Modeling crowd behavior

Natalie Fridman
Computer Science Department

Ph.D. Thesis

Submitted to the Senate of Bar-Ilan University
Ramat Gan, Israel
May 2012

This work was carried out under the supervision of Prof. Gal A. Kaminka, Computer Science Department, Bar-Ilan University.

Acknowledgments

It is very difficult to put into words the appropriate acknowledgments that reflect my appreciation and gratitude to all the people with whom I had the great honor to work and to my family and friends who always support me in every decision that I make.

First and foremost, I would like to express my deepest gratitude to my advisor Professor Gal A. Kaminka, for his constant support, outstanding guidance and excellent advice throughout this research. He has been everything that one could want in an advisor and I consider myself very lucky and most honored to have been one of his students. Gal has always provided me with direction while allowing total freedom to explore new ideas and he has always managed to keep me in focused. Gal's continuous support has not only been from a scientific perspective but also from a personal one. I admire his energy, enthusiasm and generosity. I really look forward to continue working with him.

I would like to thank the Maverick group for their friendly, supportive and particularly fun atmosphere. I would like to especially thank Inbal Rika, Avishay Zilka, Tomer Zilbershtain and Lior Spyer for their collaboration on different projects. It has always been a pleasure to work with them.

My deepest appreciation to the Teamcore Research Group especially Professor Milind Tambe and Jason Tsai for their great collaboration on the Evacuation project and for the interesting and fruitful discussions we held. It has been a great honor to work with them.

My deepest appreciation to my parents and my brother for their love, support and encouragement throughout my entire life. I would like to thank them for providing me a warm home and enabling me to maintain a clear mind for me so that I could concentrate on my studies by always ensuring that I would have no financial worries.

Most importantly, my deepest thanks to the most wonderful person in my life, my closest friend and my husband, Alex. I would like to thank him for all his love, for being patient with me and for tolerating my long working hours. His support and encouragement were a large part of what kept me going. Many thanks to my children Alon and Maya who have made my life complete. My family has always been a great source of strength for me and most importantly, they make my life more enjoyable.

Finally, I would like to express my gratitude to the Israeli Ministry of Defense especially to Head of HFE Branch LTC Michal Rottem-Hovev for providing partial support for this research.

This research was supported in part by ISF Grant #1357/07

Contents

List of Figures

List of Algorithms

Abstract

Modeling crowd behavior is an important challenge for cognitive modelers, multi-agent systems and social simulation. In this dissertation we explore several techniques for modeling crowd behaviors and address important challenges concerning each technique.

In the first part of the work we use the agent based approach for modeling crowd behavior. We present the extended Social Comparison model (SCT), which we believe is a general cognitive process underlying the social behavior of each individual in a crowd. In this work we provide a qualitative evaluation of the SCT model, as well as others, in contrast to human pedestrian behavior. The results clearly demonstrate that the SCT model is superior to others in its fidelity to human pedestrian behavior. Moreover, we have focused on an open question which has arisen from the SCT model, namely *when* should the SCT process be used at the architectural level in order to guide action-selection of agents. We have extended the SCT model to address this open question. We argue that comparisons take place all the time (i.e., differences are perceived and processed), but the cognitive architecture limits actions taken to minimize differences in cases where the comparisons yield significant differences. In addition we examine the impact of cultural differences on the macro level behavior produced in pedestrian and evacuation domains. In this work we have advanced by treating culture as a first-class object in models of physical crowds and have extended the SCT model accordingly. We introduced cultural individual-level parameters into the simulations, and then examined the effects of these individual level parameters on the emergent crowd dynamics.

In the second part of the work we use the qualitative reasoning approach for modeling demonstrations. We present a first attempt to use qualitative reasoning techniques in order to model crowd behaviors. To the best of our knowledge, such techniques have never been applied to modeling and reasoning regarding crowd behaviors, nor in particular demonstrations. We have developed qualitative models consistent with the partial, qualitative social science literature, which has enabled us to model the interactions between different factors that influence violence in demonstrations. We then utilized the qualitative simulations to predict the potential eruption of violence, at various levels, based on a description of the demographics, environmental settings, and police responses. In addition to providing predictions, the resulting qualitative simulation graphs were analyzed to determine the factors

that are most important in influencing the outcome. These factors can be used to support decision-makers.

Chapter 1

Introduction

A crowd is a large group of people who are in close geographical or logical states. Individuals in crowds are affected by each other's presence and actions, often acting in a seemingly coordinated fashion, as if governed by a single mind. However, this coordination is achieved with little or no verbal communication.

Modeling crowd behavior is an important challenge for cognitive modelers, multi-agent systems and social simulation. Models of crowd behavior facilitate analysis and prediction, are sought in training simulations [75], safety decision-support systems [12], traffic management [37, 68], business and organizational science.

Computer simulation is considered to be a promising approach for modeling and reasoning of different social phenomena [26] in particular crowd behavior. There are several micro (individual) and macro (group) level techniques that enable such modeling (see e.g., exp. 3.3.1, 8.1). Using a micro level approach we model the behavior of each individual and his/her responses to other agents. Collective behavior is the compound of many individuals' behaviors. Using a macro level approach we can model collective behavior by modeling macro level interactions between macro level factors.

Nonetheless, there are many open challenges. Micro level approaches require detailed individual modeling and are difficult to use in domains where such detailed information is unavailable. Moreover, existing models for this approach are usually not validated against human data and do not yet account for cultural factors affecting crowd behavior, and even more so, for crowds composed of members of different cultures. Macro level approaches enable reasoning about produced social behavior by modeling macro level interactions between different macro level factors. However, these techniques are not yet applied for modeling crowd behaviors.

In this dissertation we explore several techniques for modeling crowd behaviors and address important challenges in each. We use both an agent based approach (micro level) and a qualitative reasoning (macro level) approach.

In the first part of the work we use the agent based approach for modeling crowd behavior. We present the extended Social Comparison model (SCT), which we believe is the general cognitive process underlying the social behavior of each individual in a crowd. We examine this model in pedestrian dynamics and evacuation behavior. We validate this model against human data. We also examine the question of *when* social comparison is triggered at the architectural level. In addition we examine the impact of cultural differences on the produced macro level behavior in pedestrian and evacuation domains.

In the second part of the work we use the qualitative reasoning approach for modeling demonstrations. We present a pioneer attempt to use qualitative reasoning techniques in order to model crowd behaviors. To our knowledge, such techniques have never been applied to modeling and reasoning regarding crowd behaviors, nor in particular demonstrations. We have developed qualitative models consistent with the partial, qualitative social science literature, which allows us to model the interactions between different factors that influence violence in demonstrations. We then utilize qualitative simulations in order to predict the potential eruption of violence, at various levels, based on a description of the demographics, environmental settings, and police responses. In addition to providing predictions, the resulting qualitative simulation graphs are analyzed in order to determine the factors most influential on the outcome. These factors can assist decision-makers.

In the following sections we discuss the contributions of the dissertation in more detail. In Section 1.1 we describe our contributions to modeling crowd behavior using an agent based approach. In Section 1.2 we introduce our work on modeling crowd behavior using a qualitative reasoning approach. We present the thesis structure in Section 1.3 and our publications in Section 1.4.

1.1 Modeling crowd behavior using an agent based approach

Using agent-based simulations, we simulated social behaviors by modeling each individual, and their individual responses to each other. By modeling agents' social cognition, we have the ability to model complicated social interactions. Such simulations have been successfully used in modeling crowd behaviors [24, 44], economic phenomena [83], and more. It is a bottom-up approach in the sense that in order to observe macro-level behavior we must model the micro-level interactions which necessitate detailed individual modeling. However, such micro level resolution enables us to examine complicated social interactions among the individuals in addition to the emergent macro level behavior.

For several decades researchers have developed micro level computational models for simulation of collective behavior. Indeed, a variety of computational models have been proposed that exhibit crowd-like behavior in different tasks. For instance, cellular automata models are used to model pedestrian movements [8, 37] or people evacuating an area during an emergency [36, 46].

The biggest challenge in modeling crowd behaviors is in their evaluation process. Unfortunately, only a handful of existing models of crowd behavior have been evaluated against real-world human crowd data. The main difficulty is the lack of human data for the evaluation of these models. Moreover, essentially no computational cognitive models have been proposed which are tied to the cognitive science theory. Instead, *existing models are domain-specific*; the algorithms used change from one crowd behavior to the next. No attempt has been made to construct a single cognitive mechanism which accounts for all of the different behaviors.

Recently, we presented an innovative cognitive model of crowd behavior [24], which has two key novelties (compared to previous models): First, there is a single computational mechanism (algorithm) used to generate different crowd phenomena and second, it is based on a prominent social psychological theory. In particular, the model is based on Social Comparison Theory (*SCT*) [21], a popular social psychological theory that has been continuously evolving since the 1950s. The key idea in SCT is that humans, lacking objective means to evaluate their state, compare themselves to others that are similar to them.

We believe that social comparison is a general cognitive process underlying the social behavior of each individual in a crowd. Unlike previous crowd models that concentrate on specific behavior, the SCT model can account for different crowd behaviors, depending on the perceptions and actions available to each individual [24].

However, while the SCT model has been proven to be superior to other computational models in behavior-specific measures (e.g., the formation of lanes in bidirectional movement), it has never been validated against human crowd data.

In this work we provide a qualitative evaluation of the SCT model, as well as of others, in comparison to human pedestrian behavior. Moreover, we expand the SCT model to account for the timing of the SCT process at the architectural level and also for cultural differences among the agents.

In Chapter 3 we evaluate the SCT model on general pedestrian movement which includes individuals, couples, and groups, all walking at different speeds. We compare the performance of the model to other popular models from the literature and explore the impact of different parameters and model components (e.g., bounds) on the generated behavior. The evaluation was carried out by 39 human subjects who compared the behavior generated from the different models to videos of real-world pedestrians. The results clearly demonstrate that the SCT model is superior to others in its fidelity to human pedestrian behavior.

In Chapter 4, we focus on a specific open question emerging from the SCT model, namely the question of *when* the SCT process should be used at the architectural level in order to guide action-selection in agents. In particular, social psychological theory advocates a model in which social comparison occurs only when the agent lacks objective means to evaluate its own progress [21]. This approach, in which the social comparison process is triggered only when the agent is uncertain as to how to pursue its task goals, works successfully when used to simulate bi-directional pedestrian movement. However, it fails when modeling uni-directional movement; here, an approach in which agents compare themselves to others at *all times* is preferable. This on-going comparison approach is also supported by evidence from social psychology and economics.

Earlier successful demonstrations of the fidelity of the SCT model were thus switched between different triggering mechanisms ad-hoc; In Section 4.1.2 we present the results of experiments which demonstrate that the two triggering mechanisms are mutually-exclusive. This interferes with a clear understanding of how social comparison processes are used within a cognitive architecture. Clearly, social comparison at the architectural level cannot be changed based on the domain.

As part of the research, we hypothesized that social comparison processes are indeed on-going, and that humans are aware of—and compare themselves to—others at all times. Our hypothesis is that action selection mechanisms in the cognitive architecture are responsible at times for selecting actions which minimize social differences (i.e., act on the social comparison results), and at other times, for selecting actions that serve other goals. We experimented with two alternative mechanisms, and we were able to rule out one of them. This result places constraints on the cognitive architecture mechanisms that are used in social comparison.

Additionally, we examined alternative SCT algorithms, extending and refining the SCT algorithm published earlier (in [24]). An important motivation for this was the fact that the previously published algorithm ignored the group size while executing comparisons. There is evidence from social psychology that shows that in fact the size of the group is an important factor in the individual's social action-selection. The refined algorithms are presented and examined experimentally.

We also examined the impact of cultural differences on the resulting macro level crowd dynamics in pedestrian and evacuation domains. Unfortunately, existing models of physical crowds do not yet account for cultural factors. Social science literature on the effects of culture on physical crowds is extensive when it comes to individual interactions (e.g., personal spaces and speed), but rarely addresses the macro-level phenomena (e.g., pedestrian flow). As a result, it is difficult to validate models against data. This is particularly true of mixed-cultural physical crowds, in which the evolving crowd dynamics from individual interactions is inherently difficult to predict. Furthermore, to date agent-based models have ignored cultural differences in physical crowd models (e.g., in pedestrians), and treat all individuals as culturally homogeneous, who adjust to cultural parameters ad-hoc.

In this work we have taken a step towards treating culture as a first-class object in models of physical crowds. We have examined the impact of cultural differences on crowd dynamics in pedestrian and evacuation domains, using proven agent-based simulations of the two domains. We introduced cultural individual-level parameters into the simulations, and then examined the effects of these individual level parameters on the emergent crowd dynamics. Moreover, we examined the effects of mixing individuals with different cultural parameters in the same physical crowd.

In the pedestrian domain (Chapter 5), we related the resulting culturally-aware simulation to pedestrian data which we recorded from videos of pedestrians in five different countries: Iraq, Israel, England, Canada, and France. We characterized these cultures along four individual- level parameters: personal spaces, speed, avoidance side (i.e., which side is preferred when avoiding an oncoming pedestrian), and group formations. We used well-known crowd-level quantitative measures (e.g., flow, number of collisions, and mean speed) to identify crowd- level effects. We show that the model can faithfully replicate the observed pedestrian behavior in these videos.

In the evacuation domain (Chapter 6), we examined individual cultural parameters (documented in social science literature) with reference to how seriously people treat possible threats, their tendency to notify others, and their tendency to form groups. We then used the simulations to explore the impact of these tendencies on the resulting crowd behavior (measured quantitatively in times of evacuation, panic levels, etc.).

1.2 Modeling crowd behavior using a qualitative reasoning approach

We propose Qualitative Reasoning (QR) as a macro-level approach which enables modeling and reasoning without simulating interactions between individuals. By using a macro level approach we can reason about produced social behavior by modeling macro level interactions between different macro level factors. Although it may seem that this resolution is too low and therefore difficult to deduce useful conclusions, QR approaches have been found to be efficient for modeling and reasoning for macro level processes. Moreover, there are domains where high resolution such as agent-based modeling may be unavailable and even unnecessary.

In this section we use a macro level approach for modeling crowd behavior in particular demonstration behavior. Indeed violent demonstrations with many casualties among participants, police forces and innocent bystanders, unfortunately are not rare. Can we know the potential nature of an event before it ends with tragic outcomes? Accurate predictions regarding the potential violence level of a demonstration can improve the decision making process of the police, and decrease the number of casualties.

In general, there are several ways that can improve the decision making process of the police: the use of expert consultants, numeric simulations (e.g., [62]) and agent based simulations [44]. The use of expert psychologists and sociologists for consultations is a very common approach. However such consultation before every demonstration is neither a practical nor affordable solution. The use of numerical simulations (e.g., [62]) requires full information regarding the domain, which in many cases such as reasoning regarding the violence level in demonstrations, simply does not exist. Agent-based simulation approaches (e.g., [44]) require modeling at the micro (individual) level which in many cases may be impractical. Indeed in general, one of the biggest challenges in the crowd modeling field is the lack of precise data. Such precise data is even difficult to obtain since crowd experiments are complex and costly.

This is not to say that social sciences do not have deep and extensive knowledge of demonstrations. On the contrary, significant literature exists on the factors that impact violence during demonstrations. However, the literature mainly reports partial, macro-level qualitative descriptions of the influential factors. Integrating these together is a formidable challenge that requires novel forms of modeling.

To enable modeling and reasoning with imprecise and partial information we propose qualitative reasoning techniques. Qualitative Reasoning (QR) [22,47] is a sub-area of artificial intelligence (AI), which enables modeling and reasoning with partial or imprecise numeric information. The QR approach shows that while having a qualitative representation of data and order values (such as little/medium/large), it is still possible to draw useful conclusions.

In this work (Part II) we describe a novel application of QR for modeling and reasoning about the potential violence level during demonstrations. We have developed qualitative models consistent with the imprecise descriptions in social science literature, which allows us to model the interactions between different factors that influence violence in demonstrations. We have developed three separate models, which incrementally increase in complexity, and in the number of factors each considers. These three models were evaluated on real-world scenarios, using news reports and Wikipedia entries as the source of information as to the values of different quantities.

We have developed an innovative technique, which considers the number of paths leading to different violent outcomes. Using this technique, we are able to provide an estimate of the likelihood of different outcomes, for each test case. We have compared the predictions of the different models, and we have demonstrated an important benefit of using QR for social simulation modeling, i.e., the ability to easily test social science theories on real-world data.

Moreover, we have examined whether decision trees, a popular machine learning approach, can be used for qualitative predictions. While the results show that the decision trees provide accurate predictions (slightly better than our QR models) they lack the ability to support hypothetical what-if reasoning, because they do not have the explanatory power of a social science model. Thus, we claim that using QR is better for reasoning in this task.

Finally, we have developed an algorithm which analyzes the qualitative simulation graph of each test-case, to determine the factors that are most important in influencing the outcomes of the specific case under consideration. The key to this algorithm is to determine simulation graph nodes with high outcome entropy, i.e., nodes which lead to different outcomes, at fairly uniform likelihoods. In the states corresponding to such nodes, it is possible to identify actionable factors that can be used to influence the outcomes. We show that for real-world cases, the algorithm identifies the causes also identified by experts.

1.3 Thesis overview

This dissertation comprises 9 chapters, organized into two main parts (see Figure 1.1). This chapter presents the introduction to this thesis and the next chapter surveys the related work. Chapters $3-6$ constitute Part 1 of the dissertation, which deals with modeling crowd behavior using an agent-based approach. Part 2 of the dissertation presents the macro level approach for modeling crowd behaviors which is a Qualitative Reasoning approach and is presented in Chapters 7 and 8. In Chapter 9 we provide our conclusions and discuss future work.

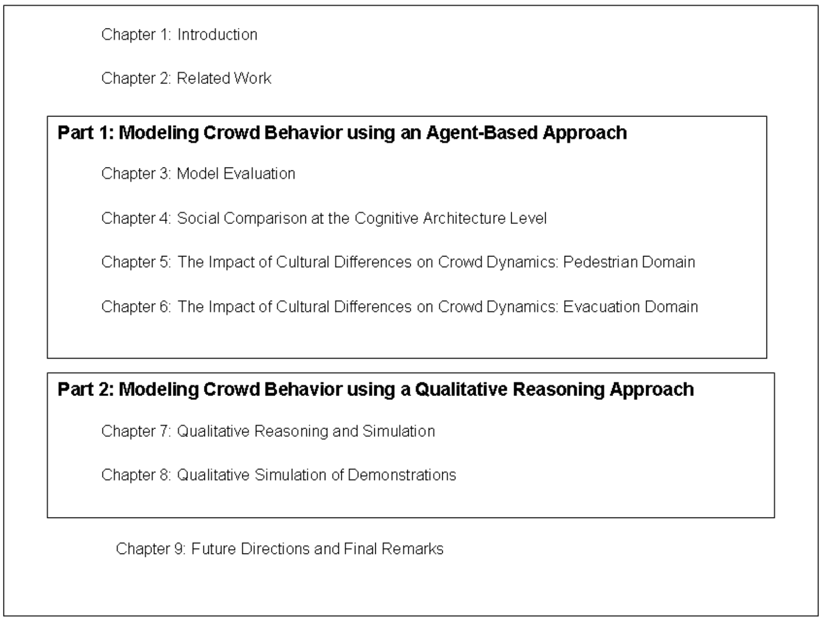


Figure 1.1: Thesis Structure.

1.4 Publications

Results that appear in this dissertation have been published in the following proceedings, journals, refereed conferences, books and workshops:

- Natalie Fridman and Gal A. Kaminka. Using Qualitative Reasoning for Social Simulation of Crowds. In ACM Transactions on Intelligent Systems and Technology, 2012, In press.
- Natalie Fridman, Avishy Zilka, and Gal A. Kaminka. The impact of cultural differences on crowd dynamics in pedestrian and evacuation domains: An Extended Abstract. In Proceedings of the Eleventh International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-12), 2012. Short paper.
- Natalie Fridman and Gal A. Kaminka. Towards a Computational Model of Social Comparison: Some Implications for the Cognitive Architecture. In Cognitive Systems Research, 12(2):186-197, 2011.
- Natalie Fridman, Tomer Zilberstein and Gal A. Kaminka. Predicting Demonstrations' Violence level using Qualitative Reasoning. In International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction (SBP-11), 2011.
- Jason Tsai, Natalie Fridman, Emma Bowring, Matthew Brown, Shira Epstein, Gal Kaminka, Stacy Marsella, Andrew Ogden, Inbal Rika, Ankur Sheel, Matthew Taylor, Xuezhi Wang, Avishay Zilka, Milind Tambe. ESCAPES: Evacuation Simulation with Children, Authorities, Parents, Emotions, and Social Comparison. In International Conference on Autonomous Agents and Multiagent Systems (AAMAS-11), 2011.
- Natalie Fridman, Gal A. Kaminka and Avishay Zilka. Towards Qualitative Reasoning for Policy Decision Support in Demonstrations. Accepted to Agent-based Modeling for PoLicy Engineering (AMPLE) workshop at AAMAS 2011.
- Gal A. Kaminka and Natalie Fridman. Using Qualitative Reasoning for Social Simulation of Crowds: A Preliminary Report. In Qualitative Reasoning (QR) workshop, 2011.

- Natalie Fridman and Gal A. Kaminka. Modeling Pedestrian Crowd Behavior Based on a Cognitive Model of Social Comparison Theory. In Computational and Mathematical Organizational Theory, Volume 16, Issue 4, Page 348. Special issue on Social Simulation from the Perspective of Artificial Intelligence. 2010.

- Natalie Fridman and Gal A. Kaminka. Comparing Human and Synthetic Group Behaviors: A Model Based on Social Psychology. In International Conference on Cognitive Modeling (ICCM-09), 2009. A previous version of this paper was presented at the proceedings of the Multi-Agent Based Simulation (MABS) workshop at AAMAS-2009.

- Natalie Fridman, Gal A. Kaminka and Meytal Traub. First Steps Towards a Social Comparison Model of Crowds. In International Conference on Cognitive Modeling (ICCM-09), 2009. A slightly different version of this paper was also presented at the Social Simulation workshop at IJCAI 2009.

- Jason Tsai, Emma Bowring, Shira Epstein, Natalie Fridman, Prakhar Garg, Gal Kaminka, Andrew Ogden, Milind Tambe, and Matthew Taylor. Agent-based Evacuation Modeling: Simulating the Los Angeles International Airport. In Workshop on Emergency Management: Incident, Resource, and Supply Chain Management (EMWS-09), 2009.

Chapter 2

Related Work

2.1 A micro-level approach to modeling crowd behavior

Social psychology literature provides several views on the emergence of crowds and the mechanisms underlying their behaviors. These views can inspire computational models, but are unfortunately too abstract to be used algorithmically. In contrast, computational crowd models tend to focus on specific crowd behaviors (e.g, flocking). A common theme in all of them is the generation of behavior from the aggregation of many local rules of interaction, e.g., [46, 66–68].

Moreover, understanding and modeling cultural differences in crowd behavior is an important challenge for social and exact science researchers. Social psychology literature provides several views on the cultural differences in micro level interactions among groups of people, but they usually do not examine the influence of these differences on the resulting macro level behavior such as pedestrian flow. Exact science researchers have been inspired by social psychology literature to develop computational models for crowd behaviors, but their focus has been on predicting the resulting macro level behavior from micro level interactions. Nonetheless, to the best of our knowledge, existing computational models for crowd behaviors have not yet taken cultural differences into account.

2.1.1 Social psychology

A phenomenon observed in crowds, and discovered early in crowd behavior research is that people in crowds act similar to one another, often acting in a coordinated fashion, as if governed by a single mind [1, 6, 9, 49]. However, this coordination is achieved with little or no verbal communication.

Le Bon [49] emphasized a view of crowd behaviors as controlled by a "Collective Mind", and observed that an individual who becomes a part of the crowd transforms to become identical to the others in the crowd. Le Bon explains the homogeneous behavior by two processes: (i) *Imitation*, where people in a crowd imitate each other; and (ii) *Contagion*, where people in a crowd behave differently from how they typically would, individually.

Blumer [9] explains that this coordinated crowd behavior occurs through a "circular reaction" process which underlies each individual who participates in the collective behavior. According to Blumer, "circular reaction" is: "a type of interstimulation wherein the response of one individual reproduces the stimulation that has come from another individual and in being reflected back to this individual reinforces the stimulation.".

According to Allport, crowd behavior is a product of the behavior of like-minded individuals. Allport's explanation of crowd homogeneous behavior is that similar people act in similar ways; otherwise they would not be a part of the same group. However, individual behavior is affected by the behavior of his surroundings, thus, according to Allport, "the individual in the crowd behaves just as he would behave alone, only more so."

Turner and Killian [80] investigated the Emergent-norm Theory, which hypothesizes that crowd members indeed imitate each other, but also create new norms for the crowd as the dynamics of the situation dictate. Thus while crowds are not entirely predictable, their collective behavior is a function of the decision-making processes of their members.

Berk's [6] explanation of crowd behavior is based on the decision making theory. According to the decision-theory, each individual always tries to maximize his reward and minimize costs. Berk argues that crowd behaviors are no exception, and that they should be understood from a game-theoretic perspective. He explained coordinated behavior of crowds as consistent with agents using a minimax strategy where the greater the number of participants who engage in a specific action, the less the cost will be for the individual engaging in the action. Thus, each individual will select the action of the majority.

Different theories provide different explanations as to what drives individual behavior when the individual is a part of a crowd. However, there is a consensus among all the theorieswhich asserts that when an individual is part of a crowd, his or her individual behavior is affected by others.

We based our work on the Social Comparison Theory [21], which (to the best of our knowledge) has never been applied to modeling crowd behavior. Nevertheless, as we show in section 3.1, key elements of the theory are at the very least compatible with those theories discussed above. Previously, Carley and Newell [17] examined the implications of SCT on computational agents and their sociability. We base our work on their observations.

Moreover, in social psychology there has been extensive research on the cultural differences in micro level interactions among groups of people. Cultural differences have been revealed in a variety of human behaviors such as in different pedestrian dynamics, evacuation behavior and more. In the pedestrian domain there are several cultural attributes that have been examined across different countries such as the distance that pedestrian keep from one another, their walking speed etc. Cultural differences have been also found in evacuation behavior such as the way the people react to the event, the way people evacuate themselves etc. In this dissertation we provide several examples of these cultural phenomena, which have been described in social psychology literature.

Hall [30–33] examined the distances that pedestrian keep from one another across different cultures. He was one of the first researchers who defined the concept of proxemics or the personal space which is an invisible boundary that people maintain from each other in different contexts. According to Hall each person is surrounded by four invisible bubbles of space: Intimate, Personal, Social and Public. Personal distance refers to interactions among good friends or family members. Social distance refers to interactions among acquaintances and public distance is used for all other interactions such as public speaking. Changes in the bubbles depend, among other things on relationships to the closest person and also on cultural background.

Beaulieu [5] also examined cultural differences in personal space where she measured personal differences in 4 cultural groups. The research showed that Anglo Saxons used the largest zone of personal space, while Mediterraneans and Latinos used the shortest distance.

Levin and Norenzayan [50] examined the cultural differences in the pace of life from 31 countries. According to their definition the pace of life comprises three indicators: average walking speed, the postal speed and the accuracy of public clocks. They showed that Japan has the fastest pace of live. They also showed people in England and France have faster walking speed than people in Jordan or Syria.

Berkowitz [7] provides a naturalistic study of urban pedestrians in six national groupings by analyzing their national social behaviors. His goal was to contribute to quantitative cross-cultural data on various pedestrian social behaviors. He examined 20 different locations in 6 different countries such as Italy, England, Iran, Turkey and more. His study shows that in Moslem countries, England and West Germany there is higher incidence of people in groups than in Italy and the United States.

Chattaraj et. al., [19] examined whether there are cultural differences between Indians and Germans, in pedestrian streams in corridors. In an experiment they performed on pedestrians walking in straight lines, they found that the speed of Indian individuals is less dependent on density than the speed of German individuals. Moreover, they also found that German groups keep greater personal space than Indian groups.

Patterson et. al., [63] examined the cultural differences in micro-interactions of pedestrians in Japan and in the United States as they walked past a confederate. They concentrated on the effect of sex of the confederate and his or her behavior when passing one another on the sidewalk such as glances, smiles, nods, greetings and more. The results show that pedestrians in Japan are less responsive in terms of smiles, nods or verbalizing a greeting than pedestrians in the United States.

Cultural differences also have been examined in the evacuation domain. Andrée and Eriksson [2] examined cultural differences between Swedish and Australian in evacuation scenarios. They conducted experiments where they examined behavior and emotional patterns of 257 students from Sweden and Australia during a fire alarm and collected data by using questionnaires, video recordings and semi-structured interviews with the subjects. The results show that the Australians students are more serious about the alarm than the Swedish and they were also more scared.

Bryan [16] also examined cultural differences in the evacuation domain. He conducted an experiment which involved 584 participants over 335 fire incidents and collected data by interviewing the subjects. He also compared the results to the findings of previous studies. He compared different parameters such as participants' awareness to the fire incident, the participants' initial action during the incident etc. The study showed among other results that people of different cultures tend to notify others about the existence of the event, to different extents. For example, in the U.S there is a higher tendency to notify others about the event than in England

2.1.2 Computational models

Work on computer modeling of collective behavior has been carried out in other branches of science, in particular for modeling and simulation. Inspired by different fields of science, researchers have been developing computational models for simulation of collective behavior. However, only a few models have been validated against human data [20, 35, 46] Indeed, only limited quantitative data exists on the behavior of human crowds at a resolution which permits accurate modeling. Moreover, a key problem with these models is that the algorithms they provide change with the crowd phenomenon modeled.

Reynolds [66] simulated *bird flocking* using simple, individual-local rules, which interact to create coherent collective movement. There are only three rules: Avoid collision with neighbors, match velocity with neighbors, and stay close to the center of gravity of all neighbors. Each simulated bird is treated as a particle, attracted and repelled by others. On the one hand there is a desire to stay close to the flock, but on the other hand, there is a desire to avoid collisions. However, this model is limited only to the interactions of the agents, and does not allow individual goals (e.g., their own steering behavior).

Tu and Terzopoulos [79] simulated the motion of artificial fish with addressed individual goals. Like Reynolds' "boids", the artificial fish are autonomous creatures with simple behaviors, but combined they are able to create a more complex, collective behavior. However, unlike Reynolds' boids, where their behavior was selected based on the current state of their neighbors, each fish revealed habits and a mental state (for example hunger, fear etc.) that also have an impact on behavior selection. Indeed, Reynolds later expanded his work on collective movement in [67] but, this time he allowed steering behavior among the autonomous agents. In the revised model, each agent has a set of simple steering behaviors such as seek, flee, pursuit, evade, etc. The combination of these simpler behaviors creates a complex steering behavior.

Similar ideas have been applied in swarm robotics. Matarić [52] sees collective (complex) behaviors as a combination of basic behaviors. Each robot has spatial behaviors (controllers) that are combined to create different kinds of group behavior: for example, flocking consisting of *safe-wandering* (moving around without bumping), *homing*, *dispersion* (moving away from other agents), and *aggregation* (moving towards other agents). The combined outputs of the basic behaviors provide a velocity vector which is used to control the robot.

Yamashita and Umemura [88] took a different approach in simulating panic behavior. While inspired by Reynolds' boid model, they proposed a model where each simulated person moves according to three instincts: escape instinct, group instinct and imitational instinct. According to Yamashita and Umemura, when a person is in a state of panic, he or she acts based on their instincts which simplifies their decision making process.

Henderson compared pedestrian movement to gaskinetic fluids. Based on experiments on real human crowds, he showed in [41] that crowd distribution is compatible with Maxwell-Boltzmann's distribution. Henderson [42] developed a pedestrian movement model based on the Maxwell-Boltzmann theory. Since each person has mass and velocity, the crowd may be compared to liquid gas and under certain assumptions, the Maxwell-Boltzmann theory may be applied. Based on Boltzmann-like equations, Helbing [34, 35] developed a general behavior model for simulation of crowd dynamics. The proposed model takes into account social forces caused by interaction between the individuals and external or spontaneous forces which are caused by the physical environment.

Helbing et al. [35, 37, 38] observed self-organization in collective motion phenomena which can be caused by interaction among pedestrians. Self-organization, in this case means that there were some behavioral phenomena which were not planned, e.g., creation of lane formation in pedestrian movement. These lanes are created as a result of pedestrians moving against the flow. The number of lanes that are created cannot be planned. It depends on the width of the street and on pedestrian density.

Helbing and Vicsek [39] expanded their physical model by using game theory. The attraction force can be expanded to profitable force which may lead to optimal self-organization in pedestrian movement. Each entity calculates the "expected success" per each possible action and the action with maximum success is chosen. In pedestrian relations, actions are possible directions in which an entity can move and optimal self-organization is minimal interaction between entities.

Moussaïd et al. [57] examined the impact of a group's motion on pedestrian crowd dynamics. They showed that social interactions among group members create different group walking patterns. They also examined the impact of such patterns on the pedestrian flow. Their results show that in low density group members tend to walk side by side, however, as the density increases the group members form a V-like pattern formation which reduces the flow due to its non-aerodynamic shape.

Brown et al. [13] examined how individual characteristics impact crowd evacuation. They expanded Helbing's physical model by adding individual parameters to each agent, such as dependence level and altruism level. According to the model, groups combining altruism and dependent agents will be formed. By changing these attributes, they examined crowd evacuation by measuring the flow of people passing the door per second, and population distribution in the flow.

Blue and Adler [8] proposed a different approach to model collective dynamics. They used Cellular Automata (CA) in order to simulate collective behaviors, in particular pedestrian movement. The focus again was on local interactions whereby each simulated pedestrian is controlled by an automaton, which decides on its next action or behavior, based on its local neighborhoods. These rules are responsible for making a decision about lane changing and forward movement. If the path ahead is free, then it is taken. If not, then the automaton seeks to proceed left or right. If both lanes are available, one is chosen arbitrarily. Blue and Adler showed that this simple rule results in the formation of lanes in movement, similar to those formed in human pedestrian movement [87].

Toyama et al. [77] expanded the cellular automata model by adding different pedestrian characteristics, such as speed, gender, repulsion level, etc. The model was examined on bi-directional pedestrian movement behavior and on evacuation behavior. The main problem with this approach is that each collective behavior is simulated with a different CA model. For example, CA for simulation of pedestrian behavior has a different set of rules than the CA for evacuation behavior.

Osaragi [60] proposed a model for simulating pedestrian flow by using the concept of pedestrian mental stress. Pedestrian mental stress increases as a result of other pedestrians (density) and whether the pedestrian is unable to move to her destination using the shortest path. To decrease her mental stress, the pedestrian may dynamically change her direction or walking velocity. Because of these dynamic changes, the simulated pedestrians are heterogeneous. The model parameters were estimated using observed data.

Kretz [46] proposed the Floor field-and-Agent based Simulation Tool model (F.A.S.T) which is a discrete-space and discrete-time model for pedestrian motion. The F.A.S.T model can be classified as an extension of Probabilistic Cellular Automata (PCA). The F.A.S.T model has been validated against human data. In particular, the model simulation results of evacuation scenarios were compared to results of an evacuation exercise at a primary school.

