speed relationship for several main IDs: 16, 18, 19, 20, 22, and 23 at GBS.

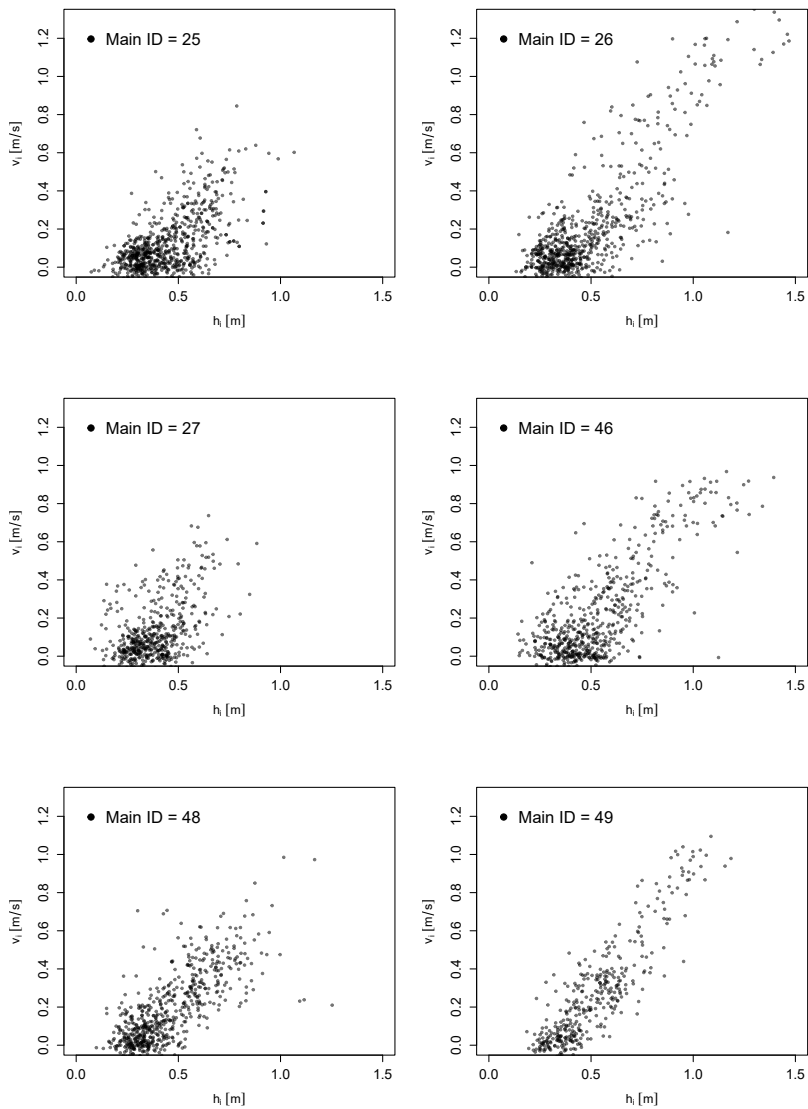


Figure A.5: Distance vs. speed relationship for several main IDs: 25, 26, 27, 46, 48 and 49 at GBS.

### A.3

The following figures illustrate the distance vs. speed relationship for several main IDs to show the approximately linear course for  $h < 1.5$  m for the data at MEH.

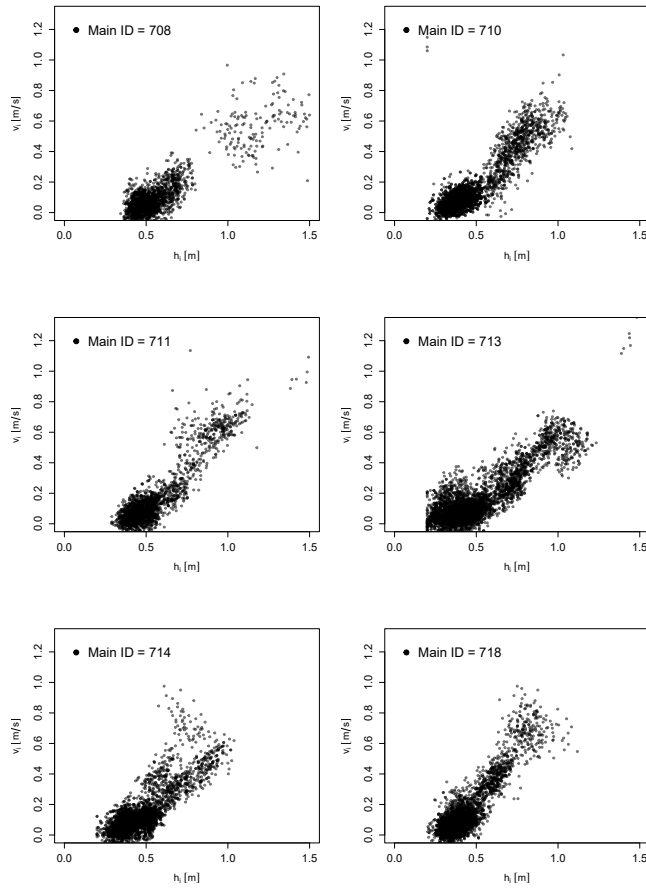


Figure A.6: Distance vs. speed relationship for several main IDs: 708, 710, 711, 713, 714, and 718 at MEH.