Daamen and Hoogendoorn [20] performed empirical experiments on human crowds, in particular in terms of movement as pedestrians. In these experiments, participants were asked to walk through a monitored area, in both directions. Their movements were recorded. One conclusion was that "during capacity conditions, two trails or lanes are formed: pedestrians tend to walk diagonally behind each other, thereby reducing the head ways and thus maximizing the use of the infrastructure supply".

In all of these previous works above, the behavior of crowds in every domain of study (pedestrian movement, flocking, evacuation, etc.) is computed using a different algorithm, yet the actions and perceptions remain largely invariant (e.g., distances to others, occupied spaces versus empty spaces, goal locations, etc.). Instead, the computation itself changes between modeled behaviors.

For instance, many models of crowd behavior utilize cellular-automata (CA), which differ between domains. One CA model for pedestrian movement [8] uses a set of 6 IF-THEN rules which work in parallel for all cells, to simulate the movement of pedestrians in cells. The rules utilize knowledge of the occupancy in adjacent (rules 1,3 in [8]) and farther cells (rule 2), as well as knowledge of the distance to oncoming pedestrians in the same lane (rules 4 and 6). The rules set the forward velocity and position of the entities, by using a set of non-deterministic choices (sub-rules 5a,5b,5c), biased by distributions which differ depending on the environmental settings (e.g., choose from a uniform 50%/50% split distribution if two nearby cells are occupied, or from a 10%/80%/10% distribution when three cells are available). Another CA model for evacuation [76] uses knowledge of adjacent cells and distances to exits, and sets the position of the entities. Thus the actions and perceptions of each entity are similar to those used in the pedestrian model. But the algorithmic computation of the new position is done in two deterministic rules [76, pp. 17], which does not involve arbitrary choices at all.

In contrast to these previous investigations, we seek a *single cognitive mechanism*—a single algorithm—that, when executed by individuals, will give rise to different crowd behaviors, depending on the perceptions and actions available to the agents. This single algorithm will account for different crowd phenomena, by virtue of the actions and perceptions available to each individual.

In our previous work [24], we presented a model of crowd behavior, based on the social comparison theory (SCT) [17, 21], a popular social psychology theory that has been continuously evolving since the 1950s. The key idea in this theory is that humans, lacking objective means to evaluate their state, compare themselves to others who are similar. We believe that social comparison is a general cognitive process underlying the social behavior of each individual in a crowd. While the model was successfully evaluated on a variety of pedestrian behaviors [24], it was not evaluated against human data.

2.2 A macro-level approach to modeling crowd behavior

There are several macro (group) level techniques that enable modeling and reasoning in relation to different social phenomena. Some of these techniques require building a model, e.g., system dynamics, qualitative reasoning. Nonetheless there are others that do not require a model as an input to enable predictions, such as machine learning techniques, in particular a decision tree.

System dynamics [26] is a macro level approach in the sense that it models an entire system. It uses defined stocks, flows and feedback loops to model system behavior. The models are basically sets of differential equations that describe changes in the system. In our domain, such accurate and full definitions are not available.

As stated there are techniques that do not require a model as input. Instead, machine learning techniques can facilitate reasoning regarding social phenomena by inducing a model from examples of the target concepts. For instance, a decision tree learning algorithm [55] uses a set of examples and the target classes to which they belong (for instance, examples of properties of demonstrations and their associated violence level) as input and induces (builds) a model, in the form of a decision tree. This decision tree allows classification of the observed data according to the given properties. However, as we show in this paper, prediction (classification) accuracy is not the only requirement for policy decision-support. In particular, we show that QR models can be much better at analyzing hypothetical settings.

Qualitative reasoning (QR) [22, 47] is a macro-level approach to modeling and reasoning with partial and imprecise numeric information. The traditional usage of QR is in modeling common-sense reasoning, e.g., in physics [22]. Indeed, for several decades extensive studies have been conducted on QR techniques in physics to enable reasoning about physical systems. However, while QR is usually associated with physics, other branches of science such as ecology [69], social science [45], politics [23] etc., are also beginning to adopt this approach. Even though social science is much less formal than physics, QR approaches have been found to be just as powerful.

For example, Kamps and Peli [45] present the use of QR in the application of social science. They showed that the QR approach was successfully applicable for modeling and reasoning about the density dependence theory of organizational ecology. They simulated the growth pattern of a population and enabled prediction regarding the population size. Salles and Bredeweg [69] used QR techniques in ecological modeling. Based on theoretical and commonsense knowledge, they built a qualitative model of population and community dynamics in the Brazilian Cerrado vegetation. They showed that QR models successfully capture ecological knowledge and enable valid prediction of different behaviors in ecological systems.

Brajnic and Lines [11] applied the QR technique in complex, socio-economic allocation problems, in particular the allocation of national income between several important resources such as capital investment, social services etc., They show that with relatively weak quantitative information about functional relationships useful prediction regarding the behavior of an economic society can be drawn.

In this work we used a qualitative reasoning approach for modeling and reasoning about the demonstration behavior of the crowd. In particular, our goal was to model and predict the violence level during demonstrations. There are several theories in social science regarding the influencing factors on the violence level. These factors are described partially and qualitatively, at a macro-level without full and precise information. Moreover, we are not aware of almost any simple model that attempts to be comprehensive. Thus, we use a qualitative reasoning approach for modeling and reasoning regarding the potential violence level of demonstrations. To the best of our knowledge, QR techniques have never been used to model crowd behavior phenomena.

Part I

Modeling Crowd Behavior using an Agent-Based Approach

Chapter 3

Model Evaluation

In this chapter we present the existing model of Social Comparison (SCT). However, while the existing SCT model has demonstrated good results in modeling pedestrian behavior [24], it has never been evaluated against human data. In this chapter we provide a qualitative evaluation of the SCT model against human pedestrian behavior. We also evaluate our model in evacuation behavior. Moreover, we discuss why we believe that the SCT model is compatible for modeling crowd behaviors.

3.1 An Existing Model of Social Comparison

In recent years we have been successfully developing the SCT model of crowd behavior, inspired by the social psychological theory of social comparison. Festinger's Social Comparison Theory (SCT) [21] served as an inspiration for the social skills necessary for our agent to be able to exhibit crowd behavior. According to the social comparison theory, when lacking objective means of appraisal of their opinions and capabilities, people compare their opinions and capabilities to those of others who are similar to them. They then attempt to correct any differences found.

Festinger presents the social comparison theory as an explicit set of axioms. The following subset of axioms (re-worded) is particularly relevant (see also [17, 21] for additional discussion):

- When lacking objective means of evaluation, the agents compare their state to that of others;
- Comparison increases with similarity;
- Agents take steps to reduce differences to the objects of comparison.

According to our existing SCT model, each observed agent A_i is considered to be a tuple of k state features $A \equiv \langle f_1^A, \ldots, f_k^A \rangle$. Each feature f_j^i of agent A_i ($1 \leq j \leq k$) corresponds to a dimension, such that agent A_i is represented by a point in a k-dimensional space, where the various dimensions correspond to state features (such as the location in x, y coordinates, color, heading, etc.)

For each such agent, we calculate a similarity value $Sim(A_{me}, A_o)$, which measures the similarity between the observed agent A_o and the agent carrying out the comparison process A_{me}. The agent with the highest value is selected.

If its similarity is between the given maximum and minimum values, the comparing agent is triggered to perform actions to reduce the discrepancy.

Algorithm 1 Argmax SCT $(O, A_{me}, S_{min}, S_{max})$

1: $S \leftarrow \emptyset$
2: **for all** $A_o \in O$ **do**
3:     **if** $S_{min} < Sim(A_{me}, A_o) < S_{max}$ **then**
4:         $S \rightarrow S \cup A_o$
5: $A_c \leftarrow \texttt{argmax}_{A_c \in S}(Sim(A_{me}, A_o)$
6: $D \leftarrow$ differences between me and agent A_c
7: $a \leftarrow$ SelectAction(D)
8: Apply action a with its Gain (Eq. 3.2) to minimize the differences in D.

This process is described in the Algorithm 1. Each agent A_i executes the following algorithm (Algorithm 1). In line 2 and 3, for each observed agent $A_o \in O$, we calculate a similarity value $Sim(A_{me}, A_o)$, which measures the similarity between the observed agent A_o and the agent carrying out the comparison process (A_{me}) (Eq. 3.1). We model each agent as an ordered set of features, where similarity can be calculated for each feature independently of the others. We measure similarity between agents independently along each dimension. The similarities in different dimensions are functions $s_{f_i}(f_i^{A_{me}}, f_i^{A_o}) : f_i \times f_i \mapsto [0, 1]$. The function s_{f_i} defines the similarity in feature f_i between the two agents A_{me} and A_o. A value of 0 indicates complete dissimilarity. A value of 1 indicates complete similarity. For instance, one commonly used feature denotes the normalized Euclidean distance, inverted: A value of 0 means that the agents are as far apart as possible. A value of 1 means that they are positioned in the same location.

To determine the overall similarity between two agents, we use a weighted sum over the functions s_{f_i}. With each feature f_i, we associate a weight $w_i \geq 0$. The similarity between two agents is then given by Eq. 3.1 below.

$$Sim(A_{me}, A_o) \equiv \sum_{j=1}^{k} s_{f_j}(f_j^{A_{me}}, f_j^{A_i}) \cdot w_j \tag{3.1}$$

For each calculated similarity value, in line 3 we check if it is bounded by S_{min} and S_{max}, and in line 5 we select the agent A_c that maximizes the similarity, but still falls within the bounds. S_{min} denotes values that are not sufficiently similar, and the associated agents are ignored. Festinger states that "when a discrepancy exists with respect to opinions or abilities there will be tendencies to cease comparing oneself with those in the group who are very different from oneself" [21]. Respectively, there is also an upper bound on similarity S_{max}, which prevents the agent from trying to minimize differences where they are not meaningful or helpful. For instance, without this upper bound, an agent that is stuck in a location may compare itself to others, and prefer those that are similarly stuck in place.

In line 6, we determine the list of features (f_i, w_i) which cause the differences between A_{me} and the selected agent A_c (list of features with $f_i < 1$). We order these features in an increasing order of weight w_i, such that the first feature to trigger corrective action is the one with the lowest weight. Thus, the correction order increases from the lowest weight to the highest weight. The reason for this ordering is intuitive, and indeed we did not find evidence for it (or against it) in the literature.

Finally, in step 7 of the algorithm, the comparing agent A_{me} takes one corrective action (a) on the selected feature. Note that we assume that every feature has one associated corrective action that minimizes its gaps with the target agent, independently of other features. Festinger asserts that "the stronger the attraction to the group the stronger will be the pressure toward uniformity concerning abilities and opinions within that group" [21]. To model this, we use a gain function *Gain* (Eq. 3.2), which translates into the amount of effort or power invested in the action. For instance, for movement, the gain function would translate into velocity; the greater the gain, the greater the velocity.

$$Gain(Sim(A_{me}, A_c)) \equiv \frac{S_{max} - S_{min}}{S_{max} - Sim(A_{me}, A_c)} \tag{3.2}$$

3.2 Can Festinger's social comparison theory be used for modeling crowd behaviors?

In order to use a computational model of social comparison theory for modeling crowd behaviors we need to address the following issues. First, we need to examine whether there is a connection between the social comparison theory and crowd behavior. Second, we need to check whether the social comparison theory can be applied to superficial comparisons, i.e., at the level of visible differences between agents, in addition to cognitive differences (e.g., intentions). We address these two issues below.

3.2.1 Social comparison in crowds

To the best of our knowledge, the social comparison theory has never been associated with crowd behavior phenomena. However, we believe that the social comparison theory may account for some important characteristics of crowd behavior, as it clearly addresses processes in groups, and no limit is placed on a group's size.

We focus on one of the primary characteristics of crowds, i.e. the similarity between individuals' behaviors. This similarity is explained by a process of *imitation* [49], convergence of like-minded individuals [1], or emerging norms [80].

Social comparison processes can give rise to this phenomenon. Festinger states that "the existence of a discrepancy in a group with respect to opinions or abilities will lead to action on the part of members of that group to reduce the discrepancy" [21, p. 124]. Indeed, one implication of SCT is the formation of homogeneous groups. Festinger notes that "the drive for self evaluation is a force acting on persons to belong to groups, to associate with others. People, then, tend to move into groups which, in their judgment, hold opinions which agree with their own" [21, p. 135]. This quote, in particular, seems to be compatible with [1].

3.2.2 Do people engage in surface comparisons?

Festinger hypothesizes that "there exists, in the human organism, a drive to evaluate his opinions and his abilities" [21, p. 117]. Thus a question which emerges with respect to the mechanisms we describe in this work is whether in fact this type of surface comparisons is associated with the social comparison theory.

There has been extensive research clarifying the concepts of "abilities" and "opinions". Smith and Arnkelsson [72] explain that evaluation ability refers to a person's performance on a specific task. Festinger himself provides a link between ability and performance as he states "abilities are of course manifested only through performance which is assumed to depend upon the particular ability" [21, p. 118]. He then provides an example "Thus, if a person evaluates his running ability, he will do so by comparing his time to run some distance with the times that other persons have taken." [21, p. 118].

Moreover, the meaning of opinion comparison, also has been extensively investigated throughout the years since the publication of [21]. Goethals and Darley [27] associate this concept with "the Related Attributes Hypothesis" which states that people will prefer to compare themselves with others similar to them on attributes that are related to their opinion or performance. Festinger provides the basis for this research by claiming that "if persons who are divergent from one's own opinion or ability are perceived as different from oneself on attributes consistent with the divergent, the tendency to narrow the range of comparability becomes stronger" [21, p. 133]. Goethals and Klein provide an example which directly acknowledges surface comparisons: "An individual evaluating his or her tennis-playing ability. He or she might compare with others who are about the same age, who have the same degree of recent practice and comparable equipment, and who are the same sex" [28, p. 25].

There is much evidence that people perform surface comparison in their everyday tasks even when they are walking down the street. For example, people use SCT to decide whether to return a lost wallet [43]. Here is another example: A well-known experiment in social sciences was performed by Milgram et al. [54]. The experiment involved one participant who stood in the middle of a busy street and stared into an empty spot in the sky. The purpose of the experiments was to examine group pressure. The results showed that when there was only one participant, only a few people that passed briefly glanced up. However, when there were several participants, almost 80 percent of those passing by also stopped and stared into the sky. Thus it seems that the application of the social comparison theory to explaining crowd behavior is justified.

3.3 Pedestrian Behavior: Validation Against Human Data

In this section we evaluate the SCT model (presented in Section 3.1) on general pedestrian movement which includes individuals, couples, and groups, all walking at different speeds in a bidirectional fashion. We compare the performance of the model with popular models from the literature, and explore the impact of different parameters and model components (e.g., bounds) on the generated behavior. The evaluation was carried out by 39 human subjects who compared the behavior generated from the different models to movies of real-world pedestrians. The results clearly demonstrate that the SCT model is superior to others in its fidelity to human pedestrian behavior.

The SCT model was previously evaluated separately on different crowd behaviors [24]. In particular, it was assessed in reference to different types of pedestrian movement phenomena such as the formation of lanes in bidirectional movement of individuals, movement in small groups with and without obstacles, etc. When evaluated on a specific behavior, community-recognized standard measures, such as flow, number of lane changes, etc. can be applied. However, when evaluating the model against human data, it must account for a more complete set of behaviors, all mixed together. For example, when watching pedestrians, we can observe people moving as groups like a family, friends and couples or as individuals, all walking at different speeds in a bidirectional fashion.

A different evaluation methodology is thus needed. One of the greatest challenges in modeling crowd behaviors is the great absence of human crowd behavior data that can be used as a basis for comparison. The main difficulty in the creation of such data is that controlled experiments are complex to design, and costly to execute, since they have to be in large scales. A standard methodology of evaluation does not exist; some researchers generate accurate behavioral data by engaging crowds in virtual environments [64], while others do qualitative comparisons of their models' predictions in comparison with movies of crowds, i.e., via observation experiments, e.g., [36, 46]. We followed the same approach. Below, we describe the observation experiments we executed to evaluate the SCT model on general pedestrian behavior.

3.3.1 Comparison of the SCT Model Generating Pedestrians' Behavior and Human Pedestrian Behavior

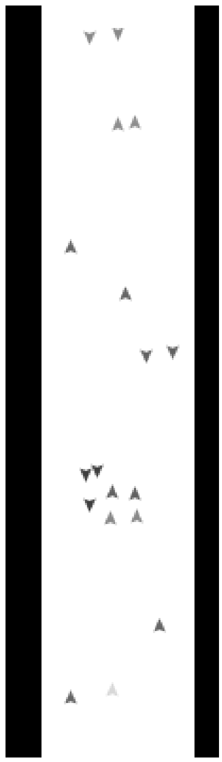
We used human crowd movies where different pedestrian behavior phenomena are presented (Figure 3.1) and created screen-capture movies of different models of the same behavior (Figure 3.2). We relied on experiments with human subjects which compare each of the resulting simulated behaviors to human crowd behavior. In addition, the subjects also voted for the most similar and dissimilar simulated behavior.

(a) (b)

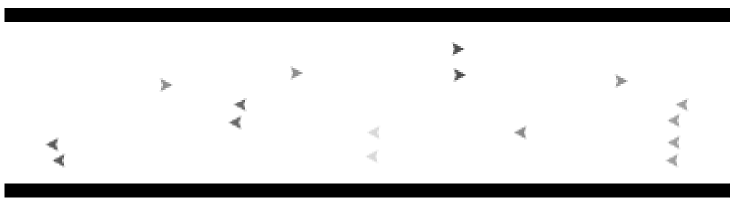
Figure 3.1: Human pedestrian behavior.

Simulated Behavior: Experiment Setup.

To simulate pedestrian behavior, we used Net-Logo. We defined a sidewalk with 104 patches in length and 10 patches in width. In order to match human crowd density the sample population comprised 30 agents. Agents were able to move in a circular fashion from east to west or in opposite directions at different speeds. Agents that belong to the same group had the same color. In order to create small groups, couples and individuals, we used 15 different colors to define our population (this was a large number considering the population size). Agents were placed in random positions at the beginning of the experiment, and each agent had a limited visual distance of 10 patches and a cone-shaped-field-of-view of 120 degrees.



(a)



(b)

Figure 3.2: **Simulated pedestrian behavior.**

Each agent had a set of features and their corresponding weights. In order to simulate pedestrian movement, we used the following features and weights: *color* (weight 3); *walking direction* east or west (weight 2); and *position* (weight 1), given global coordinates. To account for the Western cultural instinct that friends (and family) walk side-by-side, rather than in columns, we used another feature: The similarity in position along the x-axis - *X-Coordinate* (weight 0.5).

The rationale for feature priorities, as represented in their weights, stems from our intuition and common experience regarding how pedestrians act. Positional difference (distance, side-by-side) is the easiest difference to correct, and the least indicative of a similarity between pedestrians. Direction is more indicative of a similarity between agents, and color (which we use to denote sub-groups within the crowds) even more so. For instance, if an agent sees two agents, one proceeding in the same direction as it (and far away), and the other very close to it (but in the opposite direction), it will calculate greater similarity to the first agent, and try to minimize the distance to it (this may cause a lane change) and only then it will try to locate itself on the same X-coordinate.

The similarities in different features (f_i) are calculated as follows. $f_{color} = 1$ if the color is the same, otherwise 0. $f_{direction} = 1$ if the direction is the same, otherwise 0, $f_{distance} = \frac{1}{dist}$, where *dist* is the Euclidean distance between the positions of the agents and finally, $f_{x-coordinate} = 1$ if the x-coordinate is the same, otherwise it equals 0. Each agent calculates the similarity value $Sim(A_{me}, A_o)$ according to the model. If the chosen feature for closing the gap is distance, then the velocity for movement will be multiplied by the calculated gain *Gain*. For other features (which are binary), the gain is ignored.

We examined the impact of the SCT model components on the quality of the simulated pedestrian behavior. In particular, we examined the impact of SCT bounds (S_{min} and S_{max}), gain function, and correction order on the generated behavior. We defined seven models, each emphasizing a different SCT component. The models are explained below, and summarized in Table 3.1.

First we examined the impact of SCT bounds on the generated pedestrian behavior. We hypothesized that narrower bounds would provide more similar behavior to the individual model. To examine this hypothesis, we defined the following models:

- SCT-B-2-6.5 We set S_{max} to 6.5 (namely, practically no agents are too similar) and S_{min} to 2 (which means that agents that differ only in distance and in the X-axis are not considered similar). The gain was calculated according to Eq. 3.2 and the correction order was from the low weight features (distance) to the high weight features. In this domain agents could not change their color, thus, the last corrected feature was direction. Our hypothesis was that this model would provide the behavior most similar to human pedestrians.

- SCT-B-5-6.5 We set the S_{min} to 5 which means that agents that are similar at least in color and direction are considered to be similar. Thus, in this model only agents with the same color and direction would move together.

Another component that we examined was the impact of correction order on simulated pedestrian behavior. In the SCT-H-L model we defined the correction order to be from high to low. Our agents could not change their colors, and in this model if the selected agent moved in the opposite direction, the agent would first change its direction and then try to close the distance gap.

Finally, we evaluated the importance of the gain in the model. We defined the following models:

- SCT-NoGain Was defined without the gain function (i.e., the gain was constant 1).

- SCT-G-C2 The gain function was constant (2).

- SCT-G-C3 The gain function was constant (3).

- SCT-G-C4.5 The gain function was constant (4.5).

The various SCT models were compared to the individual choice model, commonly used in crowd research [8, 36]. In the individual model, each agent makes its decisions independently of its peers and in the pedestrian domain, when forward movement is blocked, an agent will arbitrarily choose a different lane. This model has been shown to be qualitatively compatible with pedestrian motion, and is often used as a baseline technique in crowd research (see, for instance, [46]).

Model	Smax	Smin	Gain	Correction order
SCT-B-2-6.5	6.5	2	Eq. 3.2 (func.)	L-H
SCT-B-5-6.5	6.5	5	Eq. 3.2 (func.)	L-H
SCT-B-H-L	6.5	2	Eq. 3.2 (func.)	H-L
SCT-B-NoGain	6.5	2	1 (const)	L-H
SCT-G-C2	6.5	2	2 (const)	L-H
SCT-G-C3	6.5	2	3 (const)	L-H
SCT-G-C4.5	6.5	2	4.5 (const)	L-H

Table 3.1: **SCT Models**

Comparison to Human Crowds.

In order to compare the simulated behavior to general behavior and not to specific video clips, we used several video clips on human pedestrian behavior and several screen-captured movies for each model. In the simulated behavior we created three screen-captured movies for each model that was randomly chosen for each subject. In human behavior we used two sets of video clips that were taken from different locations and at different times. The first set of movie clips were taken in the morning in downtown Vancouver, during rush hour. People mainly walked individually, and only a few moved in small groups. The second set of movie clips were taken in the afternoon on a street that leads to the Eiffel tower in Paris, during leisure time. Most of the pedestrians were families and friends that moved in small groups, or as couples. Each real-world video clip was cut to one minute clips. To generate a one-minute clip in the simulated behaviors, each model was executed for 5000 cycles (6 minutes), and the last minute was used.

We built a web-based experiment which enabled the subjects to participate during their free time. First we presented a brief description about the experiments. The subjects were told that the purpose of the experiment was to compare each of the simulated behaviors to human crowd behavior. However, the purpose of the simulation was not to simulate each pedestrian observed in the human crowd, but to simulate the general pedestrian behavior. The experiment was carried out in two phases, a training phase that was presented to the subjects after the description of the experiment, and an experiment phase.

The experiment was carried out with 39 subjects (ages: 20–84, mean: 29, number of males: 28). An additional 6 subjects were dropped due to technical reasons (such as network problems that prevented them from viewing the clips). The subjects were asked to watch the human pedestrian movie which was randomly chosen in each experiment. Then, they were asked to view a screen-captured movie of each model that was also chosen randomly. After each simulated movie, the subjects were asked to rank the observed behavior, which was followed by a question: To what degree was the observed simulated behavior similar to previously observed human behavior? (1 – not similar, 6 – similar). At the end of the experiment, we asked the subjects two additional questions: Which simulated movie was the most similar to human behavior and which simulated movie was the most dissimilar? To control the effects of order, the order of the presentation on the page was randomized.

Initially we had planned to compare eight different simulated behaviors to human pedestrian behavior, the individual choice model and seven SCT models. We ran a short pilot where we presented the experiment to three subjects and then we asked their opinion. All the subjects claimed that the experiment was too long. Moreover, they claimed that the SCT-B-2-6.5 model provided very similar behavior to the behaviors in the SCT-H-L model and similar behavior was also observed in the SCT-NoGain, SCT-G-C2, SCT-G-C3 and SCT-G-C4.5 models. Consequently, we reduced the number of different models that were presented to the subjects. In the experiment phase we compared four simulated behaviors. We used the Individual-choice model, SCT-B-2-6.5, SCT-B-5-6.5 and one of the randomly chosen models, SCT-NoGain, SCT-G-C3 and SCT-G-C4.5. The SCT-H-L and SCT-G-C2 models were used only in the training phase, thus their results were not included in this work.

3.3.2 Results

We first examined the ranking of the models in comparison to the actual crowd. The results are summarized in Figure 3.3. The categories on the X-axis correspond to different models. The Y-axis corresponds to grades of the compared models. Each set of bar shows the mean and median results. A higher result indicates improved fidelity, i.e., greater similarity to human pedestrian behavior.

The results clearly demonstrate that the SCT-B-2-6.5 model provides higher results than the compared models. While it may seem that the SCT-B-2-6.5 model results are close to the results of the Individual and the SCT-

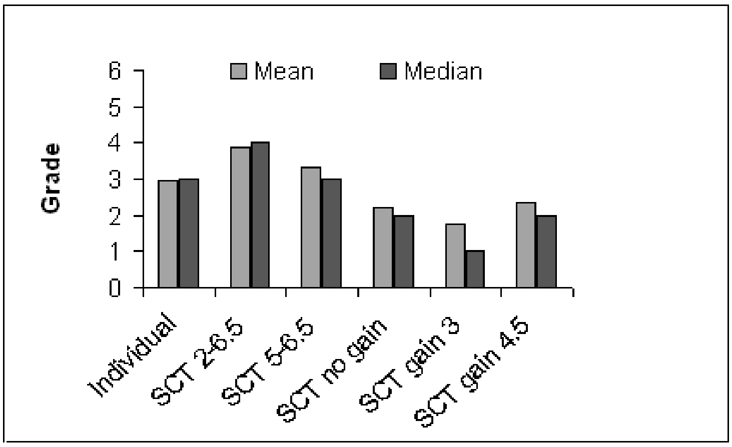


Figure 3.3: **Results of the comparison to human pedestrians**

B-5-6.5 models, according to a t-test (two-tailed) the SCT-B-2-6.5 was found to perform significantly different than both the Individual model ($p = 0.001$) and the SCT-B-5-6.5 model($p = 0.03$).

Another hypothesis underlying the experiment was that the SCT model with narrower bounds (S_{min}, S_{max}) would provide behavior closer to individual model behavior, but not the same. Indeed, the results demonstrate that SCT-B-5-6.5 lies between the SCT-B-2-6.5 and individual models. According to the t-test (two-tailed) SCT-B-5-6.5 was found to be significantly different from SCT-B-2-6.5 ($p = 0.03$) and significantly different from the Individual model ($p = 0.017$).

Our last hypothesis was that the SCT models without the gain function would provide behavior less similar to human pedestrian behavior. The results clearly demonstrate that the SCT-NoGain, SCT-G-C3 and SCT-G-C4.5 models in which the gain is fixed, attain the lowest results.

When we asked the subjects: "which simulated behavior was most similar to human behavior?" The SCT-B-2-6.5 model received the highest number of votes. To the question of "which simulated behavior was the most dissimilar to human behavior?", the subjects responded that the SCT-NoGain, SCT-G-C3 and SCT-G-C4.5 models were the least similar. The responses to these two questions are depicted in Figure 3.4.

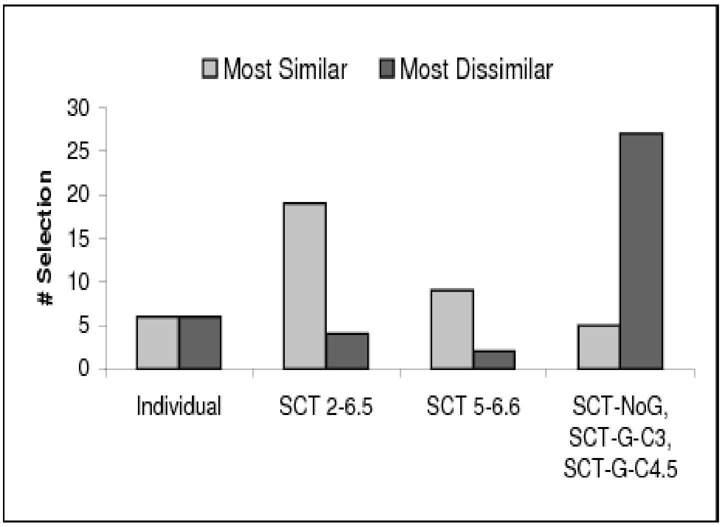


Figure 3.4: **Most similar/dissimilar results.**

3.4 Evacuation Behavior

In this section we evaluate the SCT model in reference to evacuation behavior. To model evacuation behavior, we used ESCAPES which is a multiagent evacuation simulation [78] that incorporates four key features: (i) different agent types; (ii) emotional interactions; (iii) informational interactions; (iv) behavioral interactions among agents. ESCAPES was used to model evacuation behavior in the International Terminal at the Los Angeles International Airport (LAX). It has been shown that ESCAPES provides good results for modeling evacuation behavior. Moreover, it received much praise from LAX security officers. We show that the use of SCT in evacuation leads to increased grouping of agents.

ESCAPES is a simulation of an airport with 4 terminals and 4 available exits where individual agents and also families wander freely around in shops or in available areas before the event. After the event, agents evacuate themselves. Moreover, in this simulation authority figures are also present who have the task of patrolling before the event and informing other agents about the event and about the available exits after the event. Each agent has a subset of 14 available behaviors from which it selects one, using a common architecture based on a BDI framework, according to its knowledge about the world and about other agents. For a more realistic simulation, agents have incomplete knowledge about their environment, in particular about the available exits and also about the event. Each agent holds an event certainty value (an integer between 0 and 2), which indicates the agent's awareness regarding the event and when the event certainty is high an agent will decide to evacuate. Each agent also has a specific level of emotions which affects its behavior during the evacuation, in particular its speed. Speed is modeled as an integer value between 0 and 3, and fear is modeled as an integer value between 0 and 2 (FearFactor) where 0 indicates that the agent has no fear. Higher levels of fear lead to higher movement speeds. Moreover, an agent's fear is affected by several factors such as its proximity to the event (which increases the agent's event certainty and also its fear), presence of authority figures (which decreases the agent's fear) and more. Agents that decide to evacuate also spread the knowledge about the event to their neighbors.

We found the utilization of the computational model of social comparison (SCT) to be helpful in developing agents with social skills crucial to accurate simulation of different crowd behaviors, in particular in evacuations. Social comparison is considered a general cognitive process, which underlies human social behavior. During emergencies, individuals face greater uncertainty, and thus the weight of social comparison in human decision-making is increased [48]. The SCT computational model can be used, for instance, by agents who wish to exit an area, urgently. If they do not know the location of a close exit, they may turn to mimicking others hoping that they will lead them to safety.

For the simulation, SCT was implemented as follows: first, the agent compared itself to others around it by measuring the similarity in a set of features, including speed, emotion state, distance, etc. The similarity values were totaled, and the agent most similar (within bounds) was selected. The agent executing SCT took actions to reduce dissimilarities to the selected maximally-similar agent.

The use of SCT in evacuation led to increased grouping of the agents, as we show in the experiment results. This grouping (herding) has been reported by [2, 36] as occurring in real-world emergencies. We thus believe that SCT, as a cognitive model, can account for herding in human crowds.

3.4.1 Evaluation of the SCT model in an evacuation scenario

First we examined the population without authority figures. To examine the impact of agents' grouping behavior on the evacuation we compared agents using the SCT process and agents not using the SCT process and we measured the mean value of the agents' connectivity. Connectivity was defined as the number of connectivity components in the adjacency matrix of the agents and we took the mean value of all the unevacuated agents. Thus a higher connectivity value means increased grouping. Graph 3.5 presents the impact of SCT on the agents' connectivity. The results show that agents' using the SCT process had much higher connectivity which indicates more grouping behavior than agents not using the SCT process. Moreover, the connectivity of agents using SCT was found to be significantly higher than the connectivity of agents not using SCT, according to the one tailed t-test, $alpha = 0.01$.

We then examined the impact of the SCT process on the population with authority figures. We defined 5 authority figures and as in previous experiments we measured the agents' connectivity. Graph 3.6 presents the impact of the SCT process on the agents' connectivity. The x-axis corresponds to

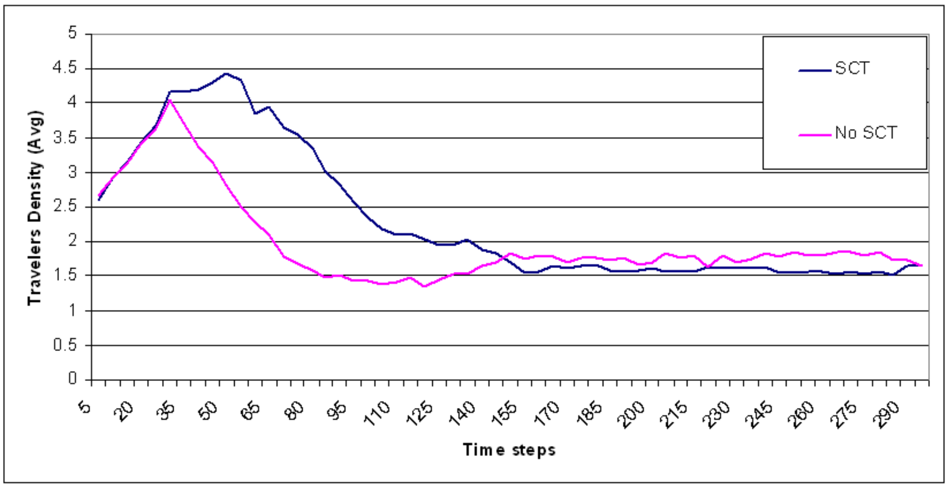


Figure 3.5: **The effect of SCT on density (without authorities)**

the time steps and the y-axis corresponds to the agents' connectivity. In contrast to the population without authority figures, in this case there was no significant difference in connectivity between the agents using the SCT process and those not, according to the two-tailed t-test, $alpha = 0.49$.

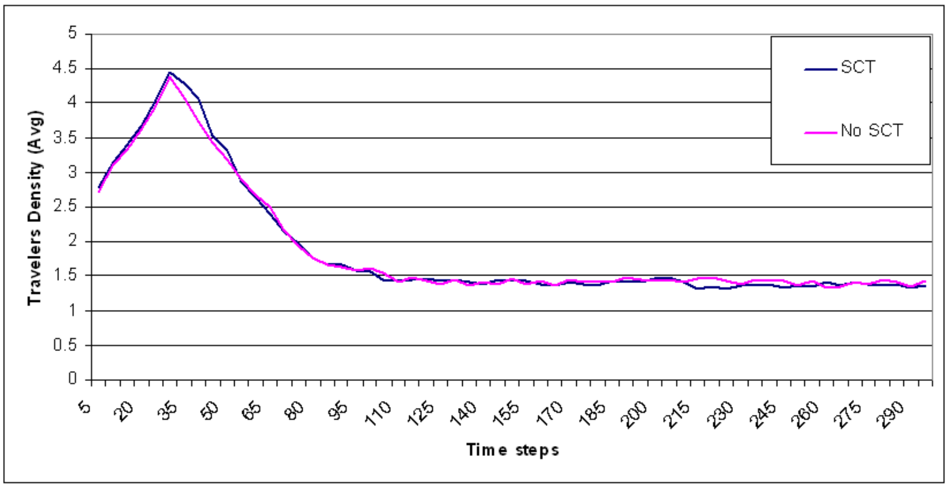


Figure 3.6: **The effect of SCT on density (with 5 authorities)**

Chapter 4

Social Comparison at the Cognitive Architecture Level

The SCT model relies on simulated entities (agents) to compare themselves to others, but the timing of these comparisons is not well understood: People clearly do not imitate others all the time. Nonetheless, there is evidence that shows that people (and therefore agents), do some comparison at all times (but do not act on them). While some progress has been made to address this question, it still remains open.

In this chapter, we address this open question. We argue that comparisons take place all the time (i.e., differences are perceived and processed), but the cognitive architecture limits actions taken to minimize differences to cases where the comparisons yield significant differences. We use both pedestrian domain experiments as well as movies of human pedestrians to argue our viewpoint.

4.1 When are social comparison processes triggered?

In this section, we address the question of *when* social comparison is triggered. We examine possible answers to this question in Section 4.1.1, and in Section 4.1.2 we present findings of experiments we conducted, which rule out some of the possible solutions and support others.

4.1.1 Social comparison at the cognitive architecture level

There are two possible implementations of the SCT process at the architectural level. The first, which seems to follow directly from Festinger's Social Comparison theory, treats the SCT process as an uncertainty-resolution method, i.e., as a weak (read: general) problem-solving method, which is *social*. The second, takes a different approach, whereby an SCT process is constantly active, in parallel to any problem solving activity which necessitates the agents to be constantly aware of others around them.

According to Festinger, people use social comparison when they lack the knowledge required to make their decisions. Thus one way of implementing the SCT process in a cognitive architecture is as a response to an uncertainty: When an agent is in an uncertain state, it may call on a comparison process to assess similarity and proposed actions.

As a result we may treat the social comparison theory as a new kind of uncertainty-resolution method. Unlike previous uncertainty-resolution (problem-solving) techniques, whereby the agent focuses on using its own resources, in this case the agent uses knowledge of others as a basis for resolving the uncertainty.

Readers familiar with the Soar integrated cognitive architecture will undoubtedly be reminded of the capabilities of Soar to detect *impasses*, situations in which the agent has no direct knowledge of how to proceed with its task, and relies on problem-solving methods to resolve the impasse [59]. In viewing SCT as a problem-solving activity, in Soar it is modeled as an impasse-resolution method.

However, elaborations on social comparison theory have expanded the view of when comparison takes place. Hakmiller [29] and Singer [71] expanded the theory and demonstrated that people tend to confirm or reassure that their actions or beliefs are the correct ones, by comparing themselves to others. Thus according to this approach people tend to use social comparison in parallel to their decision making process.

We thus offer an account in which a second hypothesis (in which a comparison process is always active) can be made compatible with Festinger's observations (that comparison occurs with uncertainty). Our hypothesis is that social comparison should always be active *alongside* any goal-oriented action-selection processes. The state of low uncertainty corresponds to the goal-oriented processes being able to produce coherent actions, which are then selected by the agents for execution. But when uncertainty increases (the goal-oriented processes do not suggest actions for execution), the social comparison processes manage to "advance" their own proposed actions for execution.

In other words, an alternative is to view the SCT as an on-going process, which takes place (at the architectural level) *in parallel* with any problem-solving activity. Whereas normally, actions are proposed (and selected) by cognitive architecture based on their suitability for the current goal (e.g., through means-ends analysis), in a socially-comparative architecture of this type, the agents' actions are also proposed based on the results of social comparisons. In other words, the agent would consider actions that advance it towards its goal, *as well as actions that seek to minimize perceived differences to other agents.*

It may seem easy to dismiss the implementation question as insignificant. However, the implementation choice carries significant impact: Since

SCT processes inherently rely on knowing about the behavior of others, the implementation question raises a more fundamental question about where modeling of others (e.g., using plan recognition) occurs in cognition: Is it a problem-solving activity, or is it carried out all the time, at an architectural level?

4.1.2 Experiments

We conducted a set of experiments to evaluate which of these two approaches (*Comparison as Problem-Solving* or *Continuous Comparison*) is more applicable in the context of pedestrian behavior simulations. We recreated the experiment setup and simulation environment used in [24, 25], and rewrote agents to operate in this environment.

We examined the two approaches in the context of pedestrian grouping behavior and bidirectional movement of individual pedestrians, described in Section 4.1.2. In these experiments we compared two types of SCT agents: the SCT-Problem-Solving agent and the SCT-Continuous agent. These agents had the same feature set and performed the SCT process in the same way as described previously. However, the main difference between them was *when* the SCT process was activated. While with a SCT-Problem-Solving agent the SCT process was activated when an agent was in an uncertain state, with a SCT-Continuous agent the SCT process was continuously active.

Simulation environment and setup

To simulate pedestrian behavior, we used the Net-Logo [86]. We defined a sidewalk comprising 104 units in length, where agents were able to move in a circular fashion from east to west (reappearing on the east side when they reached the boundaries of the west site) or in the opposite direction. Each agent had a limited visual distance of 10 patches and a cone-shaped-field-of-view of 120 degrees.

Each agent had a set of features and their corresponding weights. For simulating pedestrian movement, we used the following features and weights: *color* (weight 3); *walking direction* east or west (weight 2); and *position* (weight 1), given global coordinates. In grouping pedestrian simulation, to account for the Western cultural norm that friends (and family) walk side-by-side, rather than one after the other in rows, we used another feature i.e., similarity in the position along the x-axis - *X-Coordinate* (weight 0.5).

The similarities of the different features ($s(f_i)$) were calculated as follows:

- $s(f_{color}) = 1$ if the color was the same, otherwise 0.
- $s(f_{direction}) = 1$ if the direction was the same, otherwise 0.
- $s(f_{distance}) = max(\frac{1}{dist}, 1)$, where $dist$ was the Euclidean distance between the positions of the agents.
- $s(f_{x-coordinate}) = 1$ if the x-coordinate was the same, otherwise 0.

Each agent calculated the similarity value $Sim(A_{me}, A_o)$. If the chosen feature for closing the gap was distance, then the velocity for movement was multiplied by the calculated gain, denoted $Gain$. For other features (which were binary), the gain was ignored (as it had no effect on categorical values).

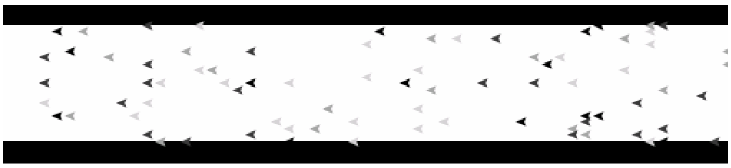
The rationale for the feature priorities, as represented in their weights, stemmed from our intuition and common experience regarding how pedestrians act. The positional difference (distance, side-by-side) is the easiest difference to correct, and the least indicative of a similarity between pedestrians. Direction is more indicative of a similarity between agents, and color (which we use to denote sub-groups within the crowds) even more so. For instance, if an agent sees two agents, one proceeding in the same direction (and far away), and the other very close to it (but in the opposite direction), it will calculate a greater similarity to the first agent, and try to minimize the distance to it (this may cause a lane change) and only then the agent will try to position itself on the same X-coordinate.

Two Crowd Modeling Tasks

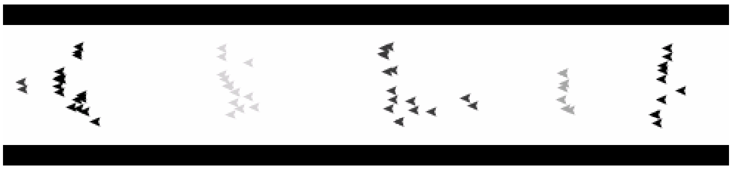
We examined two pedestrian crowd tasks. In the first, all the simulated pedestrians moved in the same direction (uni-directional traffic), and were divided into five groups, based on their color. Each agent was placed randomly, so that initially they were dispersed. Successful execution of the task involved moving while creating clusters of groups of the same color. In the second task, the simulated pedestrians moved in opposite directions (randomly assigned to agents). Each agent was independent of the others—no grouping was expected or desired.

To illustrate, Figure 4.1 depicts screen shots of the simulation running this task. The screens show the initial positions of the agents in one of the trials 4.1(a), their positions after moving 5000 cycles using the continuous

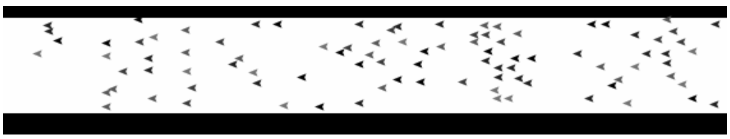
SCT approach 4.1(b) and their positions after 5000 cycles using the problem-solving approach 4.1(c). The results in figure shows that the continuous SCT approach takes into account grouping behavior while the problem-solving approach does not.



(a) Initial random positions.



(b) After 5000 cycles of continuous comparison.

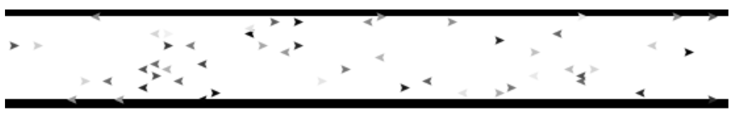


(c) After 5000 cycles, social comparison only when stuck.

Figure 4.1: **Screen shots of the comparison of implementation of the approaches for grouped pedestrian movement.**

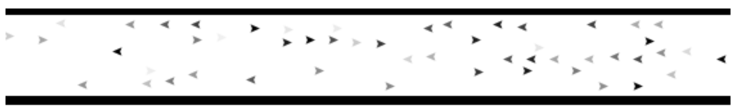
Figure 4.2 displays the screen shots of the individual pedestrian experiments. The screens show the initial positions of the agents in one of the trials 4.2(a), their positions after moving 5000 cycles using the SCT-Continuous approach 4.2(b) and their positions after 5000 cycles using the SCT-Problem-Solving approach 4.2(c).

For each of the two tasks, we compared the two trigger types, i.e., the problem-solving, and the continuous comparison. The only difference in the



(a) Initial random positions.

(b) After 5000 cycles of continuous comparison.



(c) After 5000 cycles, social comparison only when stuck.

Figure 4.2: **Screen shots of the comparison of implementation of the approaches for individual pedestrian bi-directional movement.**

runs concerned the activation of the SCT process. In the problem-solving trigger, the social comparison process was activated only when the agent was stuck and was unable to proceed towards its movement goal. In the continuous mode, the agent constantly compared itself to others and acted on this comparison. Namely the SCT process was continually active.

The results of the pedestrian grouping experiment where all the agents were divided into five groups and moved in the same direction is presented in Table 4.1. We used a hierarchical social entropy to measure the grouping behavior [4]. Lower values indicate improved grouping. Table 4.1 shows the measurement results for the SCT-Continuous approach and for the SCT-Problem-Solving approach. The results depicted in the table reveal that the SCT-Continuous approach provides improved grouping compared to the SCT-Problem-Solving approach. Each entry in the table averages the results of 15 runs; the standard deviation is provided in parentheses.

However, the conclusion is reversed when the task is changed. The SCT-

SCT-Continuous	SCT-Problem-Solving
102.36 (*sd*.19.15)	171.75 (*sd*.11.9)

Table 4.1: **Grouping measurements of the SCT-Continuous approach and the SCT-Problem-Solving approach. Lower values indicate improved grouping.**

Problem-Solving approach can perform much better. Figure 4.3 presents the results of a bidirectional traffic experiment where individual agents walked in two opposite directions on a simulated sidewalk and there were no groups, i.e., each agent constituted an individual. As is common in the literature on bidirectional pedestrian behavior, we measured the total number of lane changes. (i.e., how many times the agents needed to move left or right). The X-axis represents the density. The Y-axis represents the number of lane changes during the course of 5000 cycles. Lower values indicate improved lane changes. Each configuration was repeated 30 times. While it may seem like the SCT-Continuous approach provided slightly lower lane changes in higher density (0.19), the changes were not found to be significantly different from the SCT-Problem-Solving approach (two-tailed t-test, $alpha = 0.258$). However, in a lower density (0.07) a significant difference was found between the two approaches (two-tailed t-test, $alpha < 0.01$).

The results in the two tasks show that neither of the two approaches are superior, even within the same domain. In grouping pedestrian behavior, simulated humans should perform social-comparison continuously; but when the movement is bidirectional in low densities, they should perform it only if they are stuck.

4.2 Continuous Social Comparison with Action Selection

Our goal was to provide a single mechanism that would account for different crowd behaviors (different tasks). The results above seemingly threaten this goal, as they seem to imply that appropriate triggering of social comparison is task-dependent, and therefore, one could argue that comparison does not take place at the cognitive architecture level.

In this section, we address this argument in depth. First, we closely

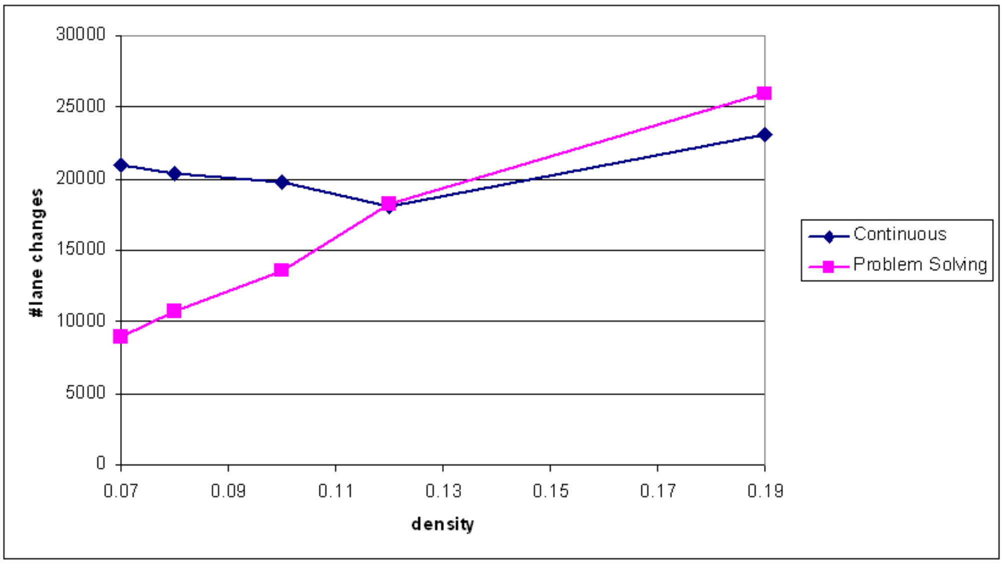


Figure 4.3: **Measurement of lane changes in the SCT-Continuous approach and the SCT-Problem-Solving approach. Lower values indicate improved lane changes.**

examine the claim and show that in fact it may still be possible to account for the results while allowing social comparison to take place within the architecture (Section 4.2.1). Then, we examine ways of weighing the proposed actions that are motivated by the social comparison process, so as to enable their selection by the architecture in a flexible manner (Section 4.2.3).

4.2.1 Social Comparison at the Architecture Level

Let us examine the conclusions of the previous section more closely. Can an architectural triggering mechanism of either type discussed above support this task-dependent behavior? The problem-solving triggering mechanism cannot emulate continuous comparison, i.e., if it is not continuously running it cannot simulate a process that is continuously running.

However, a continuous comparison mechanism may emulate a sometimes-triggered process, if the action-selection mechanism will at times ignore the actions chosen by the social comparison process. According to this view, the social comparison process should be implemented as a secondary parallel process within the cognitive architecture. Whereas normally, actions are proposed (and selected) by the architecture based on their suitability to the current goal (e.g., through means-end analysis), in our agent actions were also proposed based on their suitability to SCT. In other words, at every cycle, an agent will consider actions that will advance it towards its goal and, it will also consider social actions that seek to minimize perceived differences from other agents. Thus, the SCT-proposed actions compete with the task-oriented actions for control of the agent.

We consider two potential action-selection mechanisms which allow the competition between goal-oriented actions and socially-oriented actions. For simplicity, we describe these using a hypothetical example whereby two actions are proposed: One goal-oriented and one socially-oriented. Let us denote the weight (activation) of the goal-oriented action by α. Let us denote the weight of the social action, stemming from the social comparison process by β. Thus the following two alternative mechanisms are possible for choosing between the actions:

$max(\alpha, \beta)$. In this approach, similar to earlier work on spreading activation techniques, the action selection mechanism simply selects the action with the greatest weight.

$threshold\beta$. In this approach, the social action is selected for execution, but only if β is sufficiently high. That is, only if $\beta > C$ for some given constant C; otherwise, the goal-oriented action is selected.

In both cases, once the action is selected, it is executed. In the next decision-making cycle, new values for α and β are calculated, and again an action is selected, ad infinitum.

The discussion on an agent's α (goal-oriented weight) is not included in the scope of this thesis. For our work, it is suffice to assume that $0 \leq \alpha \leq 1$, where $\alpha = 0$ when the agent has no motivation to carry out the action, and $\alpha = 1$ when the agent is fully motivated to carry out the action. In our implementation of pedestrian behaviors (individual pedestrian and group pedestrian), all the agents in both of the implementations have identical goals (movement in their assigned direction), and their α value varies between $\alpha = 1$ when their path is clear, and $\alpha = 0$ when they are blocked. Therefore, when analyzing these behaviors we can disregard the constant α measures and focus only on the changing β measures.

4.2.2 Extended SCT algorithm

Algorithm 2 revises the earlier algorithm presented in section 3.1 to allow our revised view of the social comparison process. It differs from the earlier algorithm whereby instead of selecting an action a and executing it, it returns a recommendation for a, with a weight β. Formally, it returns a tuple $\langle a, \beta_a \rangle$.

Among all observed agents ($A_o \in O$), we calculate a similarity value $Sim(A_{me}, A_o)$, which measures the similarity between the observed agent A_o and the agent carrying out the comparison process (A_{me}). Between all agents that are not too dissimilar or too similar (i.e, agents in S), a representative agent (A_c) is selected by GetAgentForComparison(S). In our implementation A_c is the agent with the maximal similarity value within the bounds (S_{min}, S_{max}). D receives a list of features which corresponds to the differences between me (A_{me}) and the compared agent A_c. Then, an agent that agent carrying out the comparison process (A_{me}) calculates the β value, which represents the agent's attractiveness to the selected agent (A_c). The function CalculateBeta(A_c,S_{min}, S_{max}) receives the compared agent (A_c), and the similarity bounds (S_{min}, S_{max}) and returns the β value, which replaces the use of the gain function in the earlier version of the algorithm.

In the basic SCT model as presented in section 3.1, the gain function ($Gain(Sim(A_{me}, A_c))$) represents the normalized distance between the similarity with the selected agent to the two extreme values of similarity (S_{max}, S_{min}). To calculate an agent's attractiveness to the selected agent (β), we use this gain to account for normalized values (between 0 and 1).

Algorithm 2 *SCT* (O, A_{me}, S_{min}, S_{max})

```
1: S ← ∅
2: for all A_o ∈ O do {Add only agents not too similar or not too dissimilar}
3:     if S_min < Sim(A_me, A_o) < S_max then
4:         S → S ∪ A_o
5: A_c ← GetAgentForComparison(A_me, S)
6: D ← CalculateDifferences(A_c)
7: β ← CalculateBeta(A_c,S, S_min, S_max)
8: a ← SelectAction(D)
9: return ⟨a, β⟩.
```

4.2.3 Calculating β

In this section we focus on determining appropriate ways to calculate β values which shall meet the following requirements:

- Facilitate good simulation of pedestrian traffic *in both tasks* described in Section 4.1. We observed such an approach in Algorithm 1.
- Work well when the social comparison process is continuously running. Algorithm 1 fails in fulfilling this requirement, since it fails in the task of unidirectional traffic in groups.
- Preferably, be justifiable or otherwise compatible with cognitive science and psychology theory.

The β measure is a function of the agent's attraction to the observed agents (with whom it has compared itself). We distinguish between two approaches. The first approach is based on selecting an individual agent from the group and calculating the attraction to it. In Algorithm 1 the agent chosen was the agent with the highest similarity, that was still less than the maximal similar threshold S_{max}. The second approach takes an entire group of agents into account when calculating β, without singling out any particular agent.

Individual Argmax Selection: Similarity Range

In our basic SCT model, an agent compares itself to one selected agent. This individual comparison approach was successfully implemented in our previous work, and provided good results in reference to different crowd behaviors (see, for instance, its evaluation with respect to human pedestrian data [25]). In this section we present compatibility of the extended model to the basic model and also propose beta calculation to account for the timing extension.

In our basic SCT model (Algorithm 1), between all observed agents $A_o \in O$, the comparing agent selects the most similar agent A_c within the similarity range and compares itself to it. We attach a $Gain$ value to the correction action o that minimized the differences to the selected agent, which indicates the amount of effort that should be invested in the action. The $Gain(Sim(A_{me}, A_c))$ function represents the normalized distance between the similarity with the selected agent to the two extreme values of similarity

(S_{max}, S_{min}). To calculate the agent's attractiveness to the selected agent (β), we will use this gain to account for normalized values (between 0 and 1).

There is some evidence of this approach in psychology literature relevant to the social comparison theory. In particular, Volkmann [82] proposed the *range theory of social judgment*, which emphasized the relationship between what is being judged and the two extreme values of the stimulus context. In social comparison, the context is the group, which includes other people with whom one's own conditions can be compared and in our implementation we compare the similarity value.

Thus, in this variant of the extended SCT model, between all observed agents in O, S receives the group of agents with the similarity value within the bounds. A_c receives one agent from the $GetAgentForComparison(S)$ method, which selects the most similar one (with the highest similarity value within the S). D receives the vector of features with the values of 0 or 1 which indicate the feature value differences from the selected agent.

To calculate an agent's attractiveness to the selected agent (β value), we calculate the normalized distance between the similarity with the selected agent and the two extreme values of similarity (S_{max}, S_{min}). The definition of $CalculateBeta(A_c, S, S_{min}, S_{max})$ in this case is presented in 4.1.

$$\beta_{argmax} = CalculateBeta(A_c, S, S_{min}, S_{max}) = \frac{(Sim(A_{me}, A_c)) - S_{min}}{S_{max} - S_{min}} \tag{4.1}$$

Group Comparisons

One area in which the individual model fails is that it ignores the size of the group being compared. There is much evidence that the size of the group has an effect on the imitational tendencies of the individual. For instance, a well-known experiment in social sciences was performed by Milgram, Bickman, and Berkowitz [54]. The experiment involved one participant who stood in the middle of a busy street and stared into an empty spot in the sky. The experiments purpose was to examine group pressure. The results showed that when there was only one participant, there were only a few people that passed and briefly glanced up. However, when there were several participants, almost 80 percent of the passers-by also stopped and stared into the sky.

We therefore seek a model in which the number of similar agents in the group impacts β. We propose two such models, both tied to psychology literature on judgment of stimulus with respect to a context of other stimuli.

Mean Agent. Inspired by Helson's adaptation-level theory [40], we propose an alternative approach. Helson proposes that the baseline of judging a stimulus should be the mean of the stimuli that provide the context, such that the rating given to a stimulus is a function of its difference from the mean. Thus, instead of selecting one agent, i.e., the most similar and ignoring all others, we would like to take into account the group factor by looking at an abstract mean agent, and determining the similarity to it.

We created the mean agent A_{mean} and calculated the agent's attractiveness to it. The function $GetAgentForComparison$(S) in this case created the mean agent A_{mean} from the selected agents in S. Each agent was assumed to be modeled by a set of features and the mean agent was modeled by features with mean values from S[1]. The compared agent A_c was then the mean agent A_{mean}. Note that this agent did not necessarily exist. D receives the vector of features with values 0 to 1 which indicate the feature value differences between A_{me} and the mean agent. The β measure is again based on the range principle which is the normalized distance between the similarity with the mean agent and the two extreme values of similarity (S_{max}, S_{min}).

Thus the β measure is calculated as before (Eq. 4.1), but with modification of the parameters. Rather than A_c being the most similar agent, it is now a hypothetical mean agent calculated as follows:

$$\beta_{mean} = CalculateBeta(A_c, S, S_{min}, S_{max}) = \frac{(Sim(A_{me}, A_c)) - S_{min}}{S_{max} - S_{min}} \quad (4.2)$$

Range-Frequency Theory. We consider a second model, inspired by Parducci's Range-Frequency theory [61]. According to this theory, overall judgment of a stimulus should not rely only on its range to the mean, but instead should take into account its relative frequency—via its percentile rank—in the group of stimuli. Thus judgment should be modeled as a weighted sum of its range and percentile rank (frequency).

We thus propose an alternative approach for the β calculations (Eq. 4.5), which, in addition to range, also takes into consideration the group distribu-

[1] For categorical features, we use mode values.

tion (via the percentile rank of the result). The β is then a weighted sum of the range to the mean (Eq. 4.3) and frequency values (Eq. 4.4), as shown in Eq. 4.5.

The range calculation (4.3) is identical to the calculation for the mean agent model; it is the range to the hypothetical mean agent. We also calculated the percentile rank (frequency, in the terms of Parducci's theory). For all agents in the selected group (S), we calculated their similarity value to the mean agent ($Sim(A_k,A_{mean})$), and we also compared the agent's similarity to the mean agent similarity (Sim(A_{me},A_{mean}) and the compared agent's similarity). We calculated the number of agents with the same similarity value as similarity value of the compared agent (my similarity) divided by the number of agents $|S|$.

$$Range = \frac{Sim(A_{me}, A_{mean}) - S_{min}}{S_{max} - S_{min}} \tag{4.3}$$

Let $|I_{Sim}|$ denote the number of agents with a similarity value identical to mine. $|S|$ is the total number of agents. Then the frequency value $Frequency$ is calculated according to following equation:

$$Frequency = \frac{|I_{Sim}|}{|S|} \tag{4.4}$$

To compromise between range and frequency, we used the weight p to determine the proportions the range and frequency components were assigned in the weighted sum. Usually, equal weight is given to both the results. The β is the weighted sum between the Range and Frequency values and is calculated according to following equation:

$$\beta_{RF} = CalculateBeta(A_c, S, S_{min}, S_{max}) = p \cdot Range + (1 - p) \cdot Frequency \tag{4.5}$$

4.2.4 Experiments

We carried out several experiments to evaluate the hypotheses discussed in this section. The experiment design and setup were already discussed in Section 4.1.2. In Section 4.2.4 we present the results of the experiments which applied social comparison continuously, using both the $max(\alpha, \beta)$ and the threshold β action-selection mechanisms. We show that one of these mechanisms works well for the two tasks. Then in Section 4.2.4 we present the results of experiments with the three β models, comparing them (unfortunately, indirectly) to human pedestrian data.

Experimenting with Action-Selection Mechanisms

Our first task was to determine which, if any, of the two hypothetical action-selection mechanisms may be used to allow social-comparison to take place in parallel to any goal-oriented action-selection process. The two proposed mechanisms were $max(\alpha, \beta)$ (in which the highest gain wins) and threshold β (in which the socially-motivated action wins if its threshold is higher than some fixed constant C). We varied the separator C value between 0.2, 0.3 and 0.4 (chosen based on pilot experiments). The weight p in the RF model was set at 0.8.

In these experiments, we used the two mechanisms in variations on the pedestrian tasks described above. These variations included bidirectional individual movement in high-density settings, bidirectional individual movement in low-density settings, and unidirectional movement in groups. As before, in the bidirectional movement tasks, we measured performance by the accumulated number of lane changes (as before); in the unidirectional grouping task, we measured clustering by means of hierarchical social entropy [4].

In the pedestrian traffic tasks, the goal-oriented α was always set according the following rule: α is 1 if the agent's path is clear, or otherwise 0. Because of this rule—fixed along all tasks and experiments—we could control the action-selection mechanism and evaluate its performance in the different tasks, with respect only to the socially-motivated actions, proposed with weight β.

Table 4.2 shows the results of the experiments. The left column in each table indicates the β variant in use. The next two columns depict the results for the bidirectional movement task, in two different densities. The last column shows the results for the unidirectional grouping task.

Several conclusions can be drawn from these results. First, the reader should note that *all* results for the unidirectional grouping task in Table 4.2(a) (last column) are higher (worse) than the respective results in Tables 4.2(b)–4.2(d). In the bidirectional movement tasks, the results are in-

β model	Bidirectional Traffic High Density (Lane changes)	Bidirectional Traffic Low Density (Lane changes)	Unidirectional (Grouping) (Hier. entropy)
β_{argmax}	25531.75	8707.48	168.89
β_{mean}	35110.93	9864.27	170.47
β_{RF}	29199.73	9592.33	172.64

(a) $max(\alpha, \beta)$.

β model	Bidirectional Traffic High Density (lane changes)	Bidirectional Traffic Low Density (lane changes)	Unidirectional (Grouping) (hier. entropy)
β_{argmax}	26086.47	9607.27	108.41
β_{mean}	37533.33	9279.07	162.44
β_{RF}	64401.2	40128.13	136.79

(b) threshold β (threshold$= 0.2$).

β model	Bidirectional Traffic High Density (lane changes)	Bidirectional Traffic Low Density (lane changes)	Unidirectional (Grouping) (hier. entropy)
β_{argmax}	25587.87	9833.53	108.3
β_{mean}	37819.13	10837.4	155.66
β_{RF}	32349.8	7697.13	147.34

(c) threshold β (threshold$= 0.3$).

β model	Bidirectional Traffic High Density (lane changes)	Bidirectional Traffic Low Density (lane changes)	Unidirectional (Grouping) (hier. entropy)
β_{argmax}	23414	8638.27	109.19
β_{mean}	39198.8	11033.91	160.58
β_{RF}	36539.53	8769.6	149.6

(d) threshold β (threshold$= 0.4$).

Table 4.2: The results of applying two action selection mechanisms in the two tasks, for the β_{argmax}, β_{mean} and β_{RF} variants. Table (a) shows the results of the $max(\alpha, \beta)$ mechanism. Tables (b)–(d) show the results when applying the threshold β mechanism, with a threshold of $C = 0.2$, $C = 0.3$, $C = 0.4$, respectively. All results are averaged over dozens of trials (15–50). The lower the results the better.

conclusive. Thus we can conclude that the $max(\alpha, \beta)$ mechanism is inferior to the threshold β mechanism.

Second, we can conclude that the RF model is superior to the mean-agent model, when using the threshold β action-selection mechanism. In all cases except one (when $C = 0.2$), the results for the RF model improve for those of the mean-agent model.

Third, in general, the β_{argmax} model is superior to the others. This was unexpected, given its failure to account for group size. However, the results show that the β_{argmax} model provides better performance than the β_{mean} and the β_{RF} models.

Comparison with Human Pedestrian Data

The previous sets of results were all based on quantitative measures of performance, on an absolute scale where the lower the result the better. These are well-recognized measures, but they are artificial; they have not been applied to human data. Thus, we do not know the nominal values for normal human pedestrian traffic. Consequently better (lower) results on the absolute scale may in fact be unrealistic.

Hence, for final evaluation, we also conducted experiments which indirectly compare the performance of the various models to human crowd data, which we published in [25]. The experiments were carried out as follows.

First, in [25] we allowed human subjects to qualitatively compare various variants of the β_{argmax} model, based on continuous comparison, with movies of human pedestrians moving bidirectionally, in groups. The models were also compared to the *random selection* process, which is often used in the literature as baseline. While a detailed discussion of the results of the paper are outside the scope of this document, we will mention that one clear winning model—one of the β_{argmax} variants—emerged. We denote this model SCT_{argmax}. This model relied on SCT as a problem-solving activity, where the social comparison process was only triggered occasionally. As shown in Section 4.1, this type of triggering mechanism is problematic.

However, given the success of the SCT_{argmax} compared to other models, we can now use it as a basis for comparison against newer models, such as those investigated in the course of this research. In particular, we compared the results of using this model on the same task with human data, with the results of applying the various variants described above.

The results of this experiment are shown in Table 4.3. The table shows both lane-changes and hierarchical social entropy results for the same task (bidirectional movement in small groups). The table compares several models: The original (which was judged by human subjects to be the closest to human movement), shown in the second column; the β_{argmax} model, in the next column; and finally the β_{mean} and β_{RF} model, in the last two columns, respectively.

The table shows that the β_{RF} model, introduced in this paper, seems to be the closest to the original winning model in terms of the number of lane changes, and is also very close to it in terms of the social entropy measure used to evaluate grouping. While we have shown that the β_{RF} model proposed, in some cases, displays superior performance to that of the β_{argmax} model, our results are less conclusive than we would have liked. Nevertheless, these results show much promise for future development.

Measure	Baseline	β_{argmax}	β_{mean}	β_{RF}
Lane Changes	5974.5	2880.67	4283.97	5267.73
Social Entropy	22.32	25.99	22.4	21.69

Table 4.3: A comparison of different crowd models in the task of bidirectional pedestrian traffic in small groups. The first column shows the baseline, which was shown in our earlier work to be the closest to human data of previous models.

Chapter 5

The Impact of Cultural Differences on Crowd Dynamics in the Pedestrian Domain

In this chapter we examine the impact of cultural differences on crowd dynamics in the pedestrian domain. In this domain we relate to recorded pedestrian data in five different countries: Iraq, Israel, England, Canada and France. We characterize these cultures based on cultural attributes at the individual level: personal spaces, speed, avoidance side and group formations. We use an agent-based simulation to investigate the impact on the resulting macro level behavior, such as pedestrian flow, number of collisions, etc. We also examine the impact of mixed-culture pedestrians on the resulting macro-level behavior. We quantitatively validate the simulation against data from movies on human crowds in different countries.

5.1 Cultural Differences in the Pedestrians Domain

In this section we define the attributes that have an impact on pedestrian dynamics among different cultures. First, we define the cultural attributes that have an impact on pedestrian dynamics across different countries. Based on literature reviews and expert consultations we refer to the following cultural attributes: personal space, base walking speed, avoidance side, and group formations (in particular gender-heterogeneity, size, and shape, e.g., whether side-by-side, or one gender in front of the other):

- Personal space is an invisible boundary that people maintain between each other. According to Hall each person is surrounded by four invisible *"bubbles"* of space [5,30,33]: Intimate, Personal, Social and Public. Intimate distance refers to embracing, touching or whispering. Personal distance refers to interactions among good friends or family members. Social distance refers to interactions among acquaintances and public distance is used for all other interactions such as public speaking. Changes in the bubbles depend, among other things on relationships to the closest person and also on cultural background. Sociologists have found that people in different cultures maintain different distances.