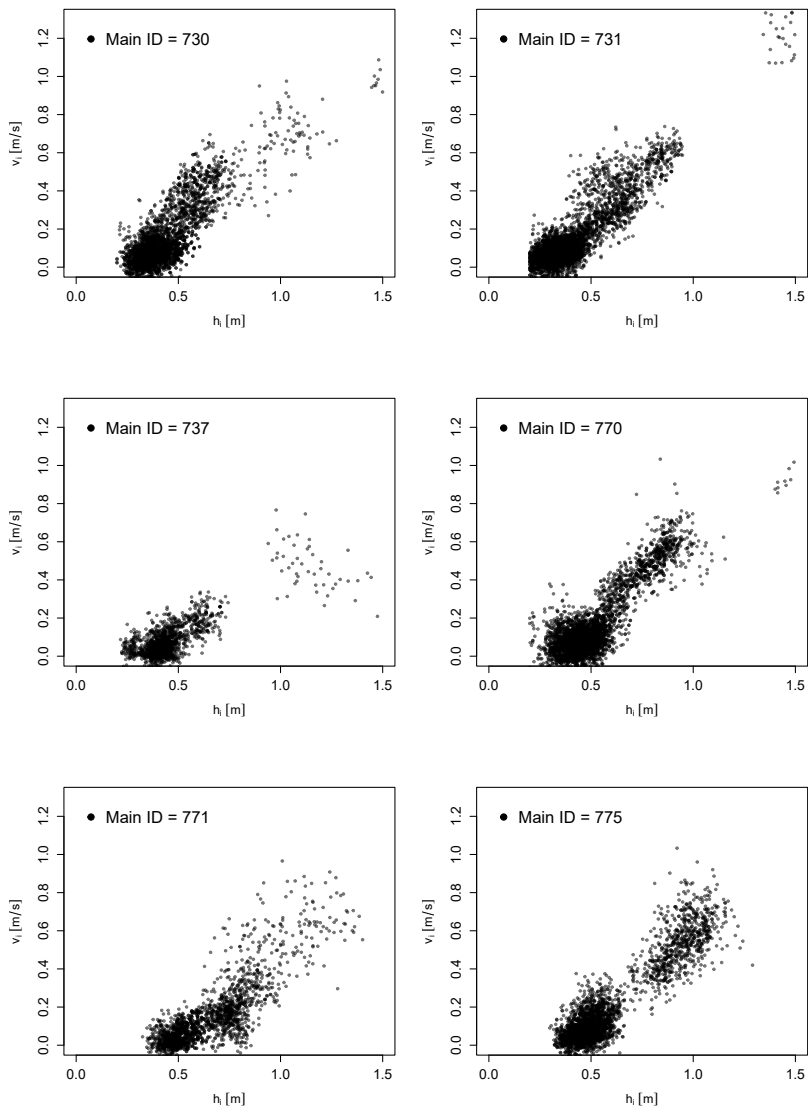


Figure A.7: Distance vs. speed relationship for several main IDs: 730, 731, 737, 770, 771, and 775 at MEH.

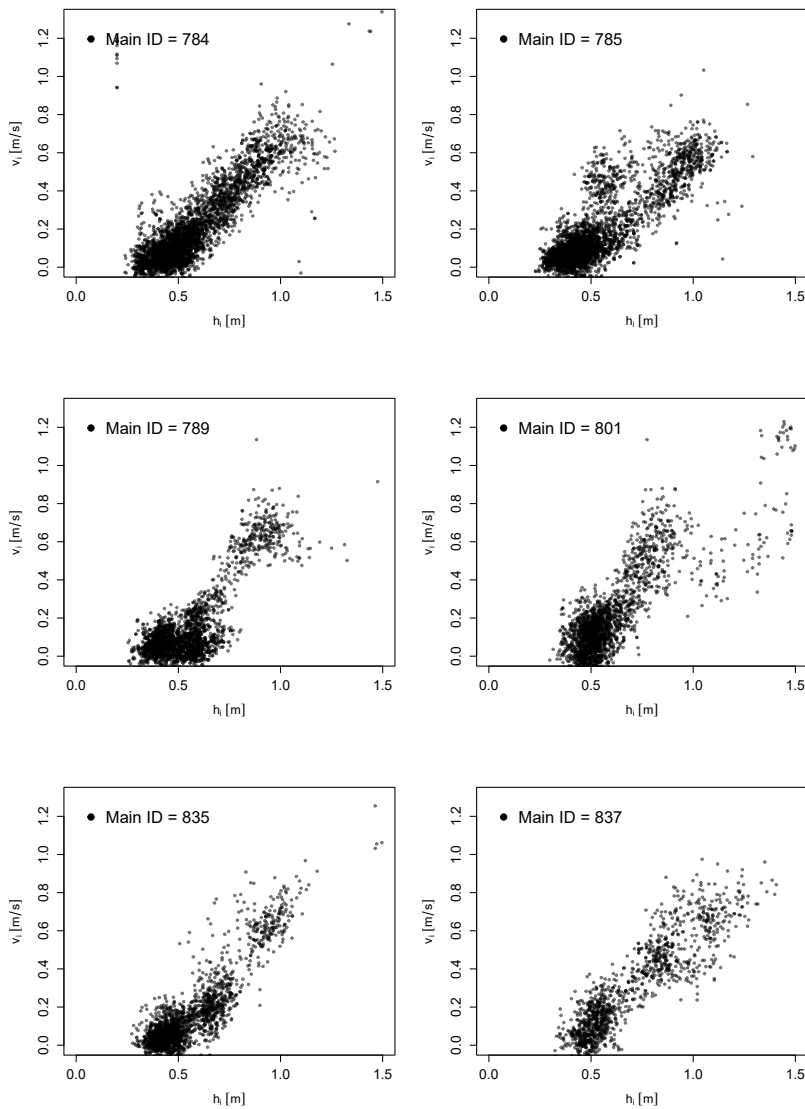


Figure A.8: Distance vs. speed relationship for several main IDs: 784, 785, 789, 801, 835, and 837 at MEH.

## A.4

The following figures illustrate velocity distributions for different gender compositions: male and gender alternating order for different global densities. Furthermore, the single velocity distributions for the various densities of the female group are shown in detail.

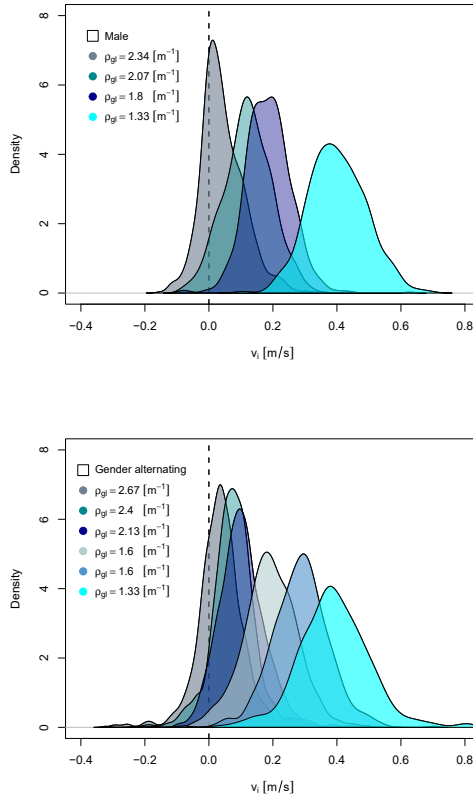


Figure A.9: Velocity distributions for different global densities. The figure above illustrates the male group composition, while the figure below illustrates a gender-alternating order.