- Pedestrian walking speed has also been found to be another cultural attribute [50]. In some cultures pedestrians walk much faster than in others. For example, people in England and France have faster walking paces than people in Jordan or Syria.

- We refer to the avoidance side as the side of passing other pedestrians in situations of collision avoidance. In order to avoid collisions pedestrians choose whether to avoid the other person on the right or left side. It has been found that side preference is also a cultural decision [56]. For example, pedestrians in the European continent tend to walk more on the right side of the sidewalk, whereas, in Japan or Korea, pedestrians have been reported to walk more on the left side.

- In group formations we examined the portion of pedestrians that walk as individuals versus as groups, and also who comprise the groups. For example, we differentiate between genders, homogeneous and heterogeneous groups. It has been reported that up to 70% of the people in a crowd move in groups such as families or friends versus individuals [57]. In this work we distinguish between individuals and groups and also size and gender in group formation. For example, according to experts, people in Arabic countries walk in larger groups than people in Europe. Moreover, it has also been observed that in the Arabic cultures there is a greater tendency for men to walk in front of women than in Europe, where men and women usually walk side by side.

To quantitatively characterize the examined cultures based on the presented cultural attributes, we analyzed videos of human pedestrian dynamics where pedestrians from different countries walk on sidewalks. We quantitatively measured the attributes in movies taken in five different cultures: Iraq, Israel, England, Canada and France. Then we use a pedestrian simulation to show the impact of these cultural attributes on the resulting macro-level crowd dynamics. In the following section we provide a detailed description of the video analysis process and present our results.

5.1.1 Extended Model of Social Comparison

The way pedestrians maneuver within group formations tends to vary between different cultures. For example, as stated in Arabic cultures men have shown a greater tendency to walk in front of women than in Europe, where men and women usually walk side by side. We propose to extend the SCT model to account for hierarchical comparison.

According to the social comparison theory the tendency of people to compare themselves to others differs between individuals. Social comparison researchers have reported that while some people prefer to make downward comparisons others may prefer to make their comparisons upward [84]. People make upward comparisons with other individuals they perceive to be better than themselves while people make downward comparisons with others who are considered to be worse than themselves. The main reason for these differences is the individual variance in personal and social variables. This tendency affects the target selection process, namely to whom people prefer to compare themselves and inevitably the different reactions that people may have following these comparisons.

We expand the SCT mechanism, presented in section 4.2.2, to account for hierarchical comparison, upward and downward. Each agent holds personal and social variables such as social class, comparison tendency etc. We define several social classes to which agents can belong. An agent performing the social comparison process will select several agents instead of one agent for comparison, i.e., one agent from each social class. Then, according to its sociological factors it will choose the final agent with which to compare. Consequently, the behavior parameters will be updated according to the social class.

This process is described in the following algorithm, which is executed by the comparing agent.

Algorithm 3 Hierarchical SCT (O, A_{me}, S_{min}, S_{max}, B, C)

$A \leftarrow \langle\rangle$
for $i \leftarrow 1$ to $|C|$ do **do**
    $S \leftarrow \emptyset$
    for all $A_o \in C_i$ do **do**
        if $S_{min} < Sim(A_{me}, A_o, C_i) < S_{max}$ **then**
            $S \rightarrow S \cup A_o$
    $A_i \leftarrow \text{ChooseAgent}(A_{me}, S, C_i)$
$(A_c, C_j) \leftarrow \text{GetAgentForComparison}(A, v)$
$D \leftarrow \text{CalculateDifferences}(A_c, A_{me}, C_j)$
$\beta \leftarrow \text{CalculateBeta}(A_c, O,\ S_{min},\ S_{max},\ C_j)$
$a \leftarrow \text{SelectAction}(D, C_j)$
return $\langle a, \beta\rangle$.

Algorithm 3 differs from the algorithm presented in section 4.2.2. In this algorithm an agent selects several potential agents for comparison, i.e., one agent from each social class. C represents a vector of social classes to which agents can belong. A is a vector of agents of size $|C|$, where A_i

corresponds to the agent chosen from each social class, C_i. Between all observed agents and for each social class C_i, we calculate the similarity value, and if the similarity value is within the bounds (S_{min}, S_{max}), the agent, A_o, is added to set S. Between all the selected agents from each social class C_i, a representative agent A_i, is selected. Then, agent selects one chosen agent for the comparison using GetAgentForComparison(A, v), which receives the vector of the representative agents of each social class A and the vector of sociological factors v. Thus, A_c represents the agent chosen for comparison, and C_j represents the social class to which the selected agent belongs.

D receives a list of features which corresponds to the differences between the agent that performs the comparison process, and A_c, i.e., the agent used for comparison. It also includes the social class to which the compared agent belongs, C_j. Accordingly, different actions may be recommended based on the different social classes to which the selected agent belongs such as walking behind or walking next to the selected agent. Then, an agent calculates the β value, which represents the agent's attractiveness to the selected group, as described in section 4.2.2. The function CalculateBeta(A_c,S, S_{min}, S_{max}, C_j) receives the compared agent (A_c), the selected group (S), the similarity bounds (S_{min}, S_{max}) and the social class to which the selected agent belongs (C_j), and returns the β value.

5.1.2 Video Analysis of Human Pedestrian Dynamics

Overall, we collected over one hundred hours of pedestrian footage in different locations. In some, we only collected a few minutes of video while in others many hours:

- The movies from France were recorded in Paris from the top of the Eiffel tower. The movies were taken in the afternoon and portray two streets that lead to the Eiffel tower. In total we analyzed two movies of two different locations that are 1:40 and 2:47 minutes long.

- The movies from Iraq were recorded from a web camera overlooking the yard in front of the Hussein mosque in Karbala. In total, we recorded over 30 different 3-hour videos (over 90 hours) in this location. The videos were recorded during different times of the day. About a third of the videos were irrelevant due to static views, or because the web camera was off, etc. Of the remaining videos, we randomly chose six of the movies and analyzed the first three minutes of each. Thus, in total we utilized 18 minutes of pedestrian dynamics in Iraq.

- The movies from Israel were similarly recorded from a web camera overlooking the Western Wall in Jerusalem. We recorded over 30 videos during different parts of the day, again each three hours long. A third of these videos were found to be irrelevant for the same reasons associated with Iraq and among the remaining ones we randomly selected four movies and analyzed the first three minutes of each. Thus, in total we utilized 12 minutes of films depicting pedestrian dynamics in Israel.

- The movies from Canada were video taped from one of the streets in downtown Vancouver in the morning and also in the afternoon. In total we analyzed four movies that are 0:15, 0:24, 1:18 and 3:36 minutes long.

- The movies from England were video taped in London in two different locations: Two movies from the London Eye (1:23 and 0:31 minutes long), and one from the Millennium Bridge (31 seconds long).

For the purposes of the analysis, we used a total of 45 minutes. We asked four subjects to analyze the movies in order to extract the group formations, speed, and avoidance side parameters. Each movie was analyzed by two different subjects and we used the mean value of each measure in our results. For example, to extract the group formations, the subjects counted the number of individuals and the number of groups. For each individual the subjects were asked to specify whether the pedestrian was a man or a woman. For each group the subjects were asked to specify the size of the group; couples, three people or more and also the gender composition of each group; two women, two men, a man and a woman, woman with child, etc. To estimate the speed, the subjects sampled 10 pedestrians in each movie, counting their steps within 15 seconds. To convert the steps into an estimated velocity measurement, we used the known average human step length for adults ($75cm$) [3].

To determine the personal spaces between people depicted in the movies, we used aerial photography and satellite image interpretation techniques which involve the estimation of the size of images. To enable successful measurement of the length, width and perimeter of a specific object, it is necessary to know the scale of the photo. Consequently, we measured the size of a few well-known objects for comparison with the unknown object. In each movie we tried to estimate personal spaces with two techniques: "Google Earth" to determine the objects' sizes, or comparison with the known sizes of familiar objects (such as cars or sports-field dimensions). If only one technique was feasible, then we used only one measurement; otherwise we took the mean value between the two measurements.

We present an example in which we used both techniques to estimate the distance. In one of the movies from Iraq a truck passed by the pedestrians (Figure 5.1). The standard size of a truck is 8ft (2.4384m). We measured the truck's width on the screen (marked in yellow) and found that it was 0.98cm. We then drew a line between the two people in the movie (marked in red) and found that it was 0.15cm on the screen. We thus deduced that the distance in reality is: $(0.98/0.15) \times 2.4384m = 37cm$.

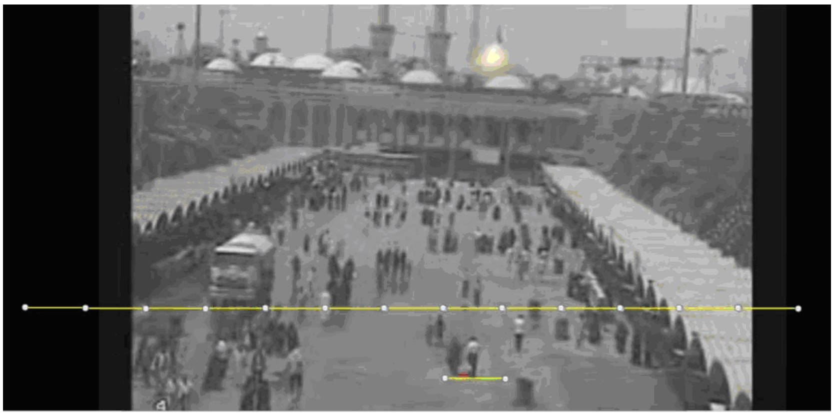


Figure 5.1: **Technique 1 for the personal space estimation**

To verify these estimations we used another method i.e., "Google Earth". We found that the width of the area is 38m (including the white shades; see Figure 5.2). Each segment in the 16-segment yellow line is therefore 2.375 meters. Again, simple math shows the distance is approximately 36 centimeters. As stated when both measurements were possible we used the mean of the two.

5.1.3 Results of Video Analysis

The results show that indeed the five countries differ from each other in the four cultural parameters. We present the analysis of the results of each

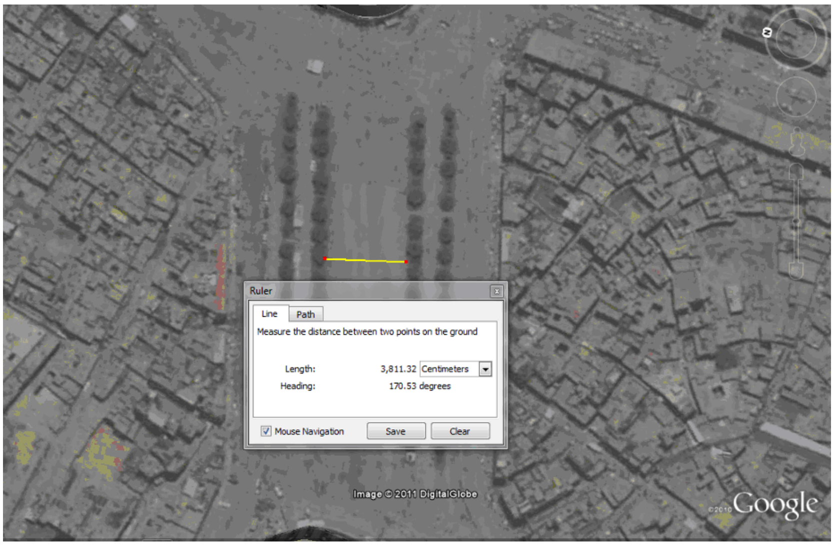


Figure 5.2: **Technique 2, using Google Earth for the personal space estimation**

parameter i.e., gender group formation, speed, passing side and personal spaces, in order to demonstrate that they actually vary between the cultures. Moreover, some of the trends were found to be consistent with the literature.

Gender Group Formations

We begin by examining the groups and their makeup. Table 5.1 presents the results of gender group formations. The first column corresponds to the examined formations. Then we present the distribution of each of these formations by culture: Iraq, Canada, Israel and France. Each value is the mean value of two measures of two subjects. In tables 5.2, 5.3, 5.4 and 5.5 we present the statistics extracted from the data received.

First we examined the portion of pedestrians that move as individuals versus groups. Table 5.2 presents the results. The first column corresponds to the formation (individuals or groups). Then we present the distribution of the pedestrians in each culture examined. The results show that in Vancouver, Canada, people move more as individuals than as groups. In every other country there is a higher tendency of pedestrians to move as groups.

We also checked whether pedestrians who move in groups tend to do so

Formation	**Iraq**	**Canada**	**Israel**	**England**	**France**
1 man	20.9%	42.4%	33.3%	12.4%	9.21%
1 woman	6.88%	17.3%	14.6%	5.53%	4.61%
2 men	15.4%	14%	15.7%	24.9%	14.5%
2 women	12.3%	9.05%	11.8%	10.1%	11.8%
1 man next to 1 woman	5.22%	4.94%	9.27%	24.9%	35.5%
1 man in front of 1 woman	2.61%	0	0.36%	3.69%	5.26%
1 man & 1 child	0.71%	0	1.78%	3.69%	1.32
1 woman & 1 child	2.14%	0	1.43%	3.69%	0
3 men	8.9%	7.41%	6.42%	5.53%	0
3 women	7.47%	4.94%	4.28%	0	1.97%
1 man next to 2 women	4.27%	0	0.53%	2.76%	1.97%
1 man in front of 2 women	1.78%	0	0	1.38%	0
2 men & 1 child	0.71%	0	0	0	0
2 women & 1 child	1.42%	0	0	0	0
1 man, 1 woman & 1 child	1.42%	0	0.18%	1.38%	5.92%
2 men & 1 woman	0.71%	0	0	0	7.89%
2 men, 1 woman & 1 child	0.47%	0	0	0	0
2 women & 2 children	0.95%	0	0	0	0
3 men & 1 child	1.42%	0	0	0	0
1 man & 3 women	0	0	0	0	0
4 women	4.27%	0	0	0	0

Table 5.1: **Distribution of gender group formation**

Formation	**Iraq**	**Canada**	**Israel**	**England**	**France**
Individuals	28%	60%	48%	18%	14%
Groups	72%	40%	52%	82%	86%

Table 5.2: **Group formation: Individuals versus groups**

in homogeneous or heterogeneous gender groups. The results are presented in Table 5.3. The findings show that in Iraq, Canada, Israel and England, pedestrians move more in homogeneous gender groups. Indeed, in France we observed many couples, i.e. a man and a woman, who move together.

We also examined the pedestrian cultural tendency concerning the sizes

Groups	Iraq	Canada	Israel	England	France
Heterogeneous	23%	12%	21%	42%	66%
Homogeneous	77%	88%	79%	58%	34%

Table 5.3: **Homogeneous versus heterogeneous gender groups**

of the groups. For all the pedestrians that move in groups we provide the statistics on their distribution into groups of several sizes, such as groups of 2, 3, 4 and more. Table 5.4 presents these results which show that in Iraq there is a higher tendency to move in larger groups than in the other countries examined.

Group size	Iraq	Canada	Israel	England	France
Groups of 2	64%	77%	84%	91%	85%
Groups of 3	30%	23%	16%	9%	15%
Groups of 4 and more	6%	0	0	0	0

Table 5.4: **Size of group formations**

According to experts, men in Arabic countries demonstrate a greater tendency to walk in front of women than in other countries, where they usually walk side by side. We checked whether this finding was supported in the videos from Iraq and compared it to the videos from the other countries. The results presented in Table 5.5 show that in Iraq this tendency was observed in 33% of the couples, demonstrating a higher tendency than in the other countries examined.

Formation	Iraq	Canada	Israel	England	France
Man next to woman	67%	100%	96%	87%	87%
Man in front of woman	33%	0	4%	13%	13%

Table 5.5: **Man and woman walking formation**

Pedestrian speed

We then examined individual speed, and its variance based on gender and grouping in the different cultures. Table 5.6 presents the results of pedestrians' speed (measured in steps per 15 seconds; the conversion to distances introduces noise that is not relevant at this stage. We will discuss the noise in the section on the simulation comparison). Again, the first column corresponds to the examined formations. Then we present the mean speed of two samples of each examined formation for each culture: Iraq, Canada, Israel and France. The values are the number of footsteps of the observed pedestrians within 15 seconds. As in noted in the previous section, we then present the statistics that we extracted from the data received.

Formation	**Iraq**	**Canada**	**Israel**	**England**	**France**
1 man	25.3	27.8	26.7	28.7	27.3
1 woman	22.1	27.6	24.9	23.5	26
2 men	23.2	27.8	24.5	26.2	26.3
2 women	20.6	31.2	22.6	24.1	26.3
1 man next to 1 woman	23	28	22.5	25	24.8
1 man in front of 1 woman					23
1 man & 1 child			31.5		
1 woman & 1 child					
3 men	23	30.5	25	28.7	
3 women	20	26.2	30		
1 man next to 2 women	23			20	
1 man in front of 2 women	22				
2 men & 1 child					
2 women & 1 child					
1 men, 1 woman & 1 child					23.8
2 men & 1 woman					24
2 men, 1 woman & 1 child					
2 women & 2 children					
3 men & 1 child					
4 woman	25.9				

Table 5.6: **Pedestrian speed**

Table 5.7 shows that men walk faster than women in all the cultures that were observed. As for the difference in the cultures, Iraqi pedestrians walk the slowest (this supports earlier research [50]). Moreover, in Iraq men as well as women walk slower than the pedestrian of the other countries observed.

Formation	Iraq	Canada	Israel	England	France
men	25.3	27.8	26.7	28.7	27.3
women	22.1	27.6	24.9	23.5	26

Table 5.7: **Speed of men versus women**

Next, we examined the effects of grouping on speed. In this case we examined the mean speed of pedestrians who move as individuals versus the mean speed of pedestrians who move in groups. The results presented in Table 5.8 show that in all cultures people as individuals move faster than people in groups.

Formation	Iraq	Canada	Israel	England	France
Individuals	25.1	28.6	25.7	26.5	26.6
Groups	23	27.3	24.6	25	24.9

Table 5.8: **Speed of individuals versus groups**

We also examined whether there is a difference in mean speed between homogeneous and heterogeneous gender groups of all the pedestrians that move in groups. Moreover, we examined whether there is a difference in speed among groups of men versus groups of women. The results summarized in Table 5.9 show that in Iraq and England the group of women are the slowest of the the groups. However, in all the cultures the groups of men are the fastest of the groups.

Avoidance side

In this section we present the results of the pedestrian avoidance side. The results depicted in table 5.10 presents . The first column correspond to right or left avoidance side and then we presents the distribution of each examined cultures. The results show that in Iraq, Canada and England the pedestrians prefer the right side while in Israel and France they prefer the left side.

Formation	Iraq	Canada	Israel	England	France
Mixed groups	23.4	25.8	22.8	24.5	24
Men homogeneous groups	24.1	28.8	26.3	26	26.8
Women homogeneous groups	21.5	26.5	24.5	23.8	25.6

Table 5.9: **Speed of homogeneous versus heterogeneous gender groups**

Passing side	Iraq	Canada	Israel	England	France
Right	62%	63%	41%	77%	45%
Left	38%	37%	59%	23%	55%

Table 5.10: **Passing side**

Personal spaces

Finally, the video analysis shows that there are cultural differences in personal spaces. Table 5.11 provides the personal spaces within groups, as well as the mean personal space. We examined whether there is a difference in the personal spaces kept by men and women in the same group. We distinguished between heterogeneous gender groups, homogeneous male groups and homogeneous female groups. The results show that in Iraq, Israel and France women keep less personal space than men. The biggest gap between the group of men and the group of women was observed in Iraq. The results of the mean space shows that in Iraq pedestrians who walk in groups maintain the smallest personal space than in the other cultures observed.

Group Type	Iraq	Canada	Israel	England	France
Mixed gender	26.5		46	50.3	35
Men only	43.8	65.8	66.5	49.5	57.5
Women only	18.3	70	50.3	52	40.5
Mean space	32.7	67.9	57.9	50.3	41.7

Table 5.11: **Personal space kept by men and women within the same group.**

5.2 The Impact of Cultural Differences on Pedestrian Dynamics: Evaluation

After establishing that the parameters chosen do indeed vary significantly between cultures, we used an agent-based simulation to examine their effect on macro-level pedestrian dynamics. We used the popular OpenSteer [65] as the simulation platform. We simulated a sidewalk where agents can move in a circular fashion from east to west, or in the opposite direction. Each agent had a limited visual distance (beyond this distance it could not see). Agents were not allowed to move through other agents, in case of possible collision the agents attempted to avoid each other. The base pedestrian model was SCT [25], which was implemented fully and then extended to support the parameters noted above. The modifications to the original algorithm are described in Section 5.1.1.

To enable visibility of group formations, each agent was given a color: Individual men marked in dark green and individual women marked in gray. In families, the husband was blue, the wife pink, and the children, who also had a smaller radius, were yellow. Agents in groups (non-family) appeared in light green if they were females and in oranges if they were males.

To account for cultural differences, each agent contained a set of cultural variables such as speed, personal spaces and avoidance side. Moreover, to account for group formations, each agent contained the following variables: group ID and social factors such as the agent's comparison tendency and the agent's social class which influence the agent's social comparison process (the SCT process) to select the agent for comparison, as described in section 5.1.1. To enable the most accuracy possible in the simulation, we translated the data on the cultural variables received from the analysis of the human movies as described below.

- **Personal space** - recall, there are four spaces which people maintain: intimate, personal, social and public. Due to the limitations of the simulation, we modeled only three of them: *personal*, *social* and *public*. Hall [30–33] reported two settings of distances for these three spaces: *close* and *far*. *Close* was defined as a personal distance of 46 cm, social distance of 120 cm, and a public distance of 370 cm. *Far* was defined as a personal distance of 76 cm, social distance of 210 cm, and a public distance of 760 cm. Because of the high possibility of noise in our estimation of the personal spaces in human pedestrian movies and since we also measured only the distances among couples (which is personal rings), we used Hall's values of close and far. For each value that we received from the analysis of the human movies we examined how close the value was to one of Hall's rings. In the translation for the simulation, we normalized all the distances based on the shortest (46cm). Thus, the three values for close ($\langle 46, 120, 370 \rangle$), were translated into simulation distances of $\langle 1, 2.6, 8 \rangle$ and for far ($\langle 76, 210, 760 \rangle$) were translated into $\langle 1.65, 4.56, 16.5 \rangle$.

- **Speed** - In our simulation we defined three speeds of walking: slow walk, average walk and faster walk. We analyzed the data extracted from the different cultures, and divided the speed samples into the three groups of speed in our simulation. Among all the received samples of speed, we computed the 33th percentile and the 67th percentile, in order to attain the two separators between these three groups, and received the values of 24.0 and 27.0, respectively. Thus, the three ranges of speed were: [20-24), [24-27), [27-31.5), where 20 and 31.5 were the minimum and maximum values, respectively, that were been sampled. Then we estimated the ratios between the speeds. Taking the mean of every range we attained: 22, 25.5, 29.25 steps per 15 seconds. Various resources suggest 75cm as a solid estimation for a human's average step [3] and we converted the number of steps we received per 15 sec. to m/s which resulted in 1.1 m/s, 1.27 m/s and 1.46 m/s. Then we examined what speed in our simulation would give us the average speed of 1.1 m/s. We found that by using the 2.27 speed in our simulation we attained an average speed of 1.1 m/s and based on the received ratios between the speeds examined in human behavior we revealed the following speed level rates in our simulations: 2.27, 2.62 and 3.01.

- **Avoidance side**. In situations of possible collision an agent will choose whether to avoid the other agent on the right or left side. Each agent has a cultural preference for the avoidance side. This variable was initialized in the beginning of the simulation according to the analyzed human pedestrian dynamics portrayed in the movies and according to the culture to which the simulated agent belonged.

In all the experiments described below, we examine the impact of individual cultural differences on the resulting macro-level pedestrian behavior, as measured by the following standard measures:

- Pedestrian flow: number of agents that cross a certain line divided by the width of the line and the time the process takes
- Mean speed: over all agents (this is the difference from the set individual speed, which each agent may or may not be able to achieve)
- Mean number of collisions: between two agents, averaged over all agents
- Mean number of lane changes: the number of direction changes of the agent that are above a predetermined threshold, averaged over all agents

5.2.1 Experiment 1: Impact of each of the cultural parameters on pedestrian dynamics

In this section we examine the impact of each of the cultural parameters on the overall pedestrian dynamics. In all the experiments in this section, we set the sidewalk as 110×20 and the number of agents at 100. To account for group formations we divided our agents into two categories, 30% individuals and 70% in groups as observed in some of the movies, and also in [57]. Furthermore, we divided the agents into different group sizes and gender formations, namely couples of women, groups of 3 men, mixed gender couples, etc. as follows:

- Individuals: 30%
- Groups: 70%
  - 5/7 in formations of groups of 2, which consisted of:
    - Men couples: 33%
    - Women couples: 33%
    - Mixed couples: 33%
  - 1/7 in formations of groups of 3, which consisted of:
    - Groups of 3 men: 50%
    - Groups of 3 women: 50%
  - 1/7 in formation of groups of 4, which we defined as husband, wife and 2 children.

Speed

First we examined the influence of the mixed speed population on the pedestrian behavior produced. We initialized the passing side of all the agents as the right side, the personal space of all the agents as close and the group formation as defined in section 5.2.1. We varied the percentage of agents who walked at a low (1.0) speed (versus a fast speed of 1.33): either 0% at a low speed, 20%, 50%, 80% or 100% and we examined the impact of the mixed speed population on the flow of the pedestrians, the mean number of collisions and the mean number of lane changes.

Figure 5.3 illustrates the effect of the mixed speed population on the mean number of collisions. The results show that the population that moves with the highest speed incurs the lowest number of collisions. The average number of collisions when all the agents walk at a high speed was 0.27. The highest number of collisions was found in the mixed population where 50% walked at a low speed and 50% at a high speed (mean value: 0.5). Moreover, this was found to be significantly different than populations that walked at homogeneous speeds where all agents either walked at a high speed or a low speed (two tailed t-test, $alpha < 0.01$ in both cases).

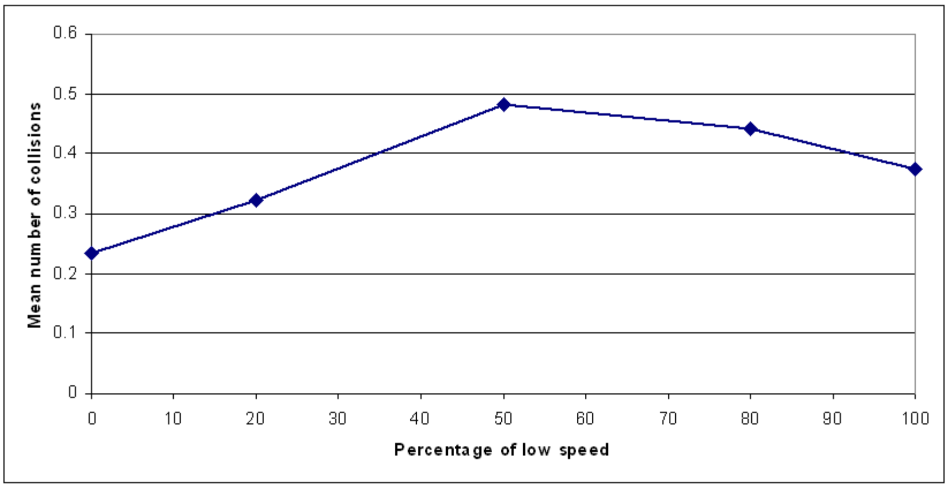


Figure 5.3: **The effect of mixed speed on the mean number of collisions**

We also examined whether the agents' mixed speeds have an effect on the number of lane changes. The findings presented in Figure 5.4 show that homogeneous speed (low or high) result in less lane changes. According to the two-tailed t-test, there is no significant difference between the high speed population and the low speed population in the number of lane changes ($alpha = 0.2$). Moreover, the mixed speed populations demonstrated the highest number of lane changes. The number of lane changes in the population where 50% of the agents walked at a high speed and 50% at a low speed, was found to be significantly different than the population of agents that all walked at a low speed and the population of agents that all walked at a high speed(two tailed t-test, $alpha < 0.01$ in both cases).

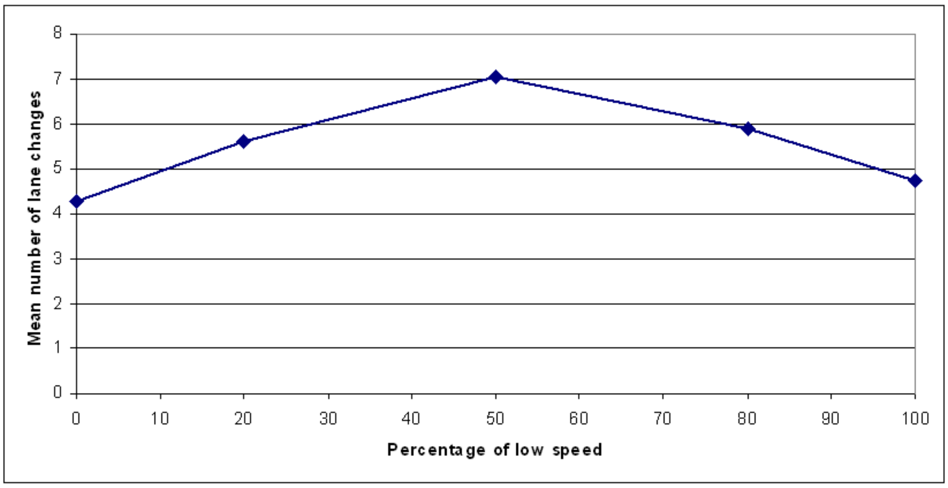


Figure 5.4: **The effect of mixed speed on the mean number of lane changes**

Figure 5.5 shows the influence of the mixed speed population on the flow of pedestrians. The results are not surprising, the more agents that move at a higher speed the higher the flow. As we can see in the results, the highest flow was found in the population of agents that all walked at the highest speed and the lowest flow was of the population with lowest speed. However, an interesting finding was the ratio between the changes in the population which caused changes in the flow. For example if we increased our population from 0% low speed to 20% low speed the flow decreased by 6%. Moreover, there was only 1% difference in the flow between the population of agents that all walked at the lowest speed and the population of the agents in which 80% walked at the lowest speed.

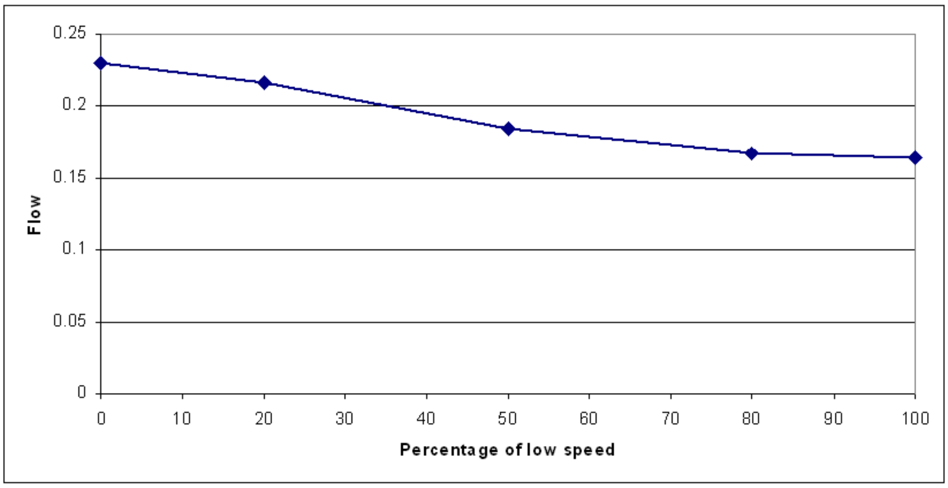


Figure 5.5: **The effect of the mixed speed on the flow**

Personal space

In this experiment we examined whether the difference in personal spaces among the agents had an impact on the pedestrian behavior produced. We initialized the avoidance side of all the agents to the right, the speed of all the agents to 1 (which is a slow pace) and the group formation was defined as detailed in section 5.2.1. We varied the percentage of agents with close personal space (versus far personal space) 0%, 20%, 50%, 80% or 100% and we examined its impact on the flow of pedestrians, the mean speed, the number of collisions and the number of lane changes.

First, we examined whether personal space had an impact on the number of collisions between the agents. The results shown in Figure 5.6 demonstrate that a significant difference was revealed in the number of collisions between the agents with close personal space and those with far personal space (two tailed t-test, $alpha = 0.01$). The mean number of collisions of agents was 0.47 for those with close personal space and 0.49 for those with far personal space. Though it seems that the difference between these values is not large it was found to be statistically significant. Surprisingly, the lowest number of collisions was found to be among the mixed group where 50% of agents maintain a close personal space while walking and 50% maintain a far space. Moreover, a significant difference was found between the homogeneous far personal space group (all agents maintained a far personal space) and the heterogeneous group (50% of the agents maintained a close space and 50% maintained a far space) according to the two tailed t-test ($alpha = 0.01$). However, no significant difference was found between the homogeneous close personal space group (all agents maintained a close personal space) and the heterogeneous group (50% of the agents maintained a close personal space and 50% maintained a far personal space) according to the two tailed t-test ($alpha = 0.09$).

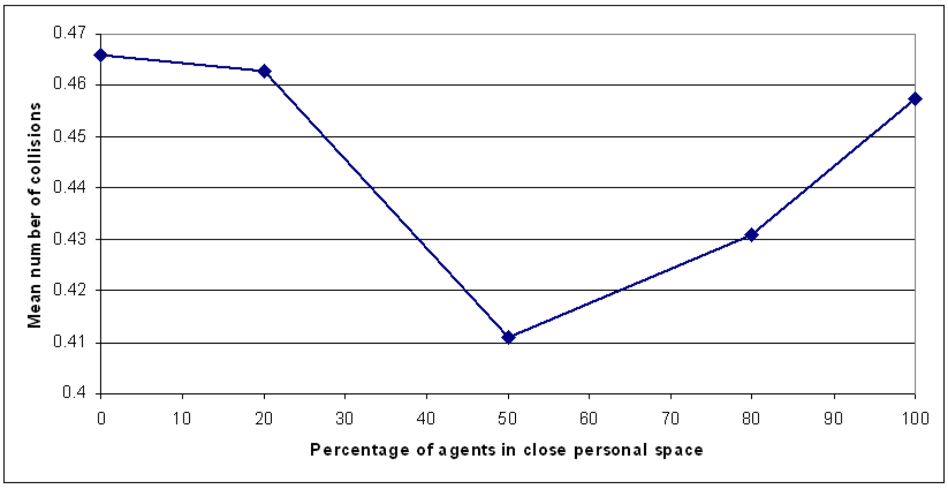


Figure 5.6: **The effect of personal space on the mean number of collisions**

We then examined whether personal space has an impact on the number of lane changes. The results are presented in graph 5.7. While it seems there is almost no difference in the results, the difference in the number of lane changes between agents maintaining close personal space and those maintaining far personal space was found to be statistically significant according to the two tailed t-test ($alpha < 0.01$). The agents that maintained close personal space made less lane changes. The results also show that there is a significant difference between the homogeneous groups (all agents with close personal space or all agents with far personal space) and the heterogeneous group (50% of the agents with close personal space and 50% with far) according to the two tailed t-test ($alpha = 0.01$ and $alpha = 0.03$, respectively).

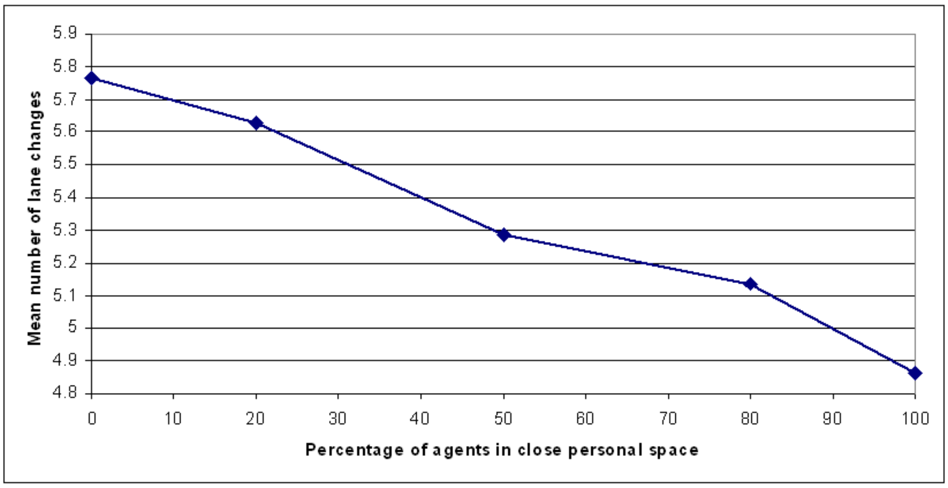


Figure 5.7: **The effect of personal space on the mean number of lane changes**

Figure 5.8 shows the results of the effect of personal spaces of the agents on their speed. The results show that agents that maintain close personal space have a higher mean speed than agents that maintain a far personal space, even though both groups were initialized with the same speed. Moreover, a significant difference was found between agents with close personal space and far personal space (two tailed t-test, $alpha < 0.01$). Differences in the mean speed were also revealed between the homogeneous groups (all agents maintained close personal space or all maintained far personal space) and the heterogeneous group (50% of agents kept a close personal space and 50% kept a far space) according to the two tailed t-test ($alpha < 0.01$ in both cases).

In addition we examined the impact of the personal spaces of the agents on the flow. As depicted in Figure 5.9, the results show that agents with close personal space demonstrate a higher flow than agents with far personal space. As we have shown in our results presented earlier, the agents that move

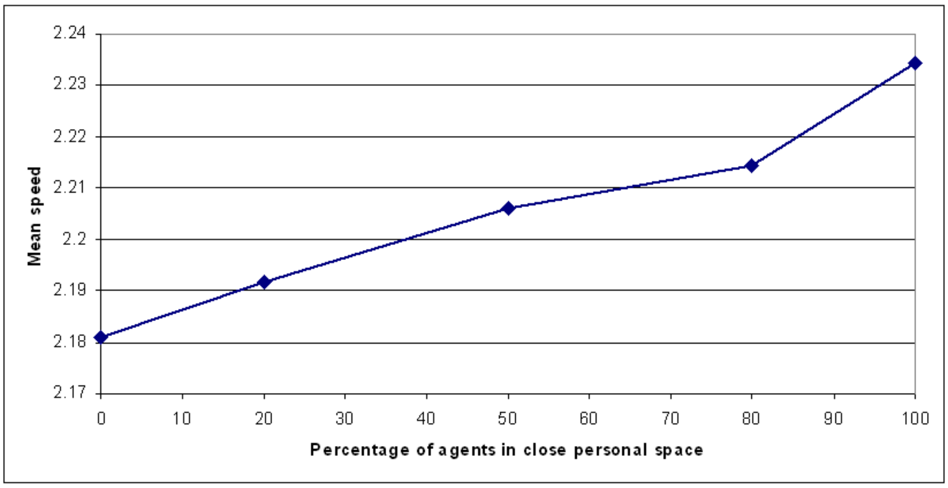


Figure 5.8: **The effect of personal space on the mean speed**

while maintaining a far personal space have a higher number of collisions and a higher number of lane changes than agents that maintain close personal space, which affect their mean speed and eventually their flow.

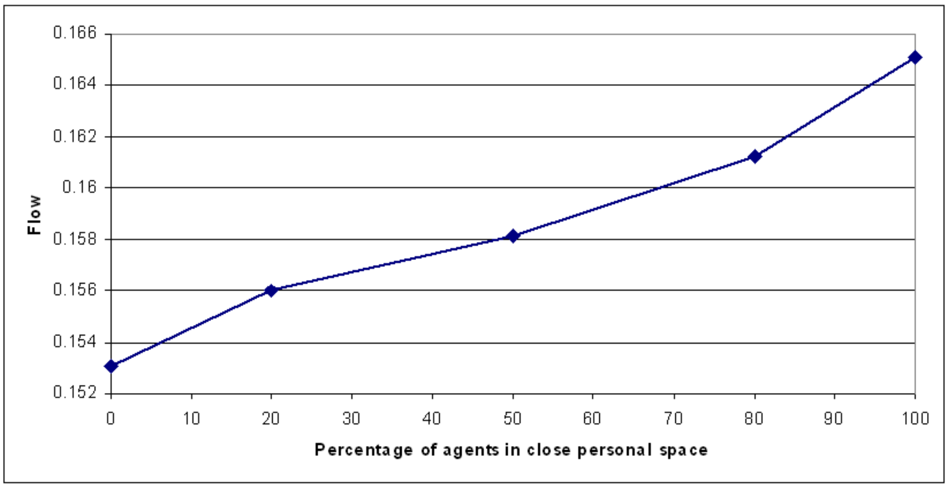


Figure 5.9: **The effect of personal space on the flow**

Avoidance side

Furthermore, we examined whether the pedestrian's avoidance side has an impact on the pedestrian dynamics. In this experiment we initialized the speed of all the agents at a slow pace, the group formation was set as detailed in section 5.2.1 and the personal space of all the agents was defined as close. We varied the percentage of agents with right avoidance side (versus left avoidance side) between 0%, 20%, 50%, 80% and 100% and we examined the impact of these mixed populations on the pedestrian's flow, mean speed, number of collisions and number of lane changes.

First, we examined whether the agent's avoidance side had any impact on the number of collisions between the agents. The results presented in Figure 5.10 show that the lowest number of collisions was found in homogeneous groups where all agents used the right avoidance side or all agents used the left avoidance side. The highest number of collisions was found in the heterogeneous group where 50% of agents used the right avoidance side and 50% used the left avoidance side. Moreover, the difference between the homogeneous groups and the heterogeneous group was found to be statistically significant according to the two tailed t-test ($alpha < 0.01$ in both cases).

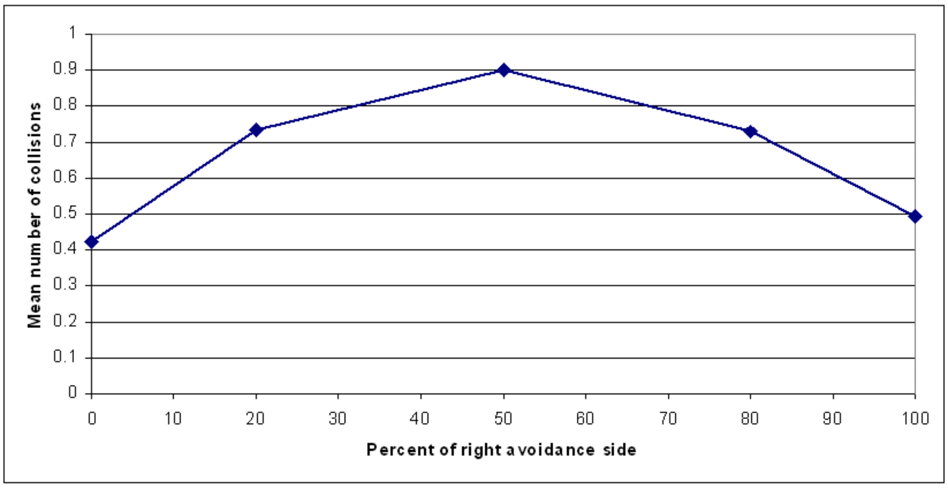


Figure 5.10: **The effect of the passing side on the mean number of collisions**

Figure 5.11 represents the results of the effect of agents' avoidance side on the number of lane changes. Similar to the previous results, the lowest number of lane changes was found in the homogeneous groups where all agents either used the right or the left avoidance side and the highest number of lane changes was found in heterogeneous group where 50% of the agents used the right avoidance side and 50% used the left avoidance side. Again a significant difference was found between the homogeneous groups and the heterogeneous group according to the two tailed t-test ($alpha < 0.01$ in both cases).

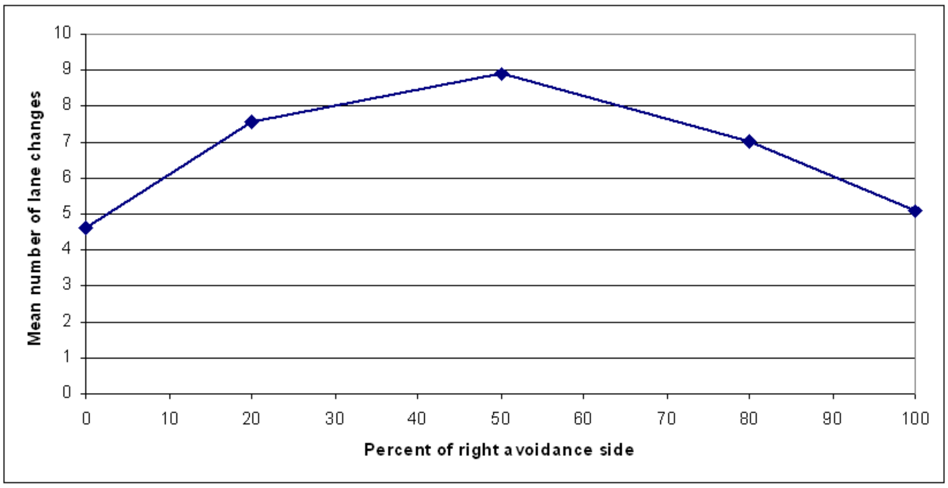


Figure 5.11: **Effect of the passing side on the mean number of lane changes**

We also examined the impact of the avoidance side on the agents' mean speed. The results, as depicted in Figure 5.12, reveal that the homogeneous group of agents had a higher mean speed than the heterogeneous group of agents. Moreover, a significant difference was revealed between these groups, according to two tailed t-test, $alpha < 0.01$ in both cases.

We then examined the impact of the agents' avoidance side on their flow. Figure 5.13 clearly shows a higher flow among the homogeneous groups of agents than the heterogeneous group of agents.

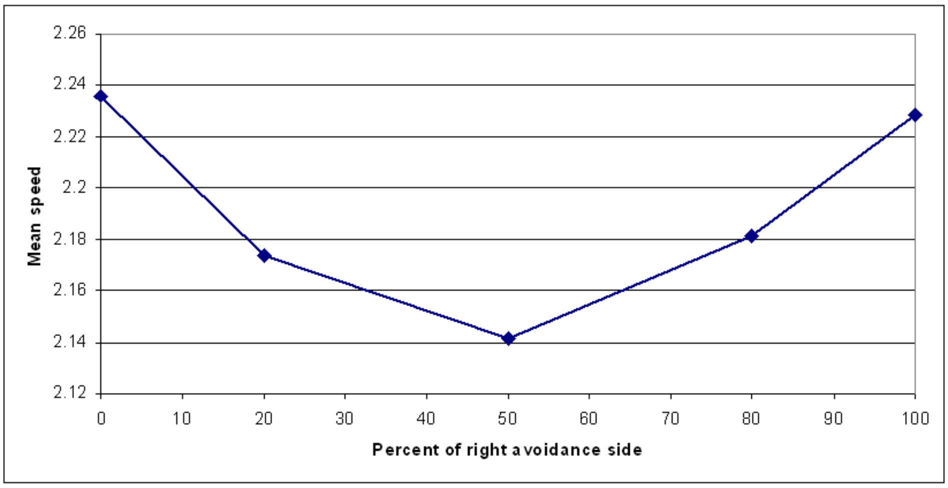


Figure 5.12: **Effect of the passing side on the mean speed**

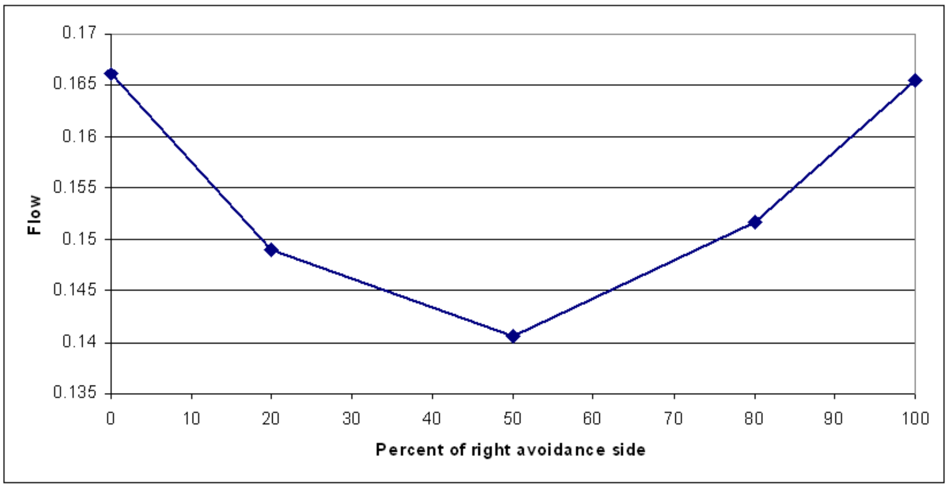


Figure 5.13: **Effect of the passing side on the flow**

Group formations with a fixed speed

In this experiment we examined whether the pedestrians movement in different groups has an impact on the pedestrian behavior produced. We initialized the avoidance side of all the agents to the right, the speed of all the agents at a slow pace and the personal space of all the agents as close. We varied the percentage of agents that move in groups (versus individuals) between 0%, 20%, 50%, 80% or 100% and we examined their impact on the pedestrians' flow, mean speed and the number of collisions. The distribution into the different sizes and gender formations of all the agents that walked in groups was the same as described above in Section 5.2.1.

Figure 5.14 displays the effect of the groups on the number of collisions. The results clearly show that the higher the number of groups in the population the higher the number of collisions. Moreover, a significant difference was found in the number of collisions between populations where all agents moved in groups and populations where all the agents moved as individuals, according to the two tailed t-test ($alpha < 0.01$).

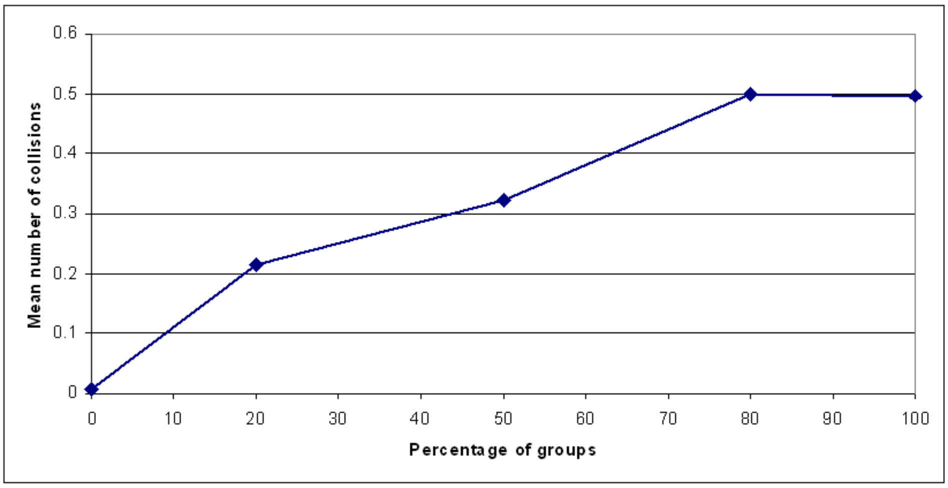


Figure 5.14: **Effect of groups on the mean number of collisions**

Moreover we examined the influence of groups on the number of lane changes. The results are illustrated in Figure 5.15. The findings clearly show that the population in which all agents walk individually have the lowest number of lane changes. There is a significant difference in number of lane changes between the population in which all the agents walk in groups and the population in which all the agents walk individually, according to the two tailed t-test ($alpha < 0.01$). However, no significant difference was found between the homogeneous population where all agents walk in groups and the heterogeneous population where 50% of agents walk in groups and 50% walk as individuals, according to the two tailed t-test ($alpha = 0.1$).

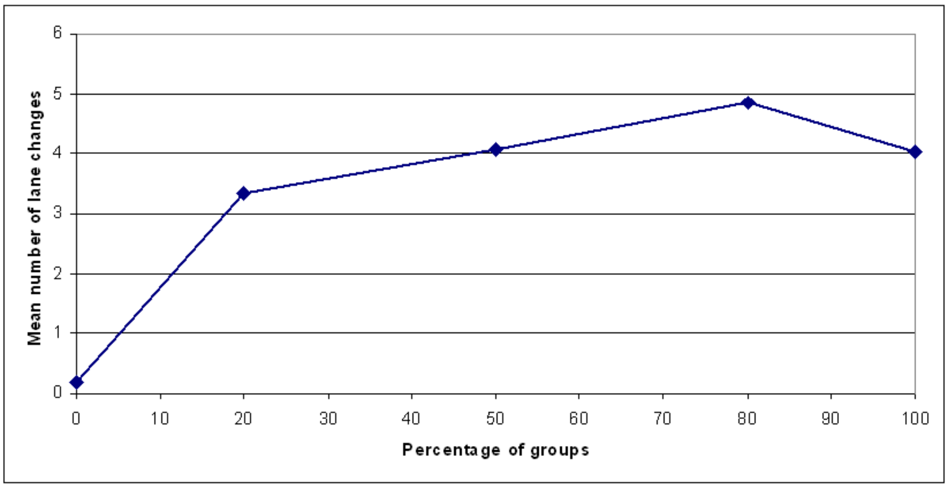


Figure 5.15: **Effect of groups on the mean number of lane changes**

Then we examined whether groups affect the pedestrian speed. The results in Figure 5.16 show that the population in which all the agents walk in groups walk at a higher speed. This finding was a bit unexpected. However, the main reason for this phenomenon is that the agents in groups occasionally accelerate to a higher speed in order to maintain the formations.

Finally we examined the influence of group formations on the pedestrian flow. The results presented in Figure 5.17 show that agents that walk individually (0% groups) display the highest flow.

Group formations with varied speeds

Similar to the previous experiment we examined the impact of groups on pedestrian dynamics. However, as we have shown in section 5.1.3, gender and different group formations walk at different speeds. In this experiment we initialized the speed of each formation (individual men, individual women, groups of men, groups of women and mixed groups) based on the analysis

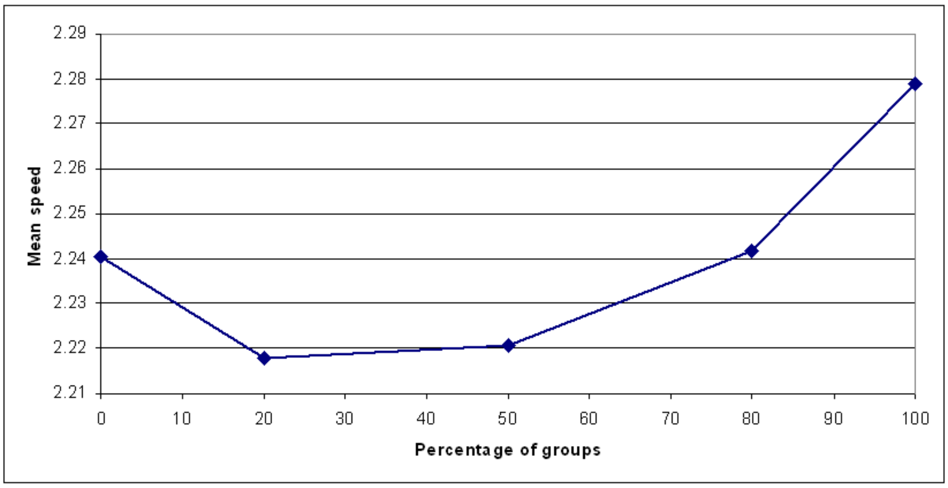


Figure 5.16: **Effect of groups on the mean speed**

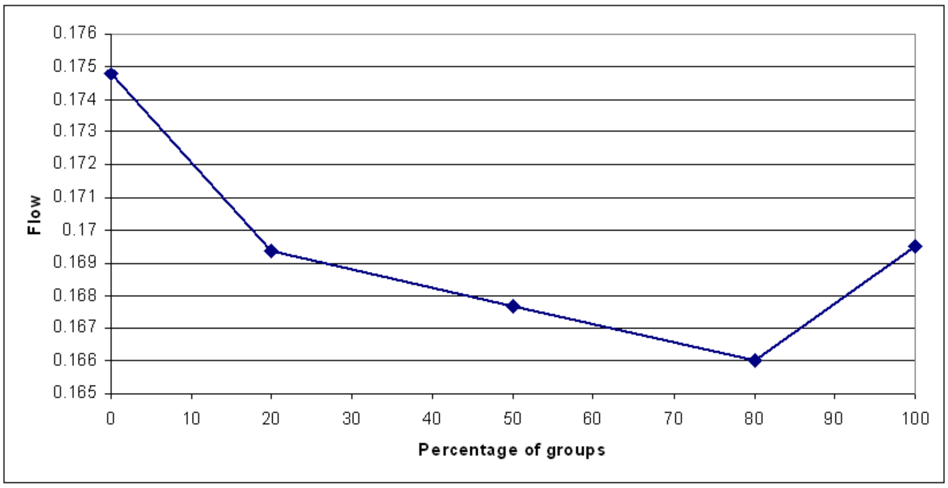


Figure 5.17: **Effect of the groups on the flow**

of the data taken from the human movies (section 5.1.3) and on the mean value of all the five cultures we sampled.

Table 5.12 presents the mean values of the different formations. The first column corresponds to the different formations and the second column corresponds to the mean speed across all five cultures in the specific formation. The results show that individual men walk at the highest speed while the group of women walk at the lowest speed.

Formation	**Mean speed (#footsteps per 15 sec.)**
Individual men	27.3
Individual women	25.3
Mixed group	23.9
Men homogeneous group	25.9
Women homogeneous group	23.7

Table 5.12: **Mean speed of different formations in the human video analysis**

Again in this simulation we initialized the avoidance side of all the agents to the right and the personal space of all the agents to close. However, the speed was set according to the formation to which the agent belonged. We varied the percentage of agents that walked in groups (versus individually): 0%, 20%, 50%, 80% or 100% in groups and we examined their impact on the pedestrians' flow, mean speed and number of collisions.

First we examined the impact of group formations on the pedestrians' number of collisions. Figure 5.18 shows that individual agents had the lowest number of collisions. The highest number of collisions was observed in the mixed population, where 80% of agents walked in groups and 20% individually (mean value of collisions 0.63), i.e. even higher than in the homogeneous population in which all the agents walked in groups (mean value of collisions 0.57). Moreover, this finding was found to be significantly higher according to the one tailed t-test ($alpha < 0.01$).

In this experiment we also examined the impact of groups on lane formation. In Figure 5.19 we can see that the higher the number of groups the higher the number of lane changes. A significant difference was revealed in the number of lane changes between the population in which all the agents walk in groups and the population in which all the agents walk individually, based on the two tailed t-test ($alpha < 0.01$). Moreover, a significant difference was found between the homogeneous population in which all the agents walk in groups or all walk individually and the heterogeneous population in

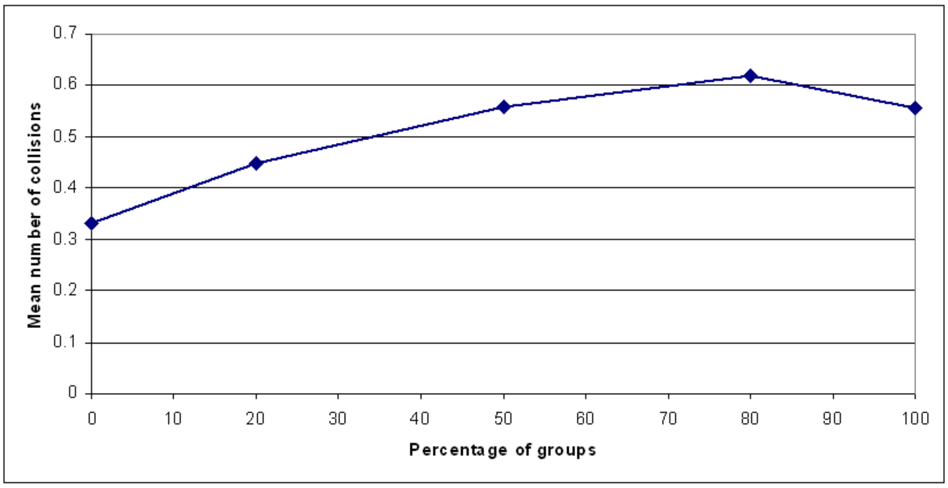


Figure 5.18: **The effect of groups on the mean number of collisions**

which 50% of the agents walk in groups and 50% walk individually, according to two tailed t-test ($alpha < 0.01$ in both cases).

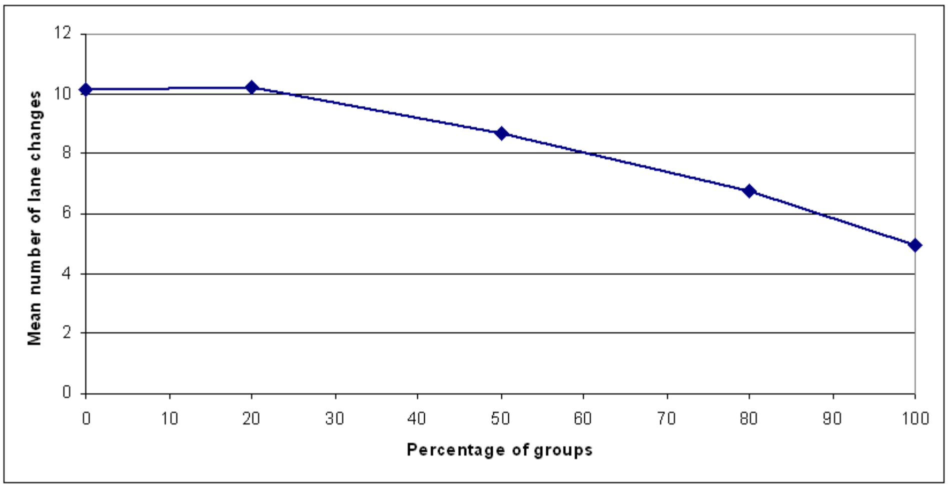


Figure 5.19: **The impact of groups on the mean number of lane changes**

We also examined the impact of groups on the pedestrian mean speed. Figure 5.20 shows that the higher the number of groups the lower the mean speed of the agents. As in the previous experiment a significant difference was revealed in the agents' mean speed between the population in which all the agents walked in groups and the population in which all the agents walked individually based on the two tailed t-test ($alpha < 0.01$). Moreover, a significant difference was also revealed between the homogeneous population in which all the agents walk in groups or all the agents walk individually and the heterogeneous population in which 50% of the agents walk in groups and 50% walk individually based on the two tailed t-test ($alpha < 0.01$ in both cases).

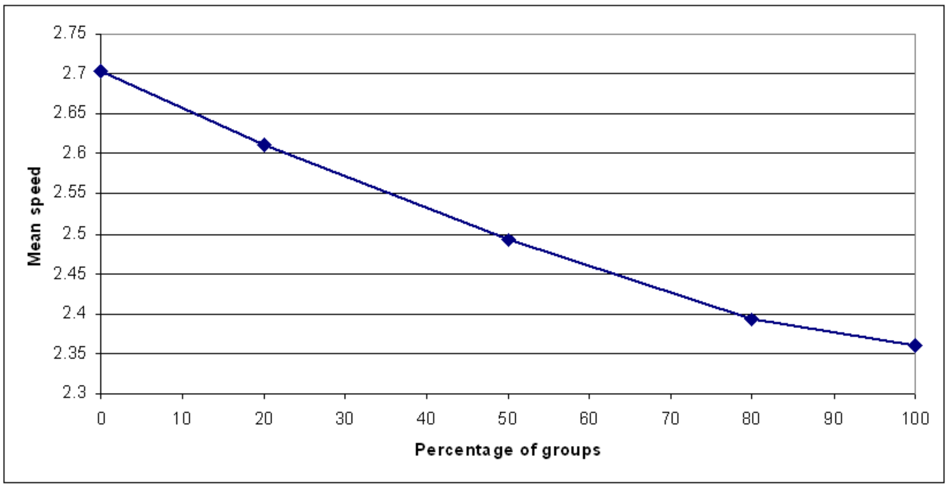


Figure 5.20: **The effect of groups on the mean speed**

Finally we examined the influence of group formation on the pedestrian flow. The results in Figure 5.21 show that the agents that walk individually (0% groups) depict the highest flow. Moreover, we found that the higher the number of groups the slower the flow.

5.2.2 Experiment 2: Differences between cultures

In this section we present our findings concerning the different cultures, i.e., Iraq, Israel, England, Canada and France. We examined whether they have

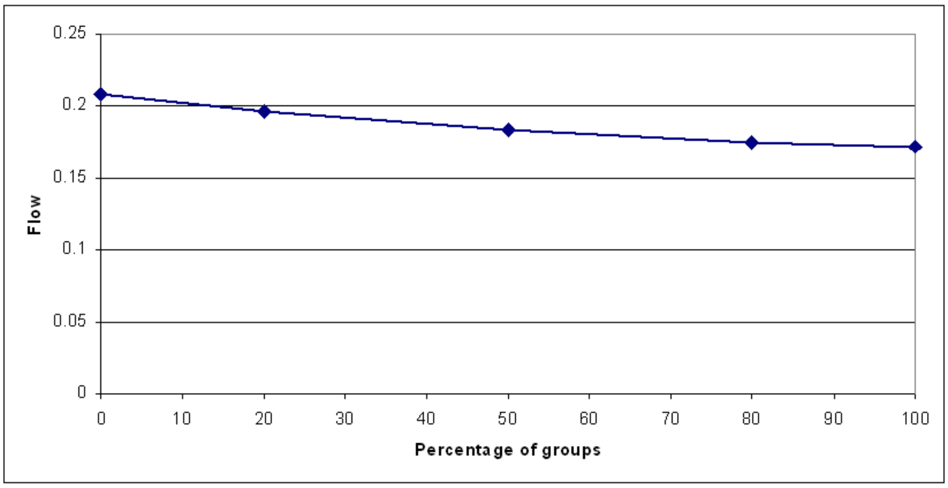


Figure 5.21: **The effect of the groups on the flow**

a different impact on pedestrian dynamics. For each culture we set each of the cultural parameters (frequencies of formations, speed, personal space and avoidance side) according to the values extracted from the real videos of the said culture as detailed in section 5.1.3.

First we examined whether there is a difference between cultures in the number of pedestrian collisions. The results presented in graph 5.22 show that France had the highest number of collisions among the pedestrians. We believe that the main cause for this is the fact that the avoidance side was more heterogeneous in France than in the other countries (45% preferred the right avoidance side and 55% preferred the left avoidance side). The lowest number of collisions was in Iraq.

We then examined whether there is a difference between cultures in reference to the number of lane changes of the pedestrians. The results in Graph 5.23 demonstrate that the lowest number of lane changes was found among the pedestrians in Iraq while the highest was in Canada. Furthermore, the pedestrians in Canada kept the greatest personal space between one another which we believe is the main reason behind this result as detailed in section 5.2.1.

In addition we examined whether there is a difference between cultures in pedestrians' speed. In Figure 5.24 we can see that the pedestrians in

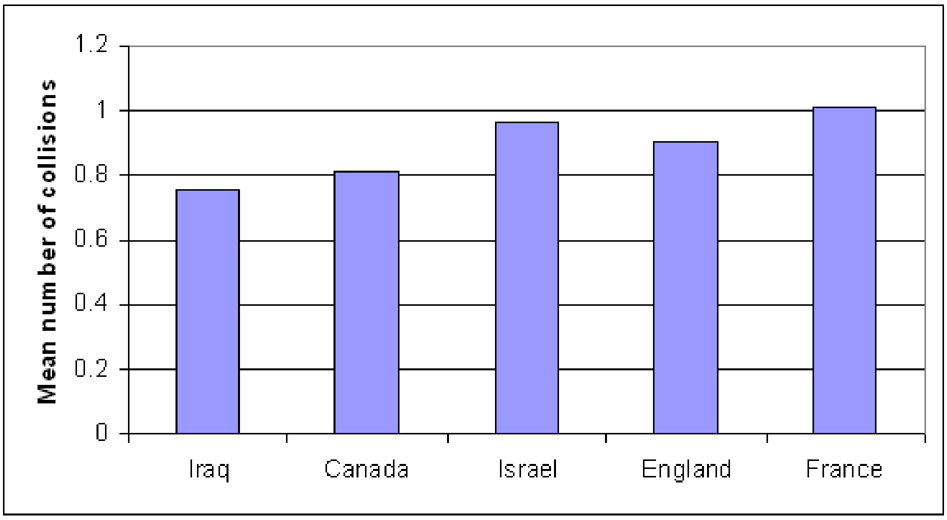


Figure 5.22: **The mean number of collisions by cultures**

Canada have the highest mean speed. The lowest mean speed was found among the pedestrians in Iraq. These results were not surprising since they were supported by the outcome of the human video analysis as detailed in Section 5.1.3. Namely, more pedestrians in Canada walk individually at a higher speed than in other cultures, while pedestrians in Iraq walk more in groups at a much slower speed than in other cultures.

Finally we examined whether there is a difference between cultures in reference to the pedestrians' flow. The results presented in Graph 5.25 indicate that the highest flow was found in Canada while the pedestrians in Iraq, Israel and France provided the lowest flow.

5.2.3 Experiment 3: Mixed Cultures

Lastly, we examined the effect of pedestrians of mixed cultures walking on the same sidewalk on pedestrian dynamics. For example, we checked the influence of a mixed population such as part of the pedestrians from Iraq and part from Canada on the pedestrian dynamics.