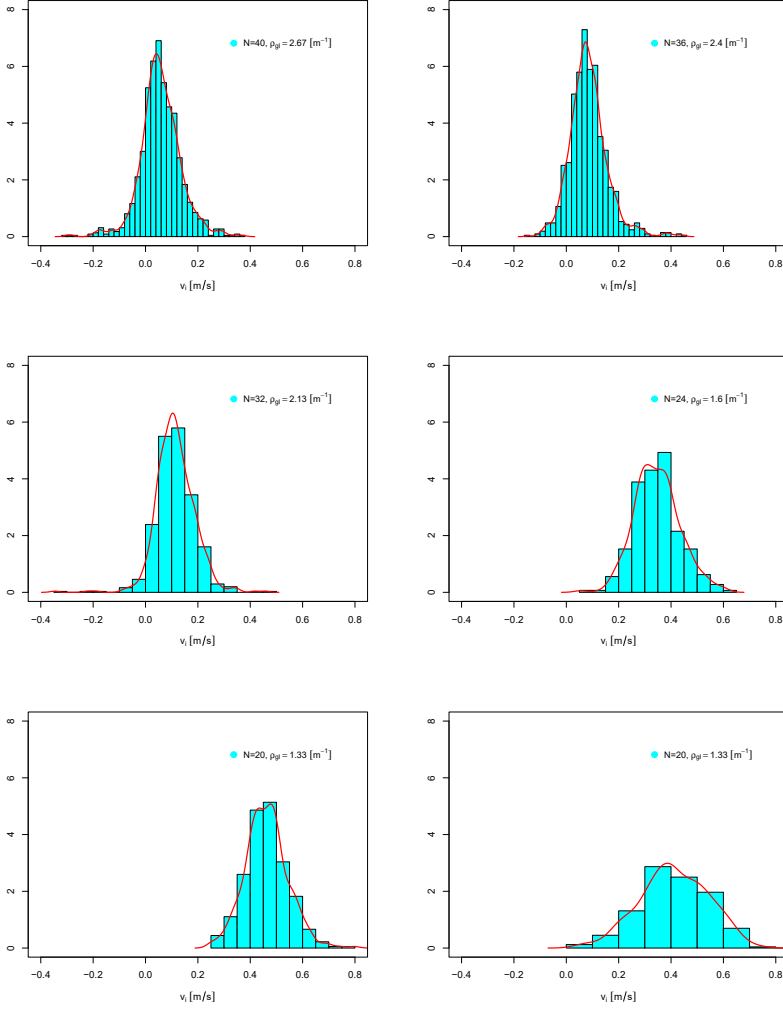


Figure A.10: Velocity distributions for the various densities of the female group are shown in detail.  $\rho_{gl} = 2.67 \text{ m}^{-1}$  top left,  $\rho_{gl} = 2.40 \text{ m}^{-1}$  top right,  $\rho_{gl} = 2.13 \text{ m}^{-1}$  middle left,  $\rho_{gl} = 1.60 \text{ m}^{-1}$  middle right,  $\rho_{gl} = 1.33 \text{ m}^{-1}$  bottom left and  $\rho_{gl} = 1.33 \text{ m}^{-1}$  bottom right.

## A.5

The following figures illustrate time-space diagrams for different gender compositions: male, gender random order and gender alternating order and densities. Red represents the stop state and blue is the go state.

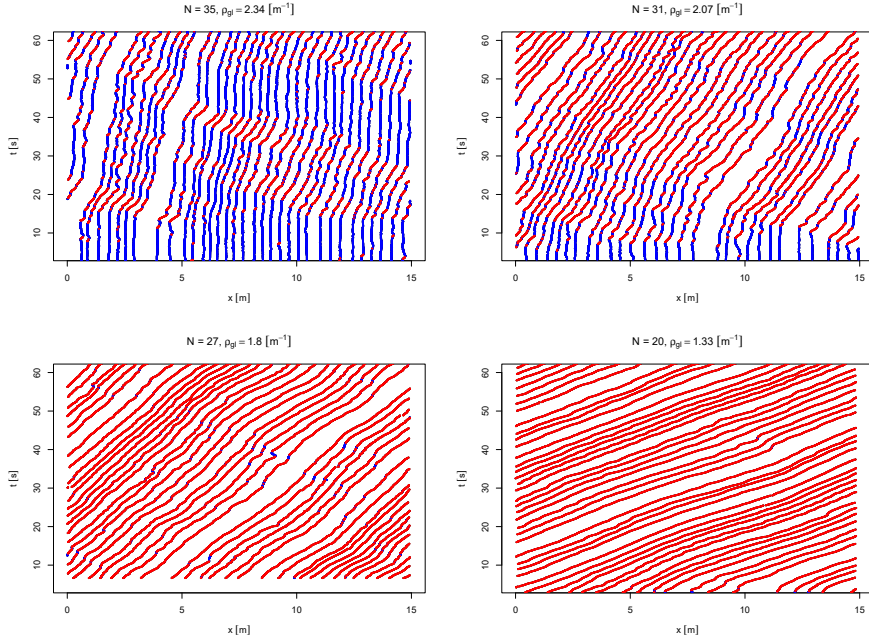


Figure A.11: Time-space diagrams for the group composition male for different global densities.  $\rho_{gl} = 2.34 \text{ m}^{-1}$  top left,  $\rho_{gl} = 2.07 \text{ m}^{-1}$  top right,  $\rho_{gl} = 1.80 \text{ m}^{-1}$  bottom left and  $\rho_{gl} = 1.33 \text{ m}^{-1}$  bottom right.