As it is infeasible to experiment with all the variations of cultures, we provide the examples of mixing between two cultures: Iraq and Canada. In

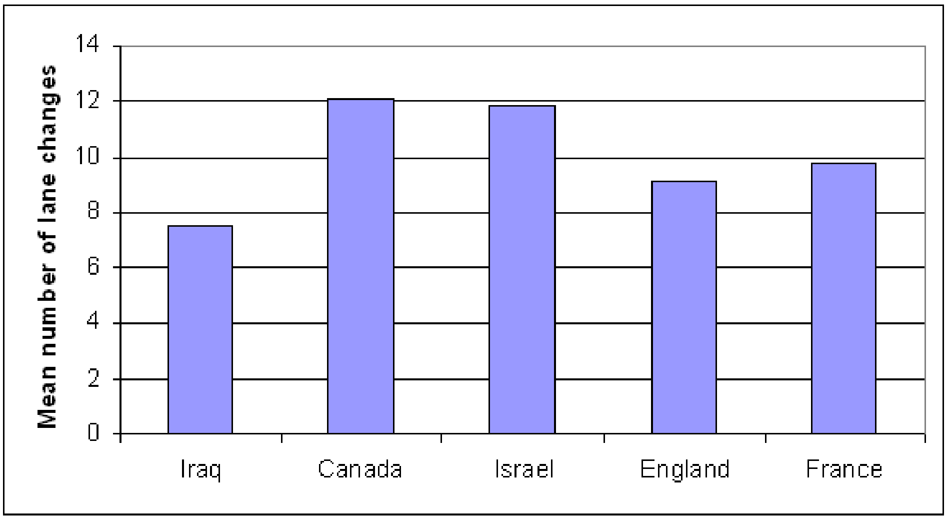


Figure 5.23: **Mean number of lane changes by culture**

this section we assert that x% of the population are from Iraq, and (100-x)% are from Canada, and we vary the values of x: 20,50 and 80. As in the previous section we set each of the cultural parameters (frequencies of formations, speed, personal space and passing side)based on the values taken from real videos of the culture as presented in section 5.1.3.

First we examined the impact of the mixed populations on the number of collisions. The results presented in graph 5.26 show that the higher the percentage of Canadians in the population the higher the number of collisions. The lowest number of collisions were found in the population comprising 20% Canadians and 80% Iraqians. Moreover, a significant difference was found between populations comprising 20% Canadians and 80% Iraqians, and populations comprising 80% Canadians and only 20% Iraqians, according to the two tailed t-test ($alpha < 0.01$).

In the next experiment we examined the impact of mixed populations (Canada and Iraq) on the number of lane changes. The results presented in Graph 5.27, as in the previous experiment demonstrate that the higher the percentage of Canadians in the population the higher the number of lane changes. Here again, the lowest number of collisions was found in population

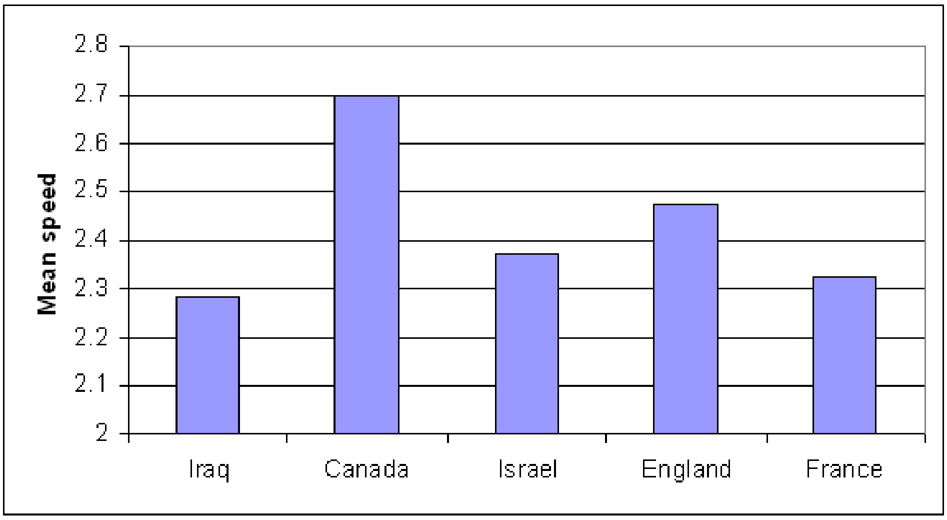


Figure 5.24: **Mean speed by culture**

where 20% were from Canada and 80% from Iraq. Moreover, a significant difference was found in the number of lane changes between populations where 20% were from Canada and 80% from Iraq, and where 80% were from Canada and only 20% were from Iraq, according to the two tailed t-test ($alpha < 0.01$).

Then we examined the impact of mixed cultures on the pedestrians' speed. Figure 5.28 illustrates that the more Canadian pedestrians in the population the higher the mean population speed. The lowest mean speed was found in the population in which 80% were from Iraq and 20% were from Canada. As in previous experiments, a significant difference was revealed in the mean speed between a population with 20% Canadians and 80% Iraqians, and a population with 80% Canadians and only 20% Iraqians, according to the two tailed t-test $alpha < 0.01$.

Furthermore, we examined the impact of mixed cultures on pedestrians' flow. The results in Graph 5.29 indicate that the highest flow was found in populations in which 80% were from Canada and only 20% from Iraq. The lowest flow was found in populations in which 80% were from Iraq and only 20% from Canada.

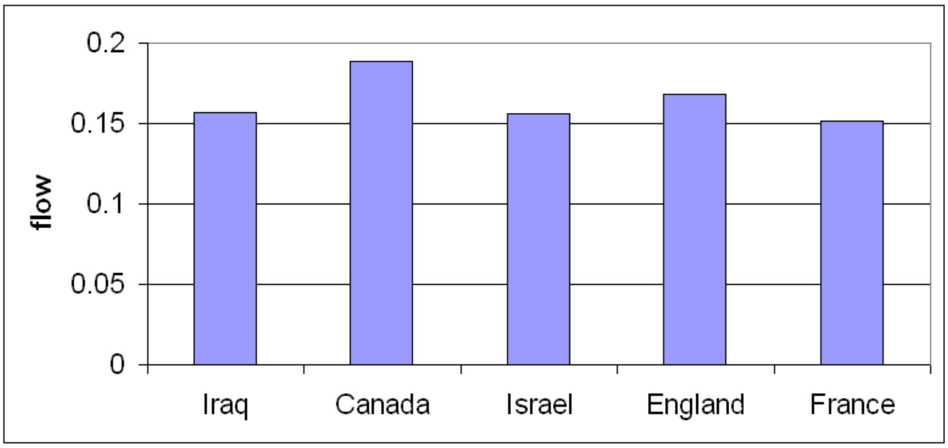


Figure 5.25: **Pedestrian flow by culture**

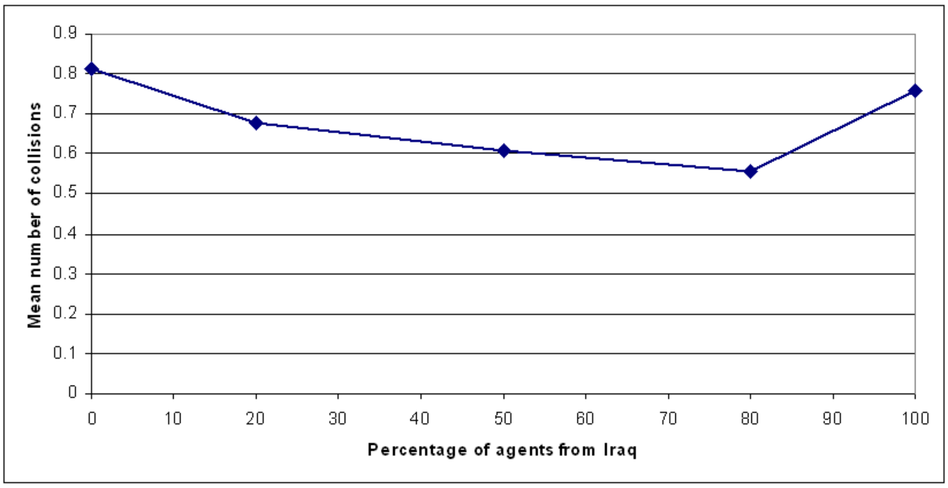


Figure 5.26: **The effect of mixed populations of Iraqians and Canadians on the mean number of collisions**

5.2.4 Experiment 4: Comparison to human data

In the previous experiments we focused on the use of simulations to investigate the effects of individual or bundled cultural parameters on overall

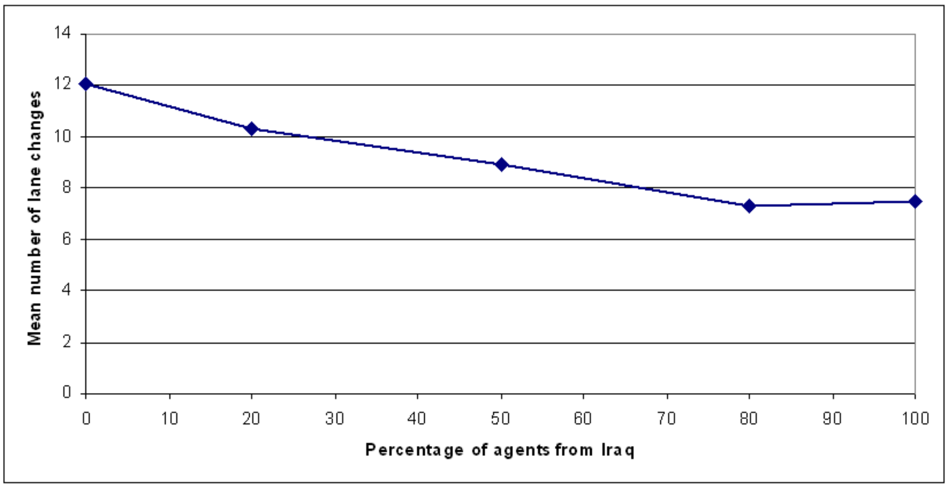


Figure 5.27: **The effect of mixed populations of Iraqians and Canadians on the mean number of lane changes**

crowd behavior. However, an important underlying question is whether the fidelity of the simulation is sufficient to support conclusions concerning human crowds.

Consequently in this section we examine whether the simulation can produce similar behavior to that of the observed human pedestrian crowd. We quantitatively compared the macro level measures (flow and mean speed) generated by the simulation to those of the crowds in the videos. We did not compare the number of collisions for this part, since humans rarely collide (not even once in the video recordings), since they employ a more sophisticated obstacle avoidance algorithm than in the simulation.

To carry out this comparison, we recreated the initial settings from four of the videos in the simulation. Specifically, we set the density of the pedestrian crowd (how many pedestrians per unit area); we set the individual parameters of agents and groups per the measured quantized values from the videos; and we ran the simulation for the same time as the videos. Note that we did not place simulated pedestrians in the initial locations of human pedestrians, so that such fine-resolution placement would not affect the macro-level of crowd dynamics. Human subjects measured the human crowd flow and mean speed by sampling pedestrians in the videos, and those sampled values were

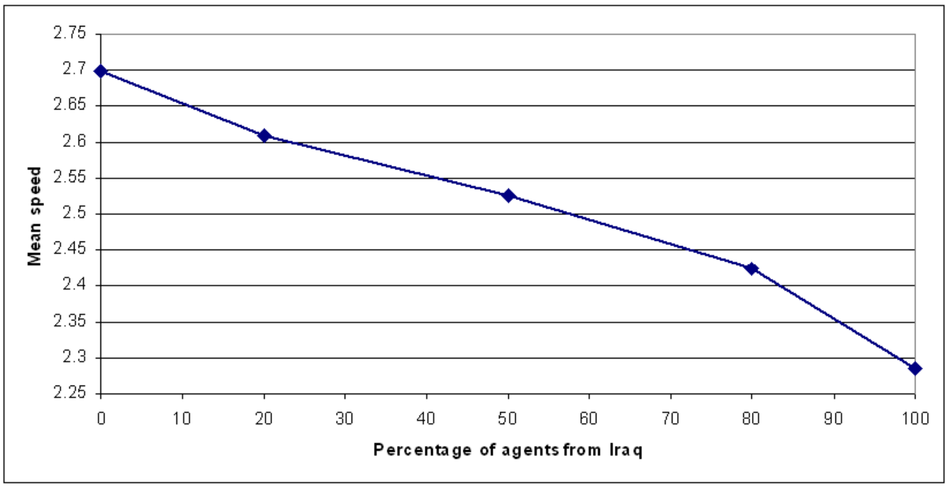


Figure 5.28: **Impact of mixed populations of Iraqians and Canadians on the mean speed**

compared to the flow and mean speed analyzed directly from the simulated trajectory data.

Flow comparison

Flow is defined as the number of persons that cross a certain line divided by the width of the line and the time the process takes. In order to extract the flow from the human pedestrian movies, we defined the sizes of the sidewalk or of the examined area. Estimating the exact sizes from the movies may constitute a great challenge, due to the position of the camera. However, there were several movies that provided very good conditions which would suffice for a good approximation of these sizes. Thus, we analyzed the flow from 4 different movies, two from France (1:40 min and 2:47 min each), one from Canada (3:36 min) and one from London (30 sec). The analysis was done only on the portions of the videos in which the deducible part was visible. It has been shown that density has a large impact on the flow [73]. Thus to quantitatively compare the simulation flow to the human pedestrian flow we had to account for the density. To extract the density from the examined human pedestrian movies, we sampled the number of pedestrians

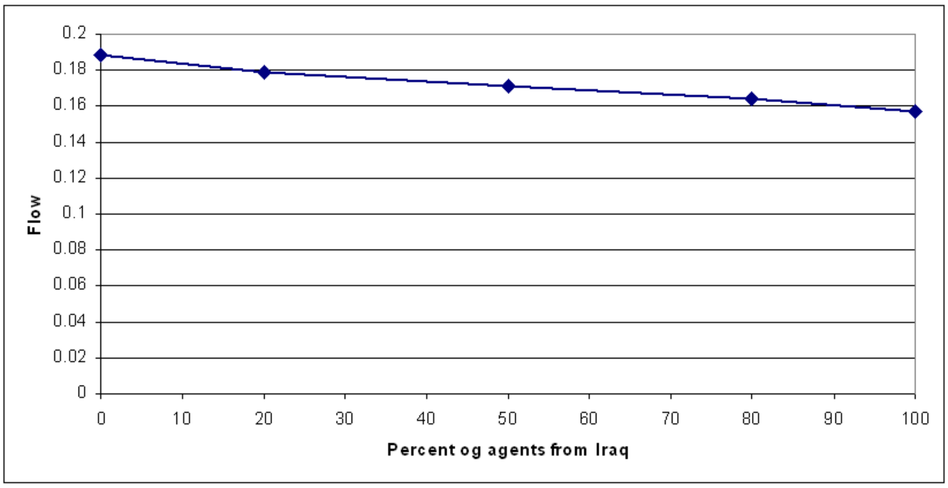


Figure 5.29: **Impact of mixed populations of Iraqians and Canadians on the pedestrian flow**

in a defined area every 5 seconds, and used the average number over all the samples.

Table 5.13 presents the densities of the examined movies. The first column presents the examined video, then we present the length and the width of the sidewalk, the resulting squared area, the average number of pedestrians in the area, and finally the density. The density was measured according to the following equation: $area/\#people$.

Movie	**Length**	**Width**	**Area**	**#People**	**Density**
France1	25m	5m	125m2	8.238	15.17
France2	16.5m	5.5m	90.75m2	5.5	16.5
Canada	9m	3.9m	35.1m2	4.428	7.92
London	12m	12m	144m2	7.4	19.4

Table 5.13: **Density analysis in human pedestrian movies**

The flow values were manually extracted from the four videos of the human pedestrians, which were analyzed, in the following manner: For each video, we counted the number of pedestrians who passed a certain line (determined as the "finish" line, i.e., one of the height borders of the sidewalk). The time variable was assigned with the number of seconds measured and the width of the sidewalk as defined in Table 5.13. The flow was calculated as follows: the number of agents that crossed a certain line divided by the width of the line and the time the process took.

To quantitatively compare the flow extracted from the human pedestrian movies to the simulation flow, we created an accurate approximation of the human pedestrians' analyzed scene. First we converted the values from the human pedestrian analysis into simulation values. We used the ratio between the width of the person in the human pedestrian scene (which was approximately 0.5 m), and the width of the agent in the simulation (which was 1). For example, in the movie from Canada the size of the measured sidewalk was 9X3.9 meters, while in our simulation it was converted to 18×7.8 based on the conversion rate.

Figure 5.30 depicts these conversions. The x-axis corresponds to the examined movie and the y-axis corresponds to the flow measurement. For each movie, we present two bars, a blue bar corresponding to the flow extracted from the human movie and a white bar corresponding to the flow received from our simulations. The results indicate a 15% error in France1, a 4% error in France2 and a 10% error in London. The maximal error we experienced was 16% which was in the movie from Canada. The mean error that we received was 11%. Note that overall a perfect match is essentially impossible due to the fact that the simulation uses a low-resolution, discrete results (e.g., only three values for speed) and mean values.

Speed comparison

In this experiment we quantitatively compared the mean speed of human pedestrians in the examined movies to the our agents' mean speed. The human pedestrians' mean speed is the mean speed values calculated from the video analysis. The simulation mean speed is the mean speed value calculated from the simulations.

Figure 5.31 presents the results. The x-axis corresponds to the examined movies and the y-axis corresponds to the mean measurement. For each movie, we present two bars, a blue bar corresponding to the mean speed that was extracted from the human movie and a white bar corresponding to the mean speed from our simulations. The results show 21% error in France1 which was the maximal error, a 16% error in France2, a 10% error in Canada and

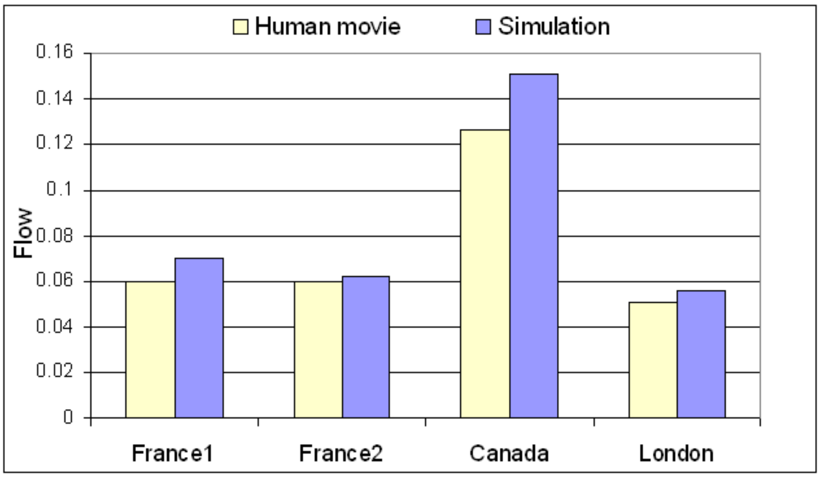


Figure 5.30: **Comparison of flow to human data**

a 6% error in London. The mean error was 13%.

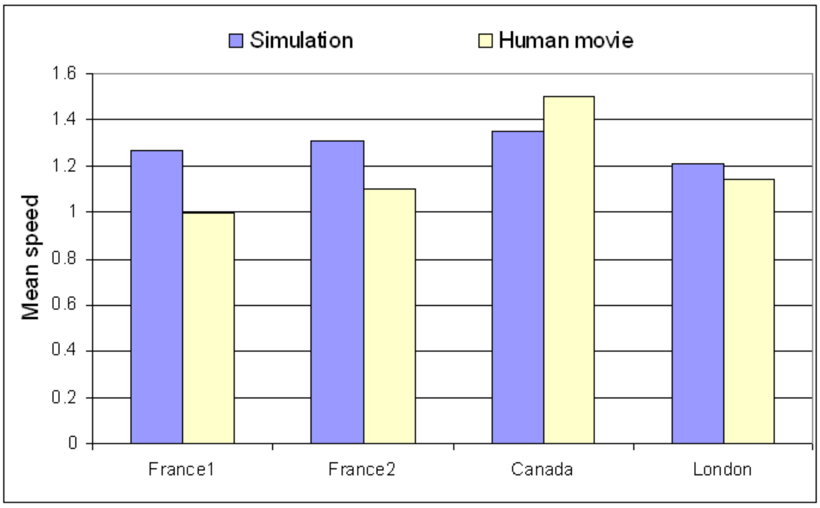


Figure 5.31: **Comparison of mean speed to human data**

Chapter 6

The Impact of Cultural Differences on Crowd Dynamics in the Evacuation Domain

In this chapter we examine the impact of cultural differences on crowd dynamics in the evacuation domain. We use an established simulation system to investigate cultural differences reported in the literature, and additionally explore the resulting macro level behavior.

6.1 Cultural Differences in the Evacuation Domain

Cultural differences also have been found in the evacuation domain. For example, it has been found that cultural differences influence the manner in which people evacuate themselves. For example, there are cultures where people tend to evacuate more in groups, and other cultures where people prefer to evacuate individually [2]. It also has been documented that some cultures perceive the event at different levels of seriousness and also with different levels of fear [2]. Moreover, it has been found that in some cultures there is more tendency to notify others about the event in comparison to other cultures [16]. Based on the literature we extracted cultural differences from evacuation scenarios such as seriousness, the tendency to notify others and group behavior. In this section we define each of these factors in detail. We use an evacuation simulation to examine the impact of these factors on the resulting macro level behavior, e.g. evacuation time, average speed.

The tendency of people to notify others about the event during an evacuation has been considered to be a cultural inclination [16]. It has been documented that different cultures tend to notify others about the existence of the event, to different extents. For example, in the America there is a higher tendency to notify others about the event than in England.

The level of seriousness participants associate to the fire alarm and also their feelings during the event, have also been found to be a cultural attribute [2]. It has been found that there is a significant difference between the Australian and Swedish populations when it comes to the emotions of fear and insecurity. People in Australia tend to take alarms more seriously and they also experience a higher level of fear and insecurity than the Swedish participants.

Another cultural difference which has also been found to influence the manner in which people evacuate is whether they evacuate individually or in groups. For example, it has been found that Swedish participants seem to prefer to evacuate in groups more than Australians who tend to evacuate individually [2].

6.2 Evaluation of the Impact of Cultural Differences on Evacuation

To examine cultural differences in evacuation behavior, we used the ESCAPES, which is a multiagent evacuation simulation, described in Section 3.4.1 and to examine the impact of these cultural differences on the resulting macro level behavior we used the following measures:

- Evacuation time: in each cycle we recorded the number of agents that are still in the terminal.
- Fear: Number of agents with HIGH fear versus LOW fear.
- Connectivity: number of connectivity components in the adjacency matrix of the agents
- Speed: the mean speed of the agents

6.2.1 Experiment 1: The impact of notifying others about the evacuation

In this section we study the impact of the tendency of agents to notify others about the event on the produced macro level of behavior. In the ESCAPES simulation, agents that are close to the event location have full knowledge regarding the event. However, agents that are at a distance from the event are unaware about what is happening. Agents aware of the event can convey the certainty of the event to other agents close to them.

In this experiment we varied the percentage of the close neighbors to whom an agent conveys its knowledge about the event and examined the impact on the evacuation time and on the agent's fear factor. Moreover, since authority figures in our simulation also notify others about the event, we examined the impact of notifying others with and without authority figures.

First we checked the impact of conveying the agents' knowledge on the evacuation time and also on their fear level, without the presence of the authority figures in the simulated environment. Figure 6.1 presents the results of the agents' evacuation time. The x-axis represents the time steps and the y-axis represents the percentage of unevacuated agents. The results clearly show that the more agents communicate the faster the evacuation time. However, no significant difference was found between agents that convey their knowledge about the event to all close neighbors (100% message conveyance) and agents that convey their knowledge to 80% of the close neighbors (80% message conveyance), according to the two-tailed t-test ($alpha = 0.26$). However, a significant difference was found between 80% of message conveyance and 50% of message conveyance (two-tailed t-test, $alpha = 0.04$). Moreover, a significant difference was also revealed between 50% of message conveyance and 20% of message conveyance (two-tailed t-test, $alpha < 0.01$). Likewise, a significant difference was also observed between agents that convey their knowledge to 20% of the close neighbors and agents that do not convey any knowledge (two-tailed t-test, $alpha < 0.01$).

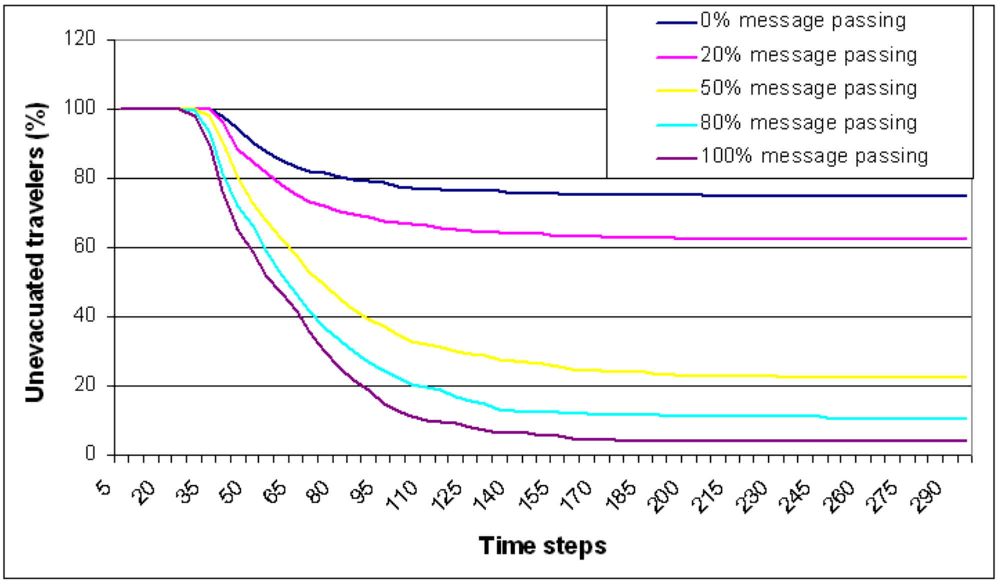


Figure 6.1: **The impact of agents' conveying their knowledge on the evacuation time (without authority figures present)**

In addition we examined the impact of agents' conveying knowledge on their fear level. Figure 6.2 presents the results. The x-axis represents the time steps and the y-axis represents the amount of unevacuated agents with a FearFactor of 2. The results show that the more the agent notifies others about the event the higher the fear level in the population.

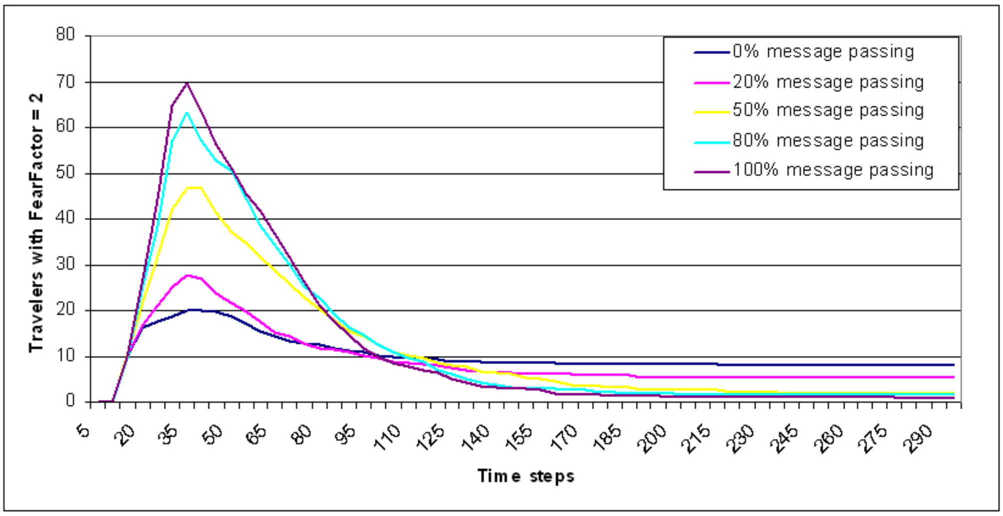


Figure 6.2: **The impact of agents' conveying their knowledge on the fear factor (without authority figures present)**

In addition we examined the impact of the agents' conveying knowledge on the evacuation behavior with the presence of 5 authority figures in the simulation environment. First we examined the impact of agents' conveying knowledge on the evacuation time. Figure 6.3 presents the results. The x-axis represents the time steps and the y-axis represents the percent of unevacuated agents. The results show that the more agents notify others regarding the event the faster the evacuation time. However, the presence of authority figures caused almost no effect on the evacuation time among fully communicative agents (i.e., 100% notify others) in comparison to Figure 6.1. For example, the mean evacuation time in a population with 5 authority figures among 100% fully communicative agents was 24.5 while the mean value among the same fully communicative agents without authorities present, as illustrated in Figure 6.1 was 23.4, which was not found to be significantly lower (according to the one tailed t-test, *alpha* = 0.42). However, among non-communicative agents (i.e., 0% notify others), the authority figures had a great impact. For example, the mean evacuation time in a population with 5 authority figures among non-communicative agents (0% notify others) was 45.05 while the mean value among the same non-communicative agents but in a population without authorities (Figure 6.1) was 80.2. The difference was found to be significantly lower (according to the one tailed t-test, $alpha < 0.01$).

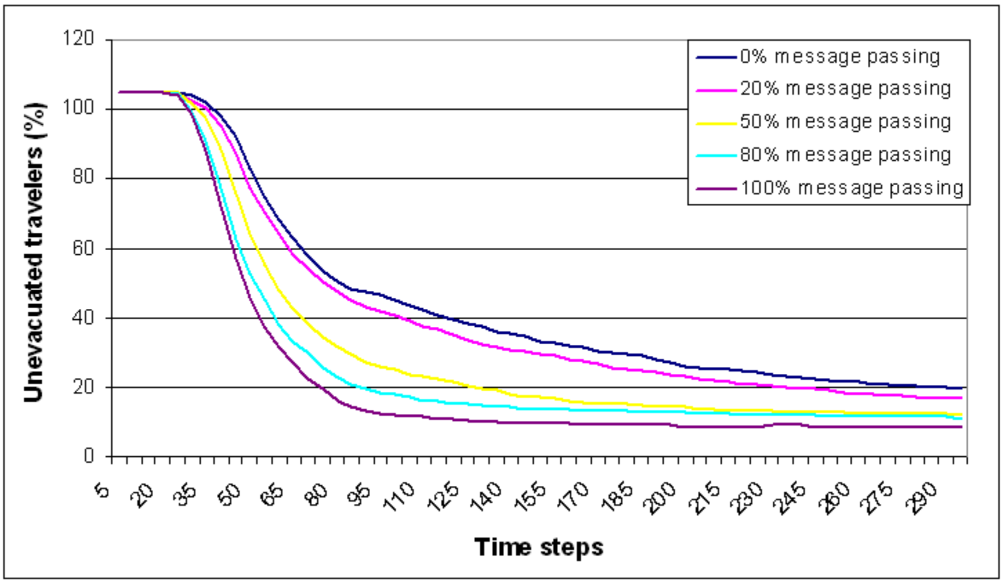


Figure 6.3: **The impact of agents conveying their knowledge on the evacuation time (with authority figures present)**

We also examined the impact of agents' conveying their knowledge on the fear factor. In this case, again, we examined the population with the presence of 5 authority figures. Figure 6.4 presents the results. The x-axis represents the time steps and the y-axis represents the amount of unevacuated agents with a FearFactor of 2. The results show that more communicative agents cause higher fear among the population. Moreover, the presence of authority figures causes lower fear among the agents in comparison to a population without authority figures (see Figure 6.2). For example the mean value of the number of agents with a FearFactor of 2 in a population with 100% fully communicative agents and 5 authority figures present was 5.2 whereas the mean value of 100% fully communicative agents without authority figures present was 12.2. This difference was found to be statistically significant (one tailed, t-test $alpha < 0.01$). In another example the mean value of the number of agents with a FearFactor = 2 in a population of non-communicative agents (0% notify others) and authority figures present was 3.3 whereas the mean value was 10.09 in the same population of agents with no authority figures present. Again this difference was found to be statistically significant(one tailed, t-test $alpha < 0.01$).

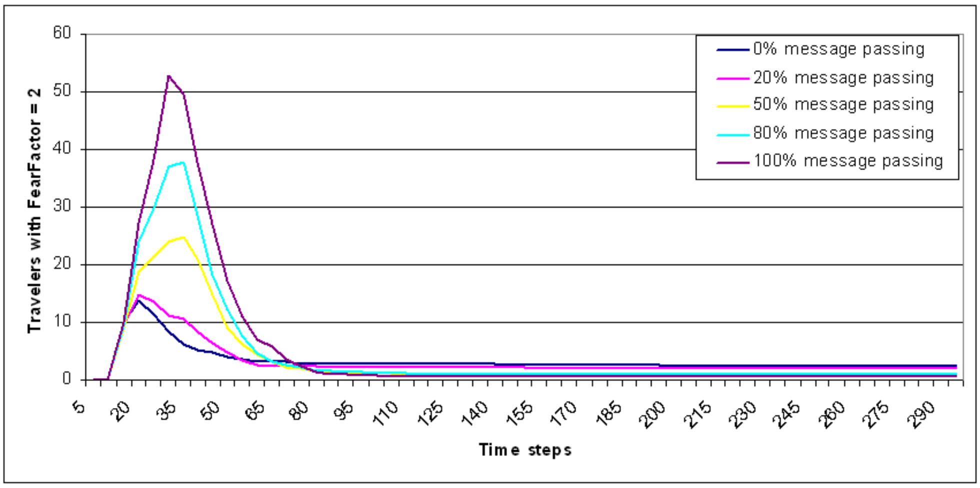


Figure 6.4: **The impact of agents conveying their knowledge on the fear factor (with authority figures present)**

6.2.2 Experiment 2: The impact of the seriousness level on the evacuation

The level of seriousness that participants associate with a fire alarm has also been found to be a cultural attribute. The level of seriousness affects the participants' level of fear during the event. In this experiment we examined the impact of an agent's seriousness on its fear and in consequence on the produced macro level of the evacuation behavior.

In our simulation each agent has an eventCertainty variable which indicates the knowledge the agent has regarding the event and also a FearFactor variable that defines an agent's level of fear. In the ESCAPES simulation the eventCertainty variable has a direct effect on the agent's FearFactor . Consequently, if an agent has a HIGH eventCertainty then its FearFactor will also be HIGH. To account for the cultural difference in the levels of seriousness that people have during an evacuation, we modified our simulation as follows: we defined a level of seriousness that an agent has (not serious, semi-serious, very serious). When an agent finds out about an event (eventCertainty = HIGH) its FearFactor will be affected by its seriousness level. Serious agents will have a high level of fear (FearFactor = HIGH), semi-serious agents will have a low level of fear (FearFactor = LOW), and non-serious agents will have no fear (FearFactor = NONE).

In this experiment we varied the percent of serious agents versus semi-serious agents and examined the impact on the evacuation time, the agents' fear factor and their mean speed (since an agent's speed is affected by its FearFactor, the more the agent is afraid the faster it will proceed to the exit). In our simulations the authority figures had a calming effect; though they notified others regarding the event they still reduced the fear level of the agents As a result, we examined the impact of seriousness with and without authority figures present.

First we examined the impact of agents' seriousness on the evacuation time, mean speed and also on their fear level, with no presence of the authority figures in the simulated environment. Figure 6.5 presents the results of the agents' evacuation time. The x-axis represents the time steps and the y-axis represents the percent of unevacuated agents. The results show that the time of evacuation of more serious agents is less than evacuation time of less serious agents. A significant difference was found between the population of all serious agents (100% seriousness) and the population of no serious agents (0% seriousness), according to the two-tailed t-test ($alpha = 0.004$).

The same pattern was also observed in reference to the evacuators' speed. Figure 6.6 shows that the more serious the agents the faster the speed. Furthermore, the difference between the population of 20% of serious agents (20% seriousness) and the population of non-serious agents (0% seriousness) was found to be statistically significant according to the two-tailed t-test ($alpha < 0.01$).

Moreover, we examined the impact of the agents' seriousness on their fear level, with no authority figures present in the simulated environment. Figure 6.7 presents the results. The x-axis represents the time steps and the y-axis represents the amount of unevacuated agents with a FearFactor of 2. The

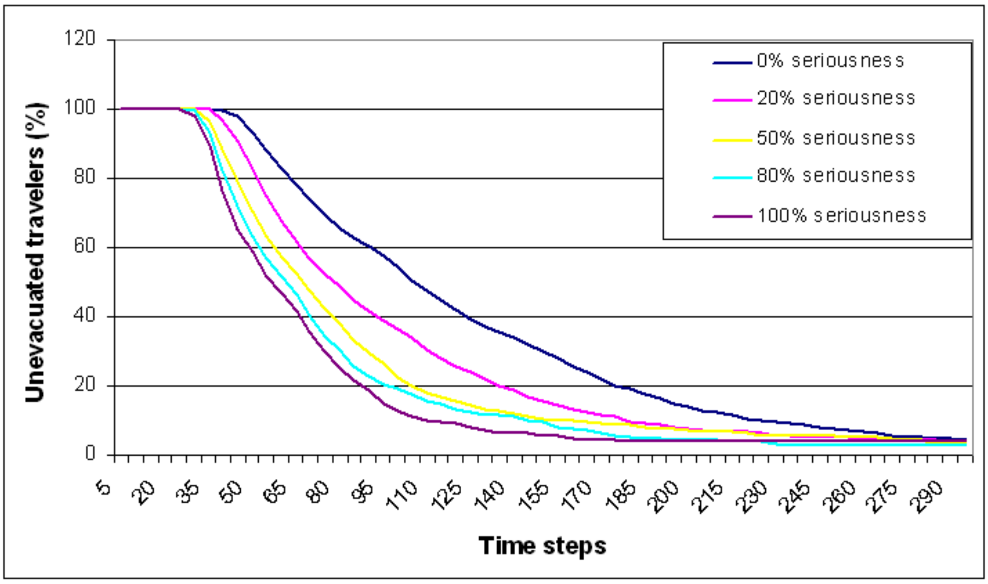


Figure 6.5: **The impact of the agents' seriousness on the evacuation time (without authority figures present)**

results show that the higher the seriousness the higher the fear.

We then examined the impact of the agents' seriousness on evacuation behavior with the presence of 5 authority figures in the simulation environment. First we examined the impact of agents' seriousness on the evacuation time. Figure 6.5 presents the results. The x-axis represents the time steps and the y-axis represents the percent of unevacuated agents. The results show that the presence of authority figures causes almost no change in the evacuation time of serious and less serious agents. Moreover, no significant difference was found between the population of all serious agents (100% seriousness) and population of no serious agents (0% seriousness), according to the two-tailed t-test ($alpha = 0.39$). Moreover, the authority figures had almost no effect on the evacuation time among serious agents (100% seriousness) in comparison to the evacuation time of the serious agents in the environment without authority figures present as depicted in Figure 6.5. For example, the mean evacuation time in populations of 100% serious agents with 5 authority figures was 24.5 whereas the mean value of the same serious agents but in a

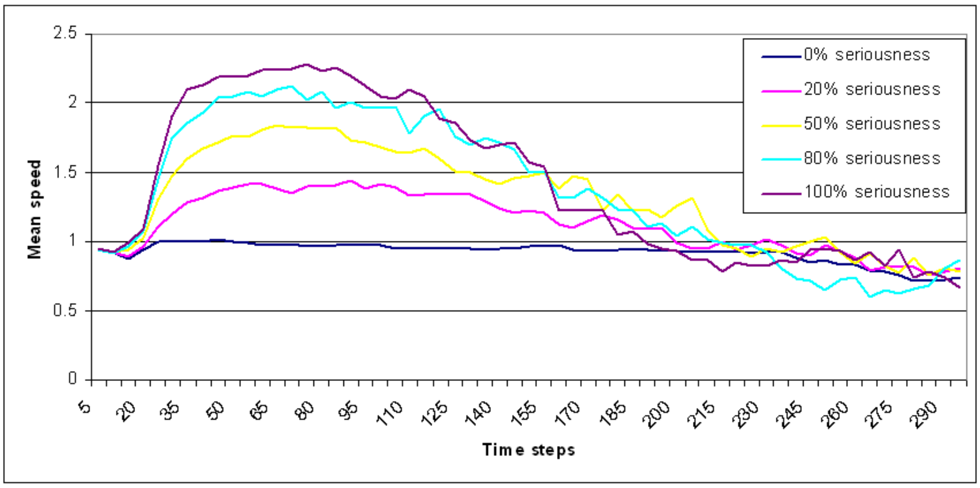


Figure 6.6: **The impact of agents' seriousness on the mean speed (without authority figures present)**

population without authorities (Figure 6.5) was 23.4. The difference was not found to be significantly lower (according to one tailed t-test, *alpha* = 0.42). In contrast, the authority figures had a great impact on the population of non-serious agents (0% seriousness). For example, the mean evacuation time in the population of non-serious agents (0% seriousness) with 5 authority figures was 29.6, whereas the mean-evacuation time of the same population without authority figures present (Figure 6.5) was 41.1. This difference was found to be statistically significant (according to the one tailed t-test, *alpha* = 0.03).

We also examined the impact of agents' seriousness on the evacuators' mean speed. The results displayed in Figure 6.9 show that the presence of authority figures cause agents to evacuate within much less time at high speeds (speed level > 2) in comparison to the agents who evacuate without authority figures present as depicted in Graph 6.6. The results also show that authority figures have almost no impact on the population comprising 100% seriousness. For example the mean speed of 100% serious agents with 5 authority figures present was 1.43 compared to a mean speed of 1.42 without authority figures present. The main reason for such slight influence of the authority figures is that on the one hand they have a calming effect which

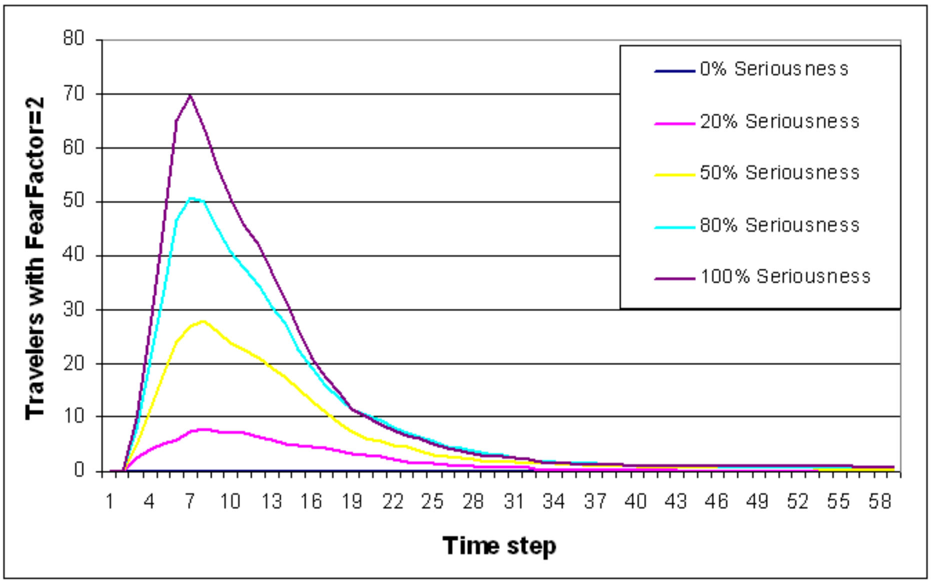


Figure 6.7: **The impact of agents' seriousness on the fear factor (without authority figures present)**

causes agents to reduce their speed and on the other hand they notify other agents regarding the event which in contrast causes them to increase their speed. Nonetheless, authority figures have an impact on the population of non serious agents (0% seriousness). For example the mean speed of 0% serious agents with 5 authority figures present was 1.3 while the mean speed of 0% serious agents without authority figures present was 0.9. This difference was found to be statistically significant (according to the one tailed t-test, $alpha < 0.01$).

Next we examined the impact of the agents' seriousness on the fear factor. Here again we examined the population with the presence of 5 authority figures. Figure 6.10 presents the results. The x-axis represents the time steps and the y-axis represents the amount of unevacuated agents with a FearFactor of 2. The results show that more serious agents have greater fear. However, the presence of authority figures lowered the fear among the agents in comparison to a population without authority figures present (Figure 6.7). For example the mean value of the number of agents with a FearFactor of

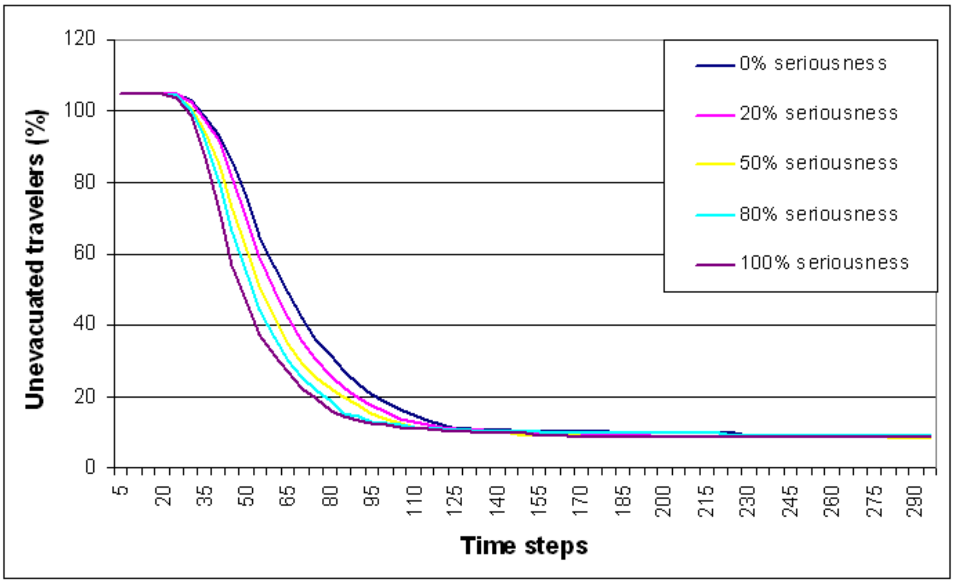


Figure 6.8: **The impact of agents' seriousness on the evacuation time (with authority figures present)**

2 in a population with 100% serious agents and 5 authority figures was 5.2 while the mean value of this population without the authority figures present was 12.2.

6.2.3 Experiment 3: The impact of group behavior on evacuation

In our final experiment we discuss the impact of how people evacuate on the evacuation behavior produced. As mentioned it has been shown that in some cultures people tend to evacuate in groups while in others they tend to evacuate individually.

In the ESCAPES evacuation simulation it has been shown that the use of SCT increases grouping behavior, as described in section 3.4.1. In this section we investigate the impact of this grouping behavior on the on evacuation time, the agents' fear factor and the agents' mean speed. Moreover, we examine the impact of the authority figures on the produced behavior.

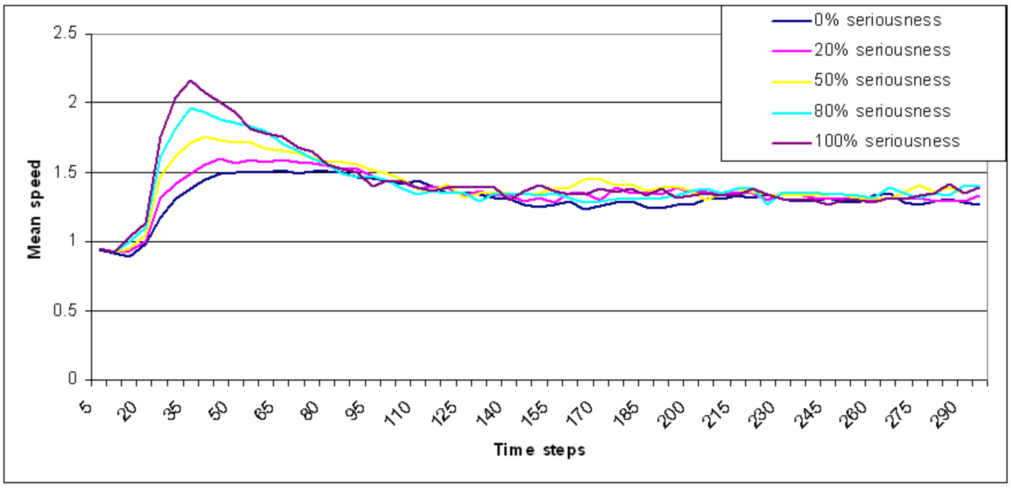


Figure 6.9: **The impact of the agents' seriousness on the mean speed (with authority figures present)**

First we examined the population without authority figures. To examine the impact of agents' grouping behavior on the evacuation we compared agents with the SCT process and agents without the SCT process and we measured the agents' evacuation time, their fear factor and their mean speed. Figure 6.11, which presents the impact of SCT on the evacuation time, shows that the evacuation time with SCT seems to be slightly longer than without SCT. Nonetheless this difference was not found to be statistically significant according to the one tailed t-test ($alpha = 0.3$).

Figure 6.12 illustrates the results of the impact on the agents' fear factor. The x-axis represents the time steps and the y-axis represents the number of agents with a FearFactor of 2. The results show that there was no significant difference in the fear factor between agents with the SCT process and those without, according to the two tailed t-test ($alpha = 0.9$).

In addition, we examined the impact of the SCT process on the agents' mean speed. Figure 6.13 presents the results of agents' mean speed. The x-axis represents the time steps and the y-axis represents the mean speed. As in the previous results, no significant difference was found in the mean speed between agents with the SCT process and those without the SCT process, according to the two tailed t-test ($alpha = 0.9$).

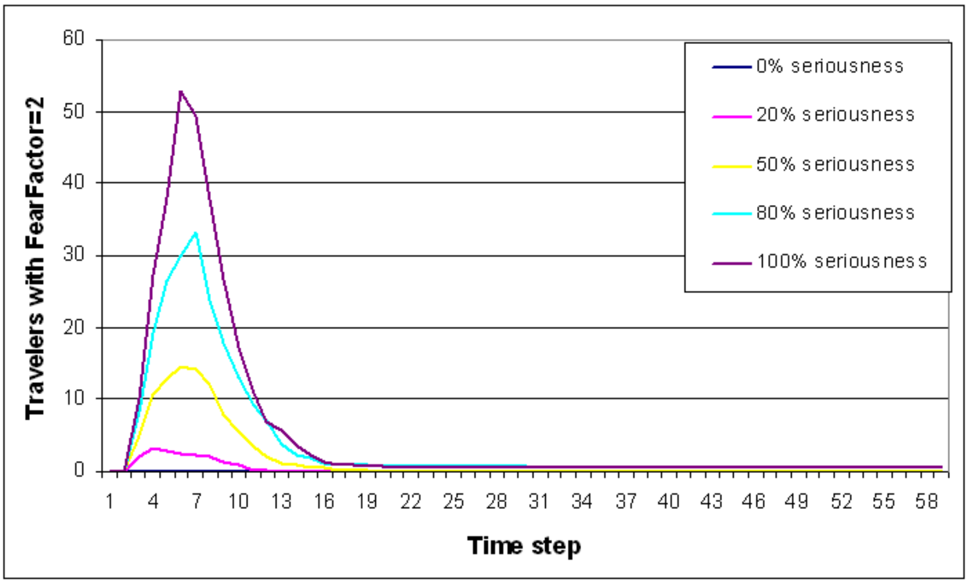


Figure 6.10: **The impact of the agents' seriousness on the fear factor (with authority figures present)**

We then examined the impact of the SCT process on the population with authority figures present. As in the previous experiments we defined 5 authority figures and measured the agents' evacuation time, fear factor and mean speed. Figure 6.14 shows the results of the evacuation time. The x-axis corresponds to the time steps and the y-axis represents the number of unevacuated agents. The results show that no significant difference was found in evacuation time between agents with the SCT process and those without, according to the two tailed t-test ($alpha = 0.48$).

Similar results were also found in reference to the agents' fear and mean speed. The results regarding the agents' fear factor presented in Figure 6.15 show that no significant difference was found in the fear factor of agents with the SCT process and those without, according to the two tailed t-test ($alpha = 0.88$). The results regarding the agents' mean speed, which appear in Figure 6.16 again reveal that no significant difference was found between the agents' mean speed of those with the SCT process and those without, according to the two tailed t-test ($alpha = 0.65$).

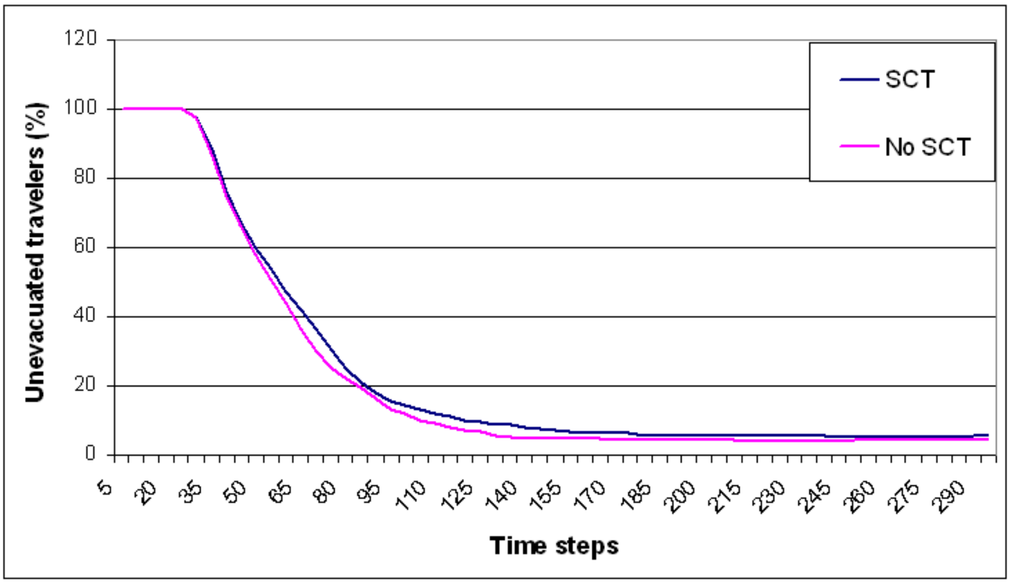


Figure 6.11: **The effect of SCT on the evacuation time (without authority figures present)**

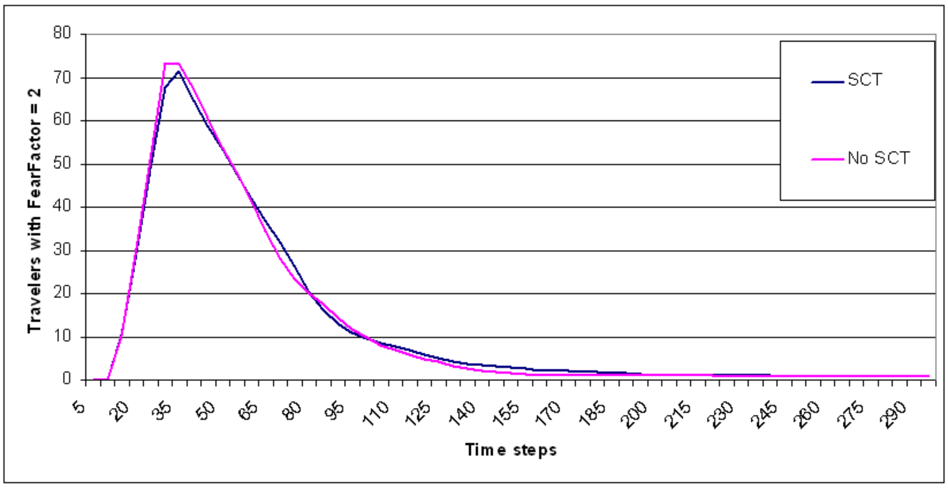


Figure 6.12: **The effect of SCT on the fear factor (without authority figures present)**

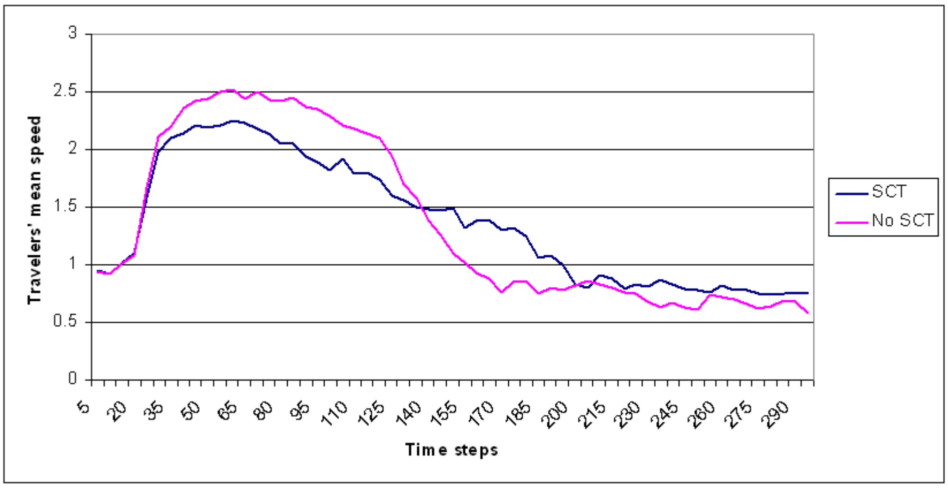


Figure 6.13: **The effect of SCT on the mean speed (without authority figures present)**

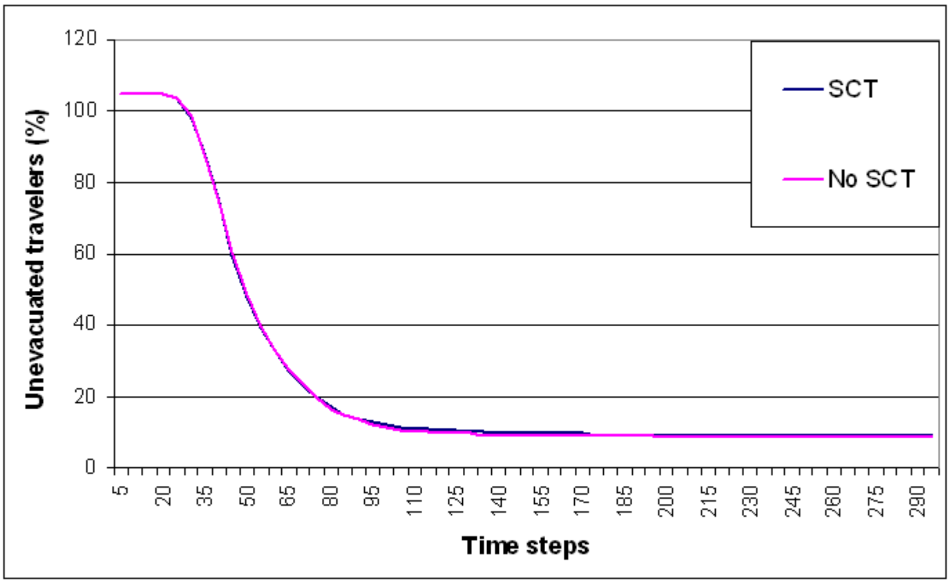


Figure 6.14: **The effect of SCT on the evacuation time (with 5 authority figures present)**

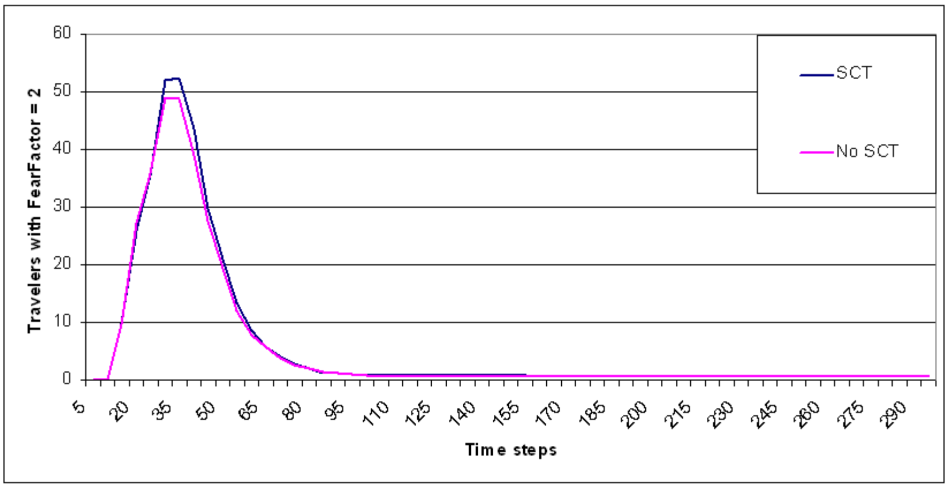


Figure 6.15: **The effect of SCT on the fear factor (with 5 authority figures present)**

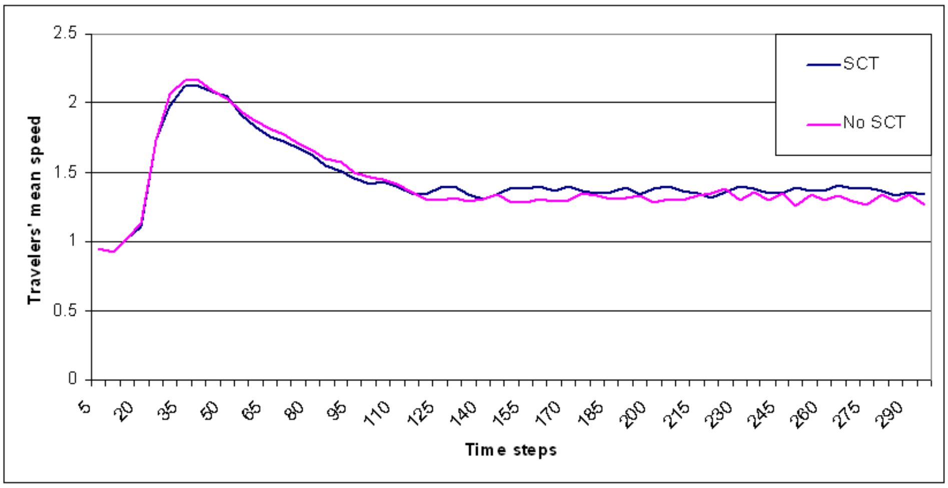


Figure 6.16: **The effect of SCT on the mean speed (with 5 authority figures present)**

Part II

Modeling Crowd Behavior using a Qualitative Reasoning Approach

In this part of the thesis we discuss a macro level approach for modeling crowd behavior in demonstrations in particular. The ability to model and reason about the potential violence level of a demonstration is extremely important to the police in their decision making process. Unfortunately, existing knowledge regarding demonstrations is composed of partial qualitative descriptions without complete and precise numerical information. In this part we describe a pioneer attempt to use qualitative reasoning techniques to model demonstrations. To the best of our knowledge, such techniques have never been applied to modeling and reasoning about crowd behaviors, nor in demonstrations in particular. We developed qualitative models consistent with the partial, qualitative social science literature. This enabled us to model the interactions between different factors that influence violence in demonstrations. We then utilized the qualitative simulation to predict the potential eruption of violence at various levels, based on a description of the demographics, environmental settings, and police responses. We incrementally present and compare three of these qualitative models. The results show that while two of the models fail to predict the outcomes of real-world events reported and analyzed in the literature, one model provides good results. We also examined whether a popular machine learning algorithm (decision tree learning) can be used. While the results show that the decision trees provide improved predictions, we demonstrate that the QR models can be more sensitive to changes, and can account for *what if* scenarios, in contrast to decision trees. Moreover, we introduce a novel analysis algorithm which analyzes the QR simulations, to automatically determine the factors that are most important in influencing the outcome in specific real-world demonstrations. We show that the algorithm identifies factors that correspond to experts analysis of these events.

Chapter 7

Qualitative Reasoning and Simulation

In this chapter we provide a brief description of the Qualitative Reasoning (QR) approach. We present our QR models of demonstration and the evaluation methods. In chapter 8 we present the results of our evaluation.

7.1 What is Qualitative Reasoning?

A comprehensive discussion on QR is beyond the scope of this work. Below we provide a brief description and refer the reader to [22, 47] for additional details.

Qualitative simulation enables reasoning about possible system behaviors that can emerge from an initial world state. A qualitative model is like any other model, in the sense that it has variables, which hold values, and there are rules or relations between the variables that govern how a given value of a variable influences another variable. Once an initial state (value assignment) is give, the model can be used to determine all possible values that can result from valid inference (i.e., given the known relations between variables).

The key idea in qualitative reasoning and simulation, however, is that the values of variables and the rules that govern their evolution are given in descriptive, qualitative terms, rather than numerical. A qualitative model is made from a set of *quantities* which are variables. Quantities are the lowest resolution representation for continuous parameters and are each composed of a pair of values: *magnitude* and *derivative*. The magnitude represents the amount of quantity and the derivative represents the direction of change. The set of possible values is described by Quantity Space (QS) which is a finite and ordered set of qualitative landmark values. Landmark values represent important real-number values.

For example, suppose we would like to represent the amount of water in a pipe. The *magnitude* represents the amount of water and we may define its quantity set (QS) with the following landmarks: $QS = zero, plus, max$, which represent *no water in the pipe*, *some water in the pipe*, or *full capacity of water in the pipe*. The *derivative* represents the direction of change and it has the following quantity set: $QS = min, zero, plus$ which represents three possible directions accordingly: decreasing, steady and increasing. If there is a state with the following quantity $< plus, plus >$ then in the current state, there is some amount of water in the pipe which is increasing.

In addition to quantities, a qualitative model is created from a set of causal relationships, which relate quantities to each other, given both their magnitude as well as their derivative. There are two types of casual relationship between quantities, *direct* and *indirect influence*. The direct influence, denoted I signifies that A causes B; if B occurs then A might be the reason for it occuring. The indirect influence, denoted P signifies that A causes B only if A causes C and C causes B. A quantity that is not influenced by any process is considered to be a constant. Each influence may be positive $(I+, P+)$ or negative $(I-, P-)$ meaning the derivative of the target quantity increases or decreases accordingly.

A qualitative simulation takes as input the initial state of the world (i.e., a value assignment for the quantities) and qualitative descriptions of processes that change the state. It produces a state transition graph. Each state is a possible unique behavior that the model develops and it contains a unique set of values and inequality statements (quantities) which describe the current behavior of the system.

To construct the state graph, qualitative simulation uses the partial models to drive changes in the quantity magnitudes and derivatives, until they no longer change. Each path of transitions, representing the application of the models to the quantities, represents a single possibility of how the state may develop from the initial state until it converges to a final state. The collection of all of these paths, leading from the initial state to the terminal state is the entire set of possible system behaviors.

Thus a state graph captures the set of all possible behaviors that the model may manifest. It consists of a set of states and the transitions between them (state-transitions). State transitions transform one state into another, by specifying the changes in values and in inequality statements. Each state may be the origin of multiple transitions which lead to multiple possible developments of the state. A sequence of states connected by state transitions where each state is the immediate successor of the one before, is called a behavior path.

In each cycle and for each quantity, all influences (direct and indirect) are combined. When positive and negative influences are combined ambiguities may occur. The qualitative simulation considers all the possible combinations thus, when qualitative description is incomplete, it provides a non-deterministic prediction.

QR is a well-established technique in artificial intelligence and computer science. There are free, user-friendly software packages that exist for specifying qualitative models, and for reasoning with them. Indeed, for most of the experiments we discuss in this work, we utilized a free software package that was available called GARP3 [14]. This software package allows visual entry and running of models.

7.2 Proposed models for Qualitative Simulation of Demonstrations

There is significant literature in social science on the factors that impact violence during demonstrations. There are many macro-level qualitative descriptions of the influencing factors on the violence level. Each theory focuses on a small sub-set of factors. Integrating all of them into a single unified model is a real challenge. Thus, in order to enable modeling and reasoning where there is imprecise and partial information requires qualitative reasoning techniques. In this chapter we show that QR modeling techniques are appropriate for the current state of knowledge in social sciences regarding demonstrations.

The Israeli police initiated a comprehensive study to address this challenge, resulting in a report [18] that provides a collection of factors and their influence on the violence level and also on each other. Their goal was to classify and analyze different kinds of demonstrations in order to propose appropriate methods to the police force for dealing with masses. They studied 102 crowd events (in particular demonstrations) during the years 2000–2003 and 87 interviews with policemen and police officers. They analyzed a variety of factors that may affect violent behavior, as well as relevant literature. Their report is thus a comprehensive, yet informal and imprecise, collection of factors which affect demonstrations.

The first contribution we make is in translating the textual results in the report into an executable model, using QR. Indeed, we developed three separate models, incrementally increasing in complexity and size, of the different components influencing violence in demonstrations. We describe these models below.

7.2.1 The `Base` Model

The first (*Base*) model was based on the literature review presented in the Israeli Police report [18] (see Figure 7.2). This was the first attempt at building a baseline model purely founded on literature review. According to the Base model the factors most influential on the violence level during demonstration are (1) the crowd's a-priori hostility towards the police; (2) willingness to pay a personal price (such as willingness to be arrested); (3) the chance of punishment for violent actions (e.g., the belief that the police will or will not respond strongly); (4) group cohesiveness; (5) previous history of violence. All of these directly increase the level of violence. However, not all have an opposite effect when reversed. For instance, the existence of previous history of violence among the specific group of demonstrators increases the potential violence level, but lack of such history does not decrease the violence level (i.e., it has no effect).

Figure 7.1 presents a description of the model and Figure 7.2 presents the graphical structure of the model. We defined one entity (termed population) comprising six quantities, five of which were based on the presented theoretical description and one on the outcome violence level. For each quantity we defined a quantity space (QS), direct influence (I) between them and the violence level. $I+$ represents the derivative of the target quantity (violence level) that increases and $I-$ means it decreases. For example, high hostility towards the police increases the violence level while low hostility towards the police decreases the level. There are also quantities with unidirectional effects such as previous history where existence of previous history of violence will increase the violence level while lack of history will not decrease the violence level; lack of such history will actually have no effect on the level.