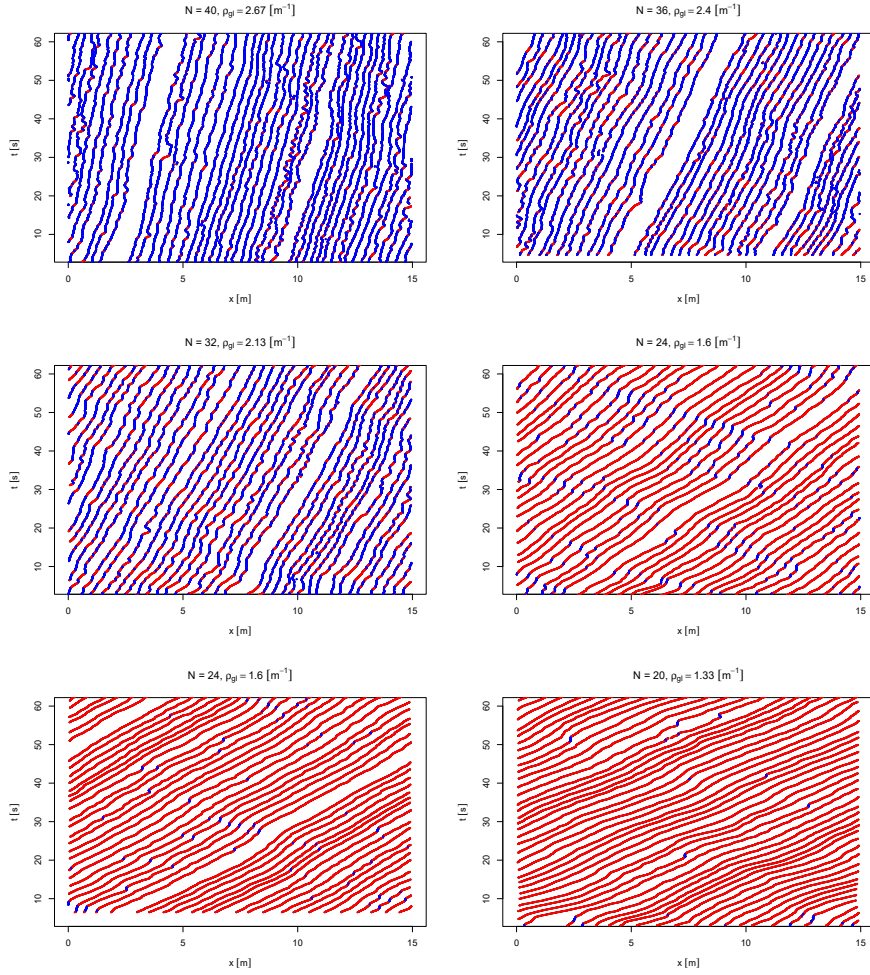


Figure A.12: Time-space diagrams for the group composition gender random order for different global densities.  $\rho_{gl} = 2.67 \text{ m}^{-1}$  top left,  $\rho_{gl} = 2.40 \text{ m}^{-1}$  top right,  $\rho_{gl} = 2.13 \text{ m}^{-1}$  middle left,  $\rho_{gl} = 1.60 \text{ m}^{-1}$  middle right,  $\rho_{gl} = 1.60 \text{ m}^{-1}$  bottom left and  $\rho_{gl} = 1.33 \text{ m}^{-1}$  bottom right.

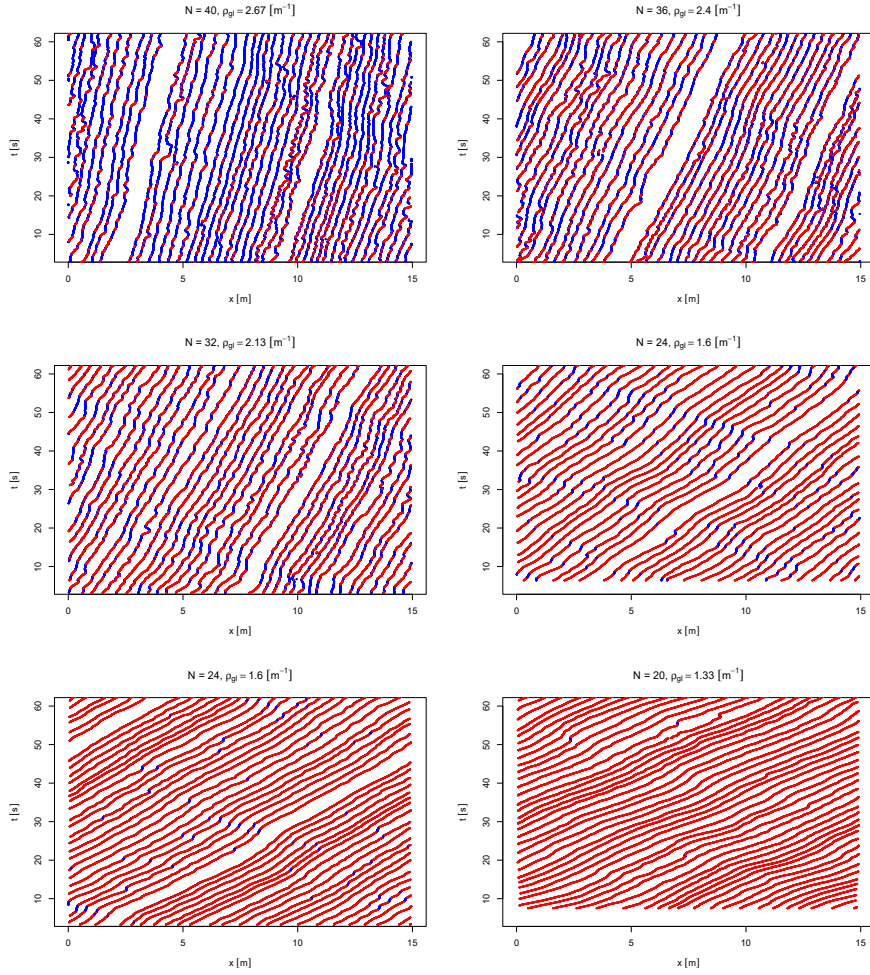


Figure A.13: Time-space diagrams for the group composition gender alternating order for different global densities.  $\rho_{gl} = 2.67 \text{ m}^{-1}$  top left,  $\rho_{gl} = 2.40 \text{ m}^{-1}$  top right,  $\rho_{gl} = 2.13 \text{ m}^{-1}$  middle left,  $\rho_{gl} = 1.60 \text{ m}^{-1}$  middle right,  $\rho_{gl} = 1.60 \text{ m}^{-1}$  bottom left and  $\rho_{gl} = 1.33 \text{ m}^{-1}$  bottom right.

## A.6

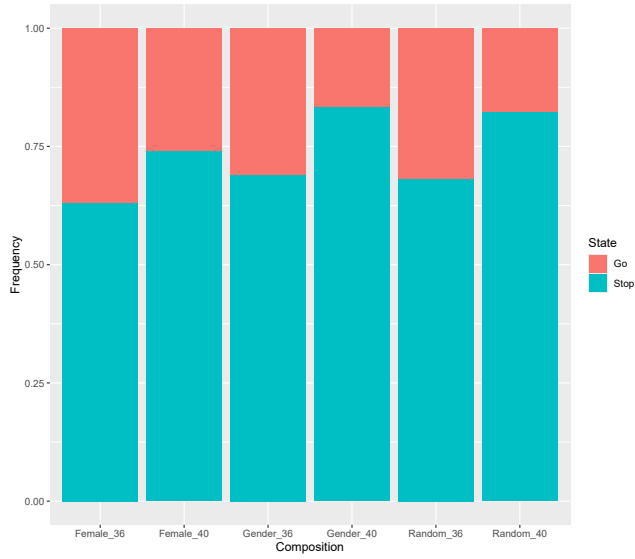


Figure A.14: The mean duration of a person's stay in the stop-and-go state for the group compositions female, gender alternating and gender random order for the global densities  $\rho_{gl} = 2.67 \text{ m}^{-1}$  ( $N = 40$ ) and  $\rho_{gl} = 2.40 \text{ m}^{-1}$  ( $N = 36$ ).



## A.7

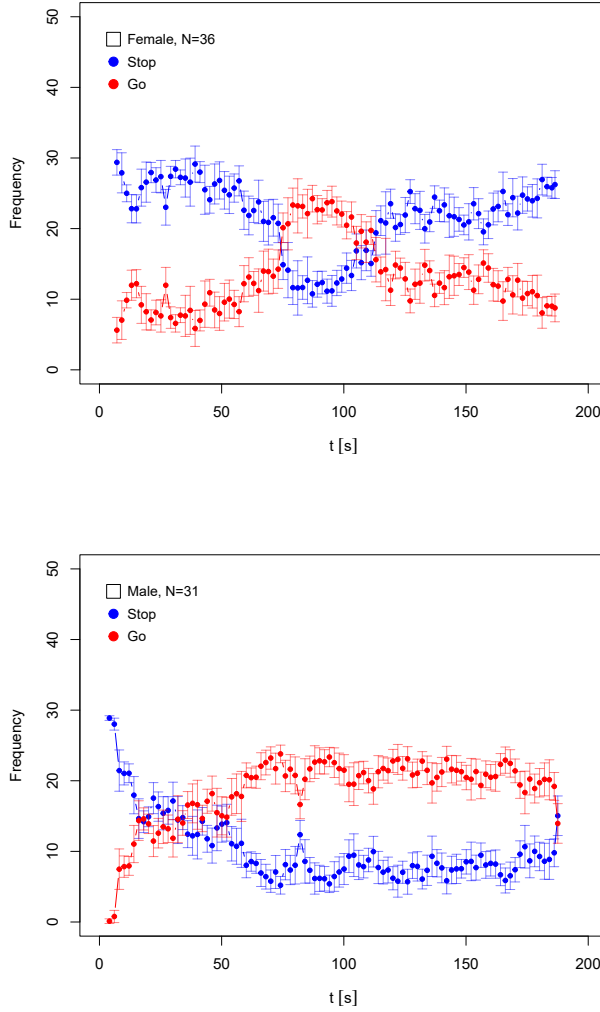


Figure A.15: The number of pedestrians in a stop-and-go state over time for the female group  $N = 40$  at the top and for the male group  $N = 35$  at the bottom. In the female group, the change between the states can be seen in the time interval of 75 to 125 seconds.

## A.8

Table A.1: An overview of experimental runs for different gender groups female, male, gender alternating, and gender random order at MEH in Germany. From the left to the right column, the following information is given: The total number of persons in a run, the global density  $\text{m}^{-1}$ , the beginning  $I_{min}$  and the end  $I_{max}$  of the steady state interval in seconds, and the mean values for individual specific time-steps  $\bar{\delta}_t$  in seconds.

	N	$\rho_{gl}$	$I_{min}$	$I_{max}$	$\bar{\delta}_t$
Female	40	2.67	6.52	185.48	0.67
	36	2.40	6.12	186.68	0.83
	32	2.14	1.16	122.28	0.73
	24	1.60	3.32	81.48	1.49
	20	1.34	7.48	122.96	1.32
	20	1.34	3.92	80.32	1.85
	16	1.07	0.84	114.08	1.53
	8	0.53	8.00	116.92	2.43
	4	0.27	2.04	118.64	1.69
Gender alternating	4	0.27	2.16	120.72	1.51
	4	0.27	3.40	45.28	1.04
	4	0.27	1.08	118.52	2.44
	4	0.27	9.96	89.20	3.78
	8	0.53	3.88	120.00	1.37
	8	0.53	3.96	122.68	1.51
	16	1.07	2.56	141.34	1.19
	20	1.34	5.08	120.32	1.56
	24	1.60	6.36	118.64	1.11
	24	1.60	3.20	123.60	1.05
	32	2.14	3.12	178.60	0.69
	36	2.40	0.76	182.12	0.66
	40	2.67	0.48	160.84	0.63
	4	0.27	1.76	119.88	0.96
	4	0.27	3.04	100.44	1.54
	4	0.27	1.48	119.56	0.92
	8	0.53	8.52	117.60	1.08
	4	0.27	1.68	117.84	0.88
	4	0.27	6.52	109.64	1.36
	8	0.53	6.08	120.60	0.88
	16	1.07	1.28	98.20	1.64
	20	1.34	0.20	120.12	1.52
	16	1.07	0.28	122.80	1.64
	8	0.53	0.64	178.56	1.16

	N	$\rho_{gl}$	$I_{min}$	$I_{max}$	$\overline{\delta_t}$
Male	35	2.34	0.52	190.16	0.76
	31	2.07	3.08	188.00	0.72
	27	1.80	6.60	121.16	0.84
	19	1.27	2.88	124.40	0.80
	15	1.00	0.68	117.12	0.76
	20	1.34	2.64	113.80	2.23
	16	1.07	0.48	120.40	1.96
	8	0.53	1.36	105.72	1.36
	8	0.53	1.40	103.20	1.16
	4	0.27	1.60	120.68	1.00
Gender random	4	0.27	3.64	116.92	1.39
	4	0.27	2.92	119.44	1.42
	4	0.27	1.84	115.72	1.12
	4	0.27	3.48	113.64	1.40
	8	0.53	2.84	119.44	1.41
	8	0.53	3.24	116.04	1.38
	16	1.07	3.28	113.68	1.75
	24	1.60	6.44	120.44	1.10
	24	1.60	1.48	121.32	0.96
	32	2.14	2.48	178.92	0.81
	36	2.40	4.60	182.88	0.78
	40	2.67	0.28	117.68	0.61
	4	0.27	0.48	117.36	2.56
	4	0.27	0.28	134.24	5.00
	4	0.27	8.68	105.32	1.48
	8	0.53	0.52	119.60	1.36
	4	0.27	1.24	112.64	1.36
	4	0.27	3.04	119.24	1.36
	8	0.53	1.36	71.60	1.36
	16	1.07	0.48	117.60	2.04
	20	1.34	0.48	110.64	0.76
	16	1.07	1.04	121.16	2.00
	8	0.53	10.00	176.64	1.56

---

# List of Publications

Publications:

I: S. Paetzke, M. Boltes, A. Seyfried, *Influence of individual factors on fundamental diagrams of pedestrians*, Phys. A Stat. Mech. Its Appl. 2022, 595, 127077.

<https://doi.org/10.1016/j.physa.2022.127077>.

II: S. Paetzke, M. Boltes, A. Seyfried, *Influence of Gender Composition in Pedestrian Single-File Experiments*, Appl. Sci. 2023, 13(9), 5450.

<https://doi.org/10.3390/app13095450>.

This dissertation contains material, such as figures and tables, already published in the above mentioned publications I and II. In addition, there is new content as well as content that was also included in papers I and II.

---

# Lebenslauf

## Persönliche Daten

Name	Sarah Paetzke (geb. Ammenhäuser)
Anschrift	Schirrmannweg 13, 42781 Haan
Geburtsdaten	08. Oktober 1992 in Wuppertal
Staatsbürgerschaft	Deutsch
Familienstand	verheiratet mit Michael Paetzke

## Bildungsweg

08.2003 – 06.2013	Abitur	Gymnasium Bayreuther Straße, Wuppertal
10.2013 – 08.2017	Bachelor-Studium	B. Sc. Wirtschaftsmathematik, BUW
	Thesis:	Bestimmung der Parameter von Modellen mit Zufallseffekten für individualisierte Fundamentaldiagramme in der Fußgängerdynamik
10.2017 – 05.2019	Master-Studium	M. Sc. Mathematik, BUW
	Thesis:	Statistische Analyse dynamischer Netzwerke in der Fußgängerdynamik
06.2019 – 08.2024	Doktorandin	IAS-7, Forschungszentrum Jülich

Band / Volume 55

**Gradient-Free Optimization of Artificial and Biological Networks  
using Learning to Learn**

A. Yeğenoğlu (2023), II, 136 pp

ISBN: 978-3-95806-719-6

Band / Volume 56

**Real-time simulations of transmon systems with  
time-dependent Hamiltonian models**

H. A. Lagemann (2023), iii, 166, XXX pp

ISBN: 978-3-95806-720-2

Band / Volume 57

**Plasma Breakdown and Runaway Modelling in ITER-scale Tokamaks**

J. Chew (2023), xv, 172 pp

ISBN: 978-3-95806-730-1

Band / Volume 58

**Space Usage and Waiting Pedestrians at Train Station Platforms**

M. Küpper (2023), ix, 95 pp

ISBN: 978-3-95806-733-2

Band / Volume 59

**Quantum annealing and its variants: Application to quadratic  
unconstrained binary optimization**

V. Mehta (2024), iii, 152 pp

ISBN: 978-3-95806-755-4

Band / Volume 60

**Elements for modeling pedestrian movement  
from theory to application and back**

M. Chraïbi (2024), vi, 279 pp

ISBN: 978-3-95806-757-8

Band / Volume 61

**Artificial Intelligence Framework for Video Analytics:**

Detecting Pushing in Crowds

A. Alia (2024), xviii, 151 pp

ISBN: 978-3-95806-763-9

Band / Volume 62

**The Relationship between Pedestrian Density, Walking Speed  
and Psychological Stress:  
Examining Physiological Arousal in Crowded Situations**

M. Beermann (2024), xi, 117 pp

ISBN: 978-3-95806-764-6

Band / Volume 63

**Eventify Meets Heterogeneity:  
Enabling Fine-Grained Task-Parallelism on GPUs**

L. Morgenstern (2024), xv, 110 pp

ISBN: 978-3-95806-765-3

Band / Volume 64

**Dynamic Motivation in Crowds: Insights from Experiments  
and Pedestrian Models for Goal-Directed Motion**

E. Üsten (2024), ix, 121 pp

ISBN: 978-3-95806-773-8

Band / Volume 65

**Propagation of Stimuli in Crowds:  
Empirical insights into mutual influence in human crowds**

H. Lügering (2024), xi, 123 pp

ISBN: 978-3-95806-775-2

Band / Volume 66

**Classification of Pedestrian Streams: From Empirics to Modelling**

J. Cordes (2024), vii, 176 pp

ISBN: 978-3-95806-780-6

Band / Volume 67

**Optimizing Automated Shading Systems in Office Buildings by  
Exploring Occupant Behaviour**

G. Derbas (2024), 9, x, 168, ccxxiii

ISBN: 978-3-95806-787-5

Band / Volume 68

**Speed-Density Analysis in Pedestrian Single-File Experiments**

S. Paetzke (2025), XIII, 107 pp

ISBN: 978-3-95806-818-6

Weitere *Schriften des Verlags im Forschungszentrum Jülich* unter  
<http://wwwzb1.fz-juelich.de/verlagextern1/index.asp>





IAS Series  
Band / Volume 68  
ISBN 978-3-95806-818-6