7.2.2 The `Police` Model

The *Police* model, described by Karmeli and Ravid-Yamin [18] (Figure 7.4), significantly expanded the Base model, based on interviews with police officers and their investigation into 102 demonstrations. In addition to the factors from the Base model, the Police model comprises additional variables which can be divided into several groups: *Environment variables* (such as time, weather and space characteristics of the location of the event), *Participant variables* (sociological and behavior characteristics of the participants such as the number of participants, the existence of speakers or leaders, etc.), *Police variables* (referring to behavior and organizational characteristics of the police such as intervention time and intervention strength), *Procedural variables* (refer to the dynamic characteristic of the event) and *Outcome variable* (refer to the outcome of the event).

The variables added to the Base model which were also found to be influential on the demonstration outcome include: (1) number of participants, (2)

Participants:
    Population
Quantities:
    Hostility for police: QS = {high, low, zero}
    Personal price: QS = {high, low, zero}
    Punishment: QS = {high, low, zero}
    Group cohesiveness: QS = {high, low, zero}
    History of violence: QS = {yes, no}
    Violence: QS = {high, low, zero}
Influences:
    I+(Hostility for police{high} ,Violence)
    I-(Hostility for police{low} ,Violence)
    I+(Personal price{high} , Violence)
    I-(Personal price {low} , Violence)
    I+(Punishment{high} , Violence)
    I-(Punishment{low} , Violence)
    I+(Group cohesiveness {high} , Violence)
    I-(Group cohesiveness {low} , Violence)
    I+(History of violence {yes} , Violence)

Figure 7.1: **Description of the Base Model**

the existence of a group representative (such as a group leader or speaker), (3) the existence of a demonstration license (permit; i.e., whether the demonstration was legal), (4) existence of a violent core among the participants, (5) the participants' united identity (such as racial minority, social groups etc.), (6) event purpose (such as emotional event or rational event), (7) police time intervention, (8) police intervention strength, (9) weather, (10) time of the day (such as morning, night etc.), (11) demonstration location sensitivity (a highly sensitive place, such as a mosque or synagogue, or a place with low sensitivity such as a city square or piazza) and (12) time of year sensitivity (e.g., Christmas). The research results showed significant relations between these variables and their impact on the event outcome (the violence level). For example, political or social demonstrations usually end with a low level of

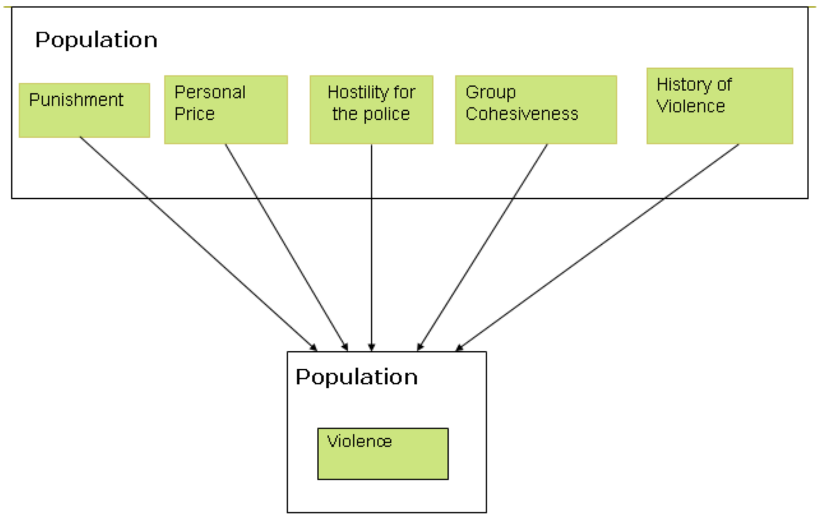


Figure 7.2: **Base Model Structure**

violence (typically without casualties). However, demonstrations of nationalistic nature intended to express emotions (letting off steam) are characterized by much more violent outcomes. It has also been found that the time of the day has an impact on the violence level, where violent demonstrations occur more at nighttime than daytime.

Figure 7.3 presents a description of the Israeli-police model and Figure 7.4 depicts its graphical structure. We defined three entities (Population, Nature and Police) and 18 quantities, where 6 are similar to the Base model and 12 are additional ones. Moreover, based on the research conclusions we defined influence (I) among the different variables. As in the previous model, $I+$ represents the derivative of the target quantity (such as violence level) increases and $I-$ means it decreases. For example, an emotional event increases the existence of a violent core among the participants.

7.2.3 The `Bar-Ilan University` Model

The third model, *BIU* (Bar Ilan University), shown in Figure 7.6, is our novel extension of the Police model. Based on interviews with social and cognitive scientists, as well as additional literature surveys [53, 70], we added four additional variables, and updated 19 influences (relations) among the

Participants:
Population
Nature
Police
Quantities:
Population:
…*Similar to Base model*…
License: QS = {have, don't have}
Group speaker: QS = {exist, not exi
Number of participants: QS = {0-10
United identity: QS = {racial minor
Violent core: QS = {exist, not exist
Event purpose: QS = {emotional, ra
Nature:
Weather: QS = {extreme, medium}
Time: QS = {morning, noon, evenii
Place sensitivity: QS = {high, low}
Time sensitivity: QS = {high, medi
Police:
Time intervention: QS = {prematur
Intervention strength: QS = {low, e:
Influences:
…*Similar to Base model*…
I+(License{don't have},Violence)
I-(License{have},Violence)
I-(Group speaker{exist}, Violence)
I+(Number participants{100-1000/1000-10
I-(Number participants {0-100/10000<}, V
I+(United Identity {racial minority/ religiou
I-(United Identity {racial minority/ religious
I-(United Identity {social&political gr. }, Ev
I+(United Identity {social&political gr.}, Li
I+(Violent core{exist}, Violence)
I-(Violent core{not exist}, Violence)
I+(Event type{emotional}, Violent core)
I-(Event type{rational}, Violent core)
I+(Weather{medium},Violence)
I-(Weather{extreme},Violence)
I+(Time{evening/night},Violence)
I-(Time{morning/noon},Violence)
I+(Place sensitivity{high},Event purpose)
I-(Place sensitivity{low},Event purpose)
I+(Time sensitivity{high/medium},Violenc
I-(Time sensitivity{low},Violence)
I+(Time intervention{premature/late },Viol
I-(Time intervention {on-time}, Violence)
I+(Intervention strength {low/high},Violen
I-(Intervention strength {exact},Violence)

Figure 7.3: **Description of the Israeli-Police Model**

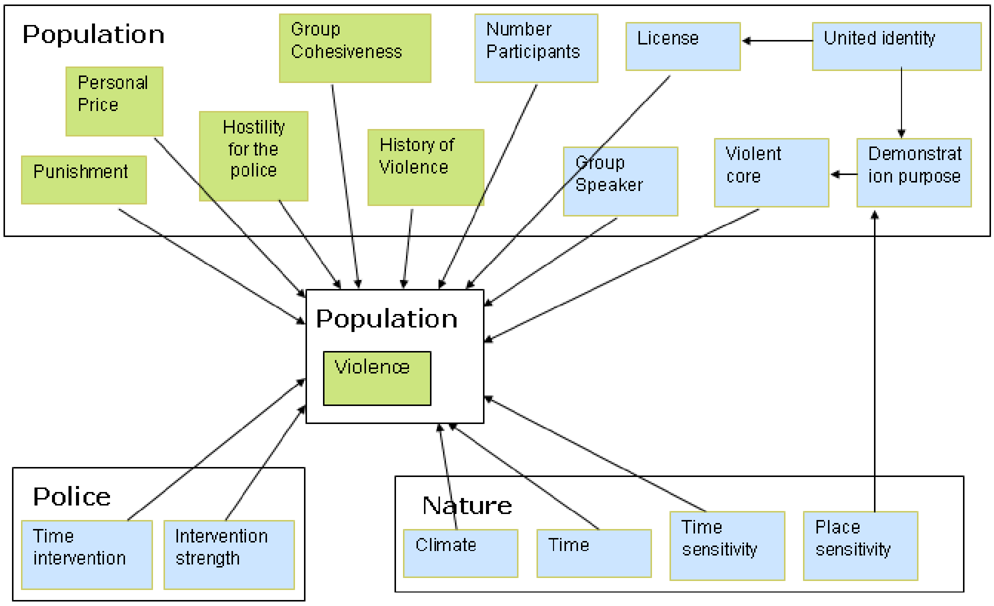


Figure 7.4: **Police Model Structure.**

variables. The factors we added include: (1) event order (indicating the amount of preparation invested in the event, such as delineation, disposition of the police forces etc.) (2) participants anonymity (indicating whether the participants believe that they can be recognized and identified), (3) participants' visual cohesiveness (such as similar outfits like football fans apparel) and (4) lighting.

Figure 7.5 presents a description of the model and Figure 7.6 depicts its graphical structure. We used the same entities (Population, Nature and Police) as in the Israeli-Police model and added four additional quantities to the existing ones. Furthermore we updated the quantity space of the police intervention strength and the influences (I) among the variables as described in detail in the model description in Figure 7.5. As in previous models $I+$ represents the derivative of the target quantity that increases and $I-$ means it decreases.

We provide several examples of the updated influences. For instance, we updated the effect of police intervention strength. Thus, instead of direct impact on the violence level as in the Police model, it impacts the participants' belief that they may be punished, and their hostility toward the police. In the BIU model, high intervention strength increases the participants' hostility toward the police and increases the participants' chance of punishment. However, low intervention strength decreases only the participants' chance for punishment without affecting their hostility towards the police factor. Another example of updates that we introduced into our model relate to the existence of a group speaker and the request (and approval) of a permit to demonstrate (demonstration license) which increase the maintenance of order, and in turn decrease the violence level. In contrast, in the Police model, the license and group speaker variables only had a direct influence on the violence level. Moreover, in the BIU model the variable of the *number participants* no longer directly influences the violence level as in the Police model. Instead this factor affects the participants' anonymity level ("the more participants around me the less I am recognizable"). Another variable added to the BIU model is the participants' visual cohesiveness which has an impact on group cohesiveness, and actually increases the sense of belonging to the same group.

7.3 Prediction and Analysis

For different demonstration cases, one can set the initial state quantities to their qualitative values, based on the demographics and environment values known at the time. Then the qualitative simulation is used to expand all possible outcomes based on the initial values. The resulting violence outcomes are used as the basis for prediction. Then, the simulation graph is used to point out specific settings in which intervention is particularly important.

7.3.1 Estimating the Likelihood of Different Outcomes.

The qualitative simulator uses the initial setting of the world state (partial state information is acceptable) as input and produces a simulation state-transition graph. Each sequence of states, following transitions from the initial state and ending with a different outcome state is a possible system trajectory—a possible sequence of qualitative state changes that may occur given the initial state and the qualitative dynamics specified. The end state in each path is where the system dynamics do not allow any further evolution

Participants:
  Population
  Nature
  Police
Quantities:
  Population:
    …***Similar to Israeli-Police model***..
    Intervention strength: QS = {low
    Event order: QS = {low high}
    Anonymity: QS = {low, high}
    Visual cohesiveness: QS = {low,
  Nature:
    …***Similar to Israeli-Police model***..
    Light: QS = {low, high}
  Police:
    …***Similar to Israeli-Police model***..
Influences:
  The influences in following quantities did
  [Hostility to police, Personal price, History
  Event purpose, Violent core, United identit
  Time sensitivity, Place sensitivity, Weather

  I+(Group speaker{exist}, Order)
  I+(License{ have},Order)
  I-(License{don't have},Order)
  I+(Order{ low},Violence)
  I-(Order{high},Violence)
  I+(Visual cohesiveness{high},Group cohesi
  I+(Group cohesiveness{high}, Anonymi
  I-(Group cohesiveness {low}, Anonymi
  I+(Anonymity {low}, Punishment)
  I-(Anonymity {high}, Punishment)
  I+(Time {morning/noon}, Light)
  I-(Time {evening/night}, Light)
  I+(Light{low}, Anonymity)
  I-(Light{high}, Anonymity)
  I+(#participants{100-1000/1000-10000/
  I-(#participants{0-100}, Anonymity)
  I+(Intervention strength {medium/high}
  I-(Intervention strength {low},Punishme
  I+(Intervention strength {high},Hostility

Figure 7.5: **Description of the BIU Model**

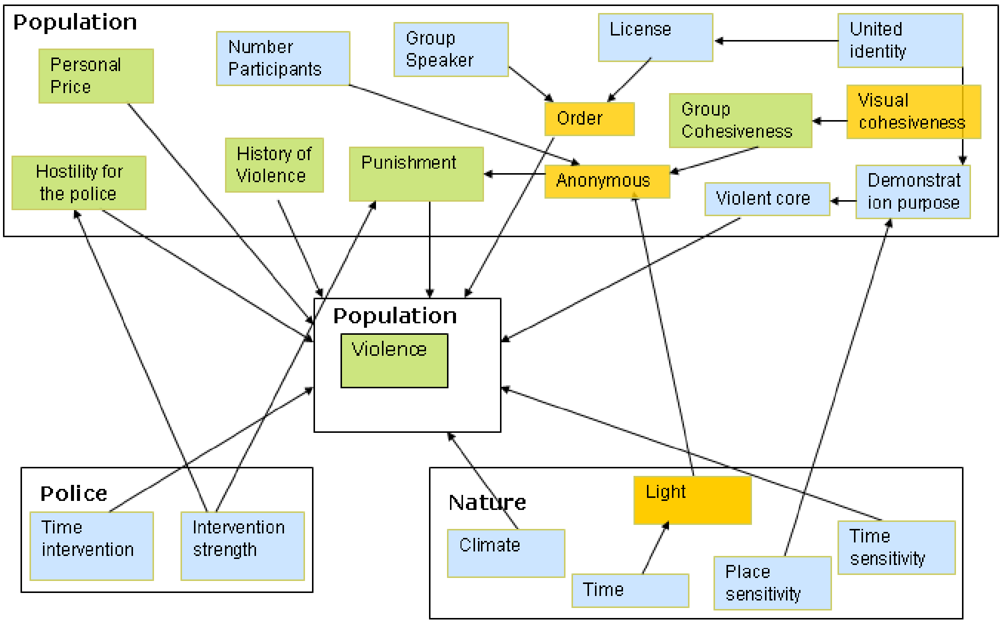


Figure 7.6: **BIU Model Structure**

(i.e., the system is stable). Taking the value of the outcome variables (in our case, violence level) in these final states enables categorical predictions.

The violence level variable can be any of three categorical values: zero, low and high. Zero value represents demonstrations that end without any causalities and also without any property damage. Low value represents demonstrations that end with property damage but without any causalities, and high represents demonstrations that end with causalities.

However, since in a sufficiently complex model, all three possible outcomes will have at least one stable state in which they appear, it is not sufficient to know whether a demonstration might be violent. Instead, our goal was to estimate the likelihood of different outcomes. To accomplish this goal, we used the received state-graph as an input and based on this developed graph we calculated the likelihood of different outcomes by counting the number of behavior paths that lead to a specific violence level, and dividing it by the total number of paths. The result is a distribution over possible violence outcomes.

For instance, suppose that there are 345 total paths leading from the initial simulation state to stable states (leaves in the simulation graph). Furthermore, suppose that 123 of these paths end up in leaves with a high level violence, *high*, 121 of the paths end up in leaves with a low level violence, *low*, and the remaining 101 paths end up with a no violence level *zero*. Then the distribution of the predicted violence levels is $\langle high, low, zero\rangle = \langle 123/345, 121/345, 101/345\rangle = \langle 0.36, 0.35, 0.29\rangle$.

7.3.2 Determining Important Factors in Specific Settings.

The a-priori predictions of the model, given initial values, do not provide decision-makers with information about factors that, *in a particular case*, will influence the level of violence. Thus we do not know which of the many different factors that may increase the level of violence are important in the specific case being simulated.

For instance, the perception of anonymity among the demonstrators may reduce their fear of being punished for breaking the law, which in turn may increase the chances of violence erupting during a demonstration. Perception of anonymity can be addressed by the police by various means: segregating the demonstrators into smaller disconnected groups, shining bright lights (if the demonstration is held when dark), etc. However, a-priori, there are few indicators of the potential anonymity perceived by the crowd. Moreover, we do not know whether tackling such perception can be effective: It could be that there are so many factors increasing the violence, that anonymity (being an indirect influencing factor) is just not worth treating. Or likewise, it could be that violence is highly unlikely, and thus bringing in bright lights is just an overkill that may incite the crowd. Thus anonymity should be addressed only in specific settings, where it becomes a determining factor in promoting violence.

To aid in this decision-making process, we describe an algorithm for determining the k most important factors in influencing the outcome of the simulation, and also for determining the conditions under which they should be addressed.

First, we traverse the simulation graph bottom up (from leaves to the root, which is the initial state). In each node, we count the number of paths resulting from it, which end up in the high level of violence, in the low level of violence, or in the zero level of violence. This process is in fact a generalization of the prediction process described above for making predictions. The number of paths of each type which is associated with the initial state is exactly the outcome distribution which we describe above. In this case, we simply generate the same count for all nodes in the graph.

Then, we identify the k nodes with the highest level of *outcome entropy* that have more than a single child[1]. The outcome entropy measures the uniformity of the distribution of different potential violence outcomes. A perfectly-uniform distribution $\langle 0.33, 0.33, 0.33 \rangle$ will have the maximal entropy; a perfectly non-uniform distribution where all paths lead to the same outcome will have minimal entropy (0).

The reason for seeking simulation nodes with a high entropy is that these are the nodes where a difference may be made, i.e., they are potentially *actionable*. Nodes with a low entropy are those in which the outcome is essentially already determined. Changing their outcome will necessarily involve making multiple changes to the state, i.e., involve more complex intervention. In contrast, nodes with a high entropy are nodes with an outcome which has not been determined, and thus provides a good opportunity for relatively

[1] A parent with a single child will have the same number of paths going through it as its child, and thus the same entropy. But we seek the state where the divergence into multiple outcomes occurs, hence we prefer the child.

simple intervention.

Given the k highest-entropy nodes, we can identify the factors that influence the outcomes. We do this by examining the simulation information saved at the node, and compare it with that of its children. We then determine which qualitative relations are at work at the node, and how they interact to lead towards the different outcomes. This significantly narrows down the list of factors that are relevant to the different outcomes, and also unravels the conditions under which these factors are important.

For instance, we may see that a node splits into different children due to the interaction between two opposing forces: Low anonymity which decreases violence (it increases the chances of punishment, as perceived by the crowd), and the lack of police response to events (i.e., the police respond too late, or with not enough force) which increases violence. Both these factors interact to cause multiple possible outcomes. Taking action (e.g., by increasing police force) can countermand the interaction, and cause the outcomes leading from this node to converge towards low or zero violence. Moreover, the state represented by the simulation node indicates the conditions under which increasing the police force will be affective (as this action may not always be the correct response to violence!).

Chapter 8

Evaluation of the Qualitative Reasoning Approach

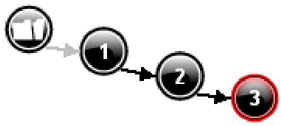
To evaluate the approach described in chapter 7 we implemented the three models in GARP3 [14], a QR engine including a user-friendly visual interface, which enables building and simulating qualitative models. GARP3 has been successfully used in many domains [15, 69]. We also developed 24 test-cases, i.e., real-life demonstrations reported in a variety of sources. Twenty two of these cases were taken from Hebrew Wikipedia under the category demonstrations [85]. The cases were taken both from the main category, and from the subcategories: "demonstrations in Israel" and "massacres of demonstrators". We excluded general descriptions which did not describe a specific event (e.g., descriptions of recurring demonstrations) and we also omitted two cases due to lack of information (resulting in a total of twenty cases). Three additional cases are well known events which have been extensively analyzed and described [18, 51, 74, 81] by experts. The last event was a peaceful demonstration that we video-taped.

To initialize the test cases, we utilized the information appearing in their descriptions in the literature and in Wikipedia. We initialized only the quantities for which we had explicit information. Quantities for which we had no information or ambiguous information were removed from the initial set. Qualitative simulation can work with partial information of this type.

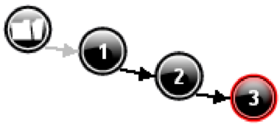
8.1 Prediction accuracy

Each model was examined on the twenty four test cases described above. We use the simulation state graph for our calculation of the numeric probability as presented earlier. Figure 8.1 presents the example of the transitions state-graph built by GARP in one of the experiments. Figure 8.1(a) presents the Base model state-graph, Figure 8.1(b) presents the Police model state-graph in the same experiment and Figure 8.1(c) presents the BIU model state-graph in the same experiment. The circles represent states and the arrows represent state transitions. The end path circles are the final states with one of the possible outcomes: zero, low or high level of violence. The figure demonstrate the results of the BIU model on a much more complex and richer state-graph than the Base and Police models, which elucidates the complexity of reasoning afforded by a rich qualitative model.

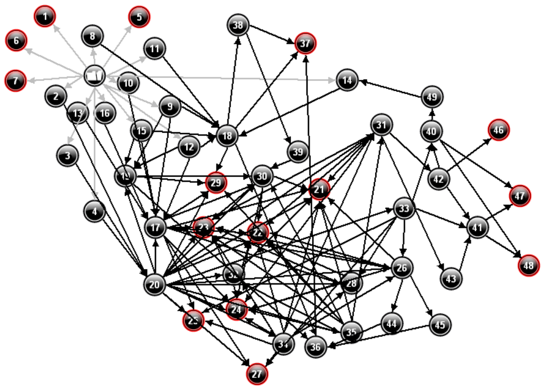
In evaluating the predictions of the different models, we observed the maximum likelihood predictions of each model, for the 24 different cases. If the maximum-likelihood prediction corresponded to the outcome of the event



(a)



(b)



(c)

Figure 8.1: Transitions state-graph

in the real world, we denoted this as an error of 0. Otherwise, we examined how far off the prediction was from the actual outcome, and we defined the following types of errors:

- Yes, denoting success. Equivalent to an error of 0.
- One level error, corresponding to one ordinal mistake level such as classification by the model as a low level of violence instead of a high level
- Two level error, corresponding to two ordinal mistake levels such as classification by the model as a high level of violence instead of zero.

Figure 8.2 summarizes the experiment results across the 24 cases. The different models are presented on the horizontal axis. The vertical axis measures the number of cases. The results of the three models are presented as stacked bars. Their total height is always equal (24 cases), but they are internally divided into 0-level errors, 1-level errors, and 2-level errors. All three models achieved 19 of 24 possible successes. While the *Base* and *Police* models had five cases of 2-level errors, the *BIU* model had one case of 2-level errors (predicted high levels of violence where there was none) and four cases of 1-level errors (predicted low levels of violence where there was none). The predictions of the *Bar Ilan* model are thus noticeably closer to the actual outcomes.

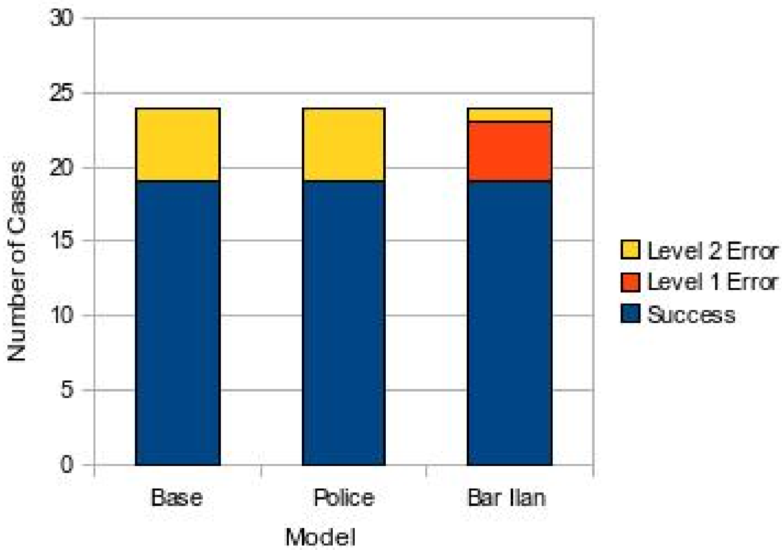


Figure 8.2: Results of the models' predictions

8.2 Comparison to the Machine Learning Techniques

There are other techniques in the field of artificial intelligence which can be used to make qualitative (ordinal, in this case) predictions. We decided to compare the QR predictions with those of a popular machine learning algorithm, i.e., decision tree learning [55].

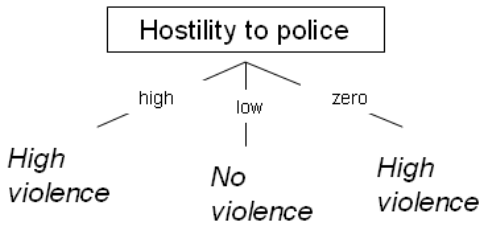
To do this, we used WEKA [10], an open source, user-friendly software that contains a collection of machine learning algorithms and we used the J48 decision tree algorithm, which is considered state-of-the-art.We used the algorithm with the default parameters (confidence threshold of 0.25, and a minimum of two instances per leaf). Testing was carried out via 10-fold cross-validation, which is a standard in this field. We built three files that were used as an input for WEKA. Each file contained a collection of attributes with their values and was built based on the quantities initialization set of each QR model (Base model, Police model and BIU model). The target class value of the violence attribute in each file was set based on the real-life event outcome. The output of the J48 algorithm is a learned decision tree and classification statistics.

Figure 8.3 presents the decision trees that were learned based on the each QR model's initial quantity set. Figure 8.3(a) presents the decision tree that was learned based on the quantity set of the Base model (Base decision tree), Figure 8.3(b) presents the tree that was learned based on the quantity set of the Police model (Police decision tree) and the same tree was learned based on the quantity set of the BIU model (BIU tree).

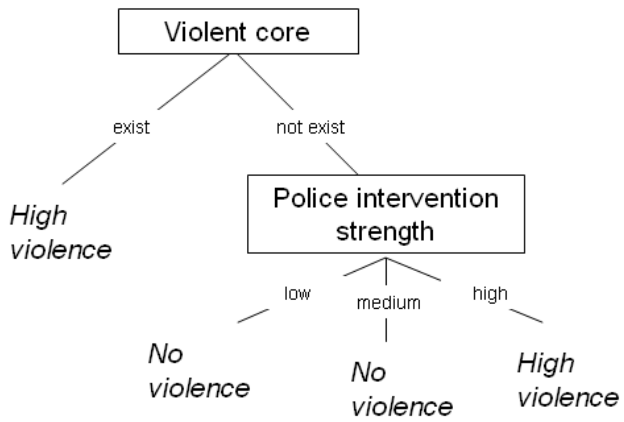
The results show that the Police tree and the BIU tree achieved 100% success in classification, while the Base tree attained only 70.8% of success. Although the decision tree technique provided an accurate prediction, which was a slightly better prediction than the BIU model with the QR approach, in the next section we claim that the QR approach is much more sensitive to changes and can account for *what if* scenarios. Thus, the QR approach is better for reasoning on the potential violence level and can improve the police decision making process.

8.3 Sensitivity Analysis

In the following experiments we demonstrate the use of QR and decision trees for a variety of hypothetical changes. According to experts [51,74,81] in several of the events we modeled (Exp. 15–17), the police intervention strength was found to be one of the important factors in the eruption of violence. Thus, in this section, we examined the presented QR model's prediction and the decision tree technique with *what if* scenarios.



(a)



(b)

Figure 8.3: Decision trees

First we examined whether the presented models with the QR approach and the decision tree technique are sufficiently sensitive to changes in terms of *what if* scenarios. We then examined the degree of influence the police intervention strength has on the event outcome, i.e., could this be the main factor that could prevent violence or the essence of the event itself is violent and the police intervention strength has little effect? Then we examined hypothetical situations of changing the probability of violence in several test case scenarios by changing different controlled factors and not only the police intervention strength.

8.3.1 Sensitivity Analysis: Experiment 1

In this experiment we examined whether the models based on the QR approach and the machine learning technique, could take into account changes in the police intervention strength. We used the same twenty four test cases described in Section 8 and examined the police intervention strength attribute with all its possible values. As in Section 8, we estimated the likelihood of different event outcomes. The model was considered to be sensitive to the changes if it provided a different outcome for different values in the attribute examined. The change could be one of the following: different distribution but no change in classification, or a different distribution, as well as a change in classification.

We compared the BIU and Police models built with QR techniques to a decision tree that was built with the BIU initialization set. The Base model built with QR techniques is irrelevant for this experiment since the Base model does not take into account the factor of police intervention strength and thus there would be no changes in the model's predictions.

The results show that the Police model changed its distribution in five of the test cases (of twenty four) and in two of them it also changed its classification. The BIU model changed its distribution in all of the examined test cases and in seven of them it also changed its classification. The decision tree is not able to provide distribution for all possible outcomes; it can only provide the final classification, thus unless there was a change in classification the prediction remains the same[1]. Out of twenty four examined test cases, the decision tree changed its classification in six.

[1]Note that there are other machine learning algorithms that result in a complete distribution that exist, and thus may be useful in this task. Exploring their use remains high on our list of future work.

These results demonstrate that the QR model (the BIU model) is much more sensitive to changes in the input conditions than the competing QR model (the Police model) as well as the decision-tree. But of course, these are all hypothetical changes and we do not know whether the modified predictions will correspond to the hypothetical results. In other words, the question remains whether the predictions were correctly changed.

8.3.2 Sensitivity Analysis: Experiment 2

To address the question raised in the previous section, we used three test cases which were explored independently by experts. The experts had already discussed the hypothetical results of the changes in the police force used in these test cases, and their conclusions could be used to validate the predictions of the models:

- The first case, Exp. #15, was the 1985 Heysel Stadium Disaster, during the European Cup Final. According to Lewis [51] who analyzed this event, one of the reasons for this violent outcome was the police's lack of intervention to prevent the emergent violence.

- The second case, Exp. #16, was the Los Angeles Riots which occurred in 1991. Useem [81] who analyzed this event, argued that the police were not properly organized and did not react in time with appropriate force to prevent the eruption which allowed a violent core to grow.

- The third case, Exp. #17, was the London Riot Disaster which occurred in 1990. As opposed to the previous two events, in this case the police used extensive force against the protesters without distinguishing between anarchists and peaceful marchers [74]. The marchers fought back, and this changed the initially-peaceful protest into a very violent event with many casualties.

Table 8.1 presents the experiment results. The first column corresponds to the examined test case. The second column corresponds to recommended police intervention strength. Then we present the models' predictions for each possible outcome: no violence, low violence and high violence. Below each prediction, we present the change, if any, in the recommended prediction. *Dist. Change* denotes a change in the distribution, but not in the overall prediction; *Classif. Change* signifies a change in the classification.

For example, if in the Heysel stadium disaster (Exp. #15), the police would have increased their intervention strength, the results show that there would be a change in the Police and BIU distributions. With this change the BIU model predicted 83% probability of high violence instead of 96% (before the recommended change). Though there was a change, the maximum posterior prediction remained as high violence. The reason for this outcome was that there were other factors in this test case that led to high violence. Another example was the London disaster. If we decrease the police intervention strength, the BIU model predicts only a 19% probability of high violence and 45% of low violence instead a 57% probability of high violence. In this test case the BIU model changed its classification. Instead of predicting a high violence, with this change it predicted a low violence. Thus, it is apparent that the police force had immense influence on the event outcome in this test case, which corresponds to the experts' expectations.

The results demonstrate that the decision tree technique is not sensitive to the examined changes that were claimed by the experts. The Police model performed slightly better than the decision tree (which changed the distribution in Exp. #15, but failed in the two other test cases). However, the BIU model provided good results which match the experts expectations, and indicate that it can account for *what if* scenarios. We remind the reader that these results are limited to the use of the decision tree. Other machine learning algorithms could potentially perform better; we leave such evaluation for future work.

8.3.3 Sensitivity Analysis: Experiment 3

In this experiment we examined whether we can further decrease the violence level in test cases 15 (Heysel Stadium disaster) and 16 (LA riots). We used the same initial settings with several updates as explained below. Some factors such as weather or history of violence cannot be changed, while others can be controlled. For example, police intervention strength, anonymity and order are examples of features that can be manipulated in the sense that actions can be done to change their values. Police may increase the intervention strength by using more manpower or by using different types of weapons. The existence of projectors and cameras in the demonstration area will decrease the participants' perception of anonymity.

Table 8.2 presents the experiment results. In this experiment we examined the BIU model and the decision tree. The first column corresponds to the

Exp. #	Recommended Change	Model Outcome	**Police Model**	**BIU Model**	**Decision tree**
15	Increase strength [51]	High violence Low violence No violence	66% 0 34%	83% 6% 10%	
	Prediction Change	High	High **Dist. Change**	High **Dist. Change**	High No Change
16	Increase strength [81]	High violence Low violence No violence	66% 0 34%	87% 3% 10%	
	Prediction Change		High No Change	High **Dist. Change**	High No Change
17	Decrease strength [74]	High violence Low violence No violence	80% 0 20%	19% 45% 36%	
	Prediction Change		High No Change	Low **Classif. Change**	High No Change

Table 8.1: **Results of changing the police intervention strength.**

examined test case and the second column corresponds to the changed initial values of the quantities. Then we present the models predictions before the change and after.

Here again the results demonstrate that the decision tree technique is not sensitive to the changes. This is not surprising, since the only components of the learned tree which can change its classification is the existence of violence core and the police intervention strength. In contrast, the BIU model was found to be sensitive to the changes.

We emphasize that these results do not lead to the conclusion that machine learning algorithms in general cannot be used in ways that will allow some hypothetical reasoning and sensitivity analysis. For example, a novel form of visualization [58] has recently been used to allow certain sensitivity analysis with another machine-learning method (*Naive Bayes*).

8.4 Determining Influential Factors

The previous section demonstrates that while QR models are sensitive to hypothetical changes to the simulated quantities, not all changes can cause a qualitative change in the predictions of the system. In other words, not all factors have equal weight on the outcome of a particular case. Attempting to test all factors, in the hope of identifying those that are important is a long computer-intensive process that is not scalable.

We now turn to evaluating the use of the algorithm described in Section 7.3 to determine the important factors influencing the outcome of the demonstrations. We ran the algorithm on the resulting simulation graphs for the three cases (Exp 15–17) for which we had expert analysis in addition to the predictions of the different models. We requested the 5 highest-entropy nodes. The algorithm analyzed the information associated with them to determine which factors interacted to cause the different outcomes to form (or more accurately, to create children leading to the different outcomes).

In Table 8.3 we report the top factors increasing violence in cases Exp. 15–17. The second and third columns in the table show the factors determined by the algorithm, and the factors determined by experts in the field, who have analyzed these cases.

In case Exp. 15 there was complete agreement between the algorithm and the expert. In case Exp. 16 there was only partial agreement: Both the algorithm and the expert agree that the police responded with insuffi-

Exp. #	Changed initializations	**Violence level**	**BIU Model before change**	**BIU Model after change**	**Dec. Tree before change**	**Dec. Tree after change**
Exp15	Police strength: medium, Punishment: high, Anonymity: low	High Low No	96% 3% 1%	80% 6% 14%		
	Prediction Change			High **Dist. Change**	High	High No Change
Exp16	Police strength: medium, Punishment: high, Order: high	High Low No	99% 1% 0%	80% 6% 14%		
	Prediction No Change		High	High **Dist. Change**	High	High No Change

Table 8.2: **Results of hypothetical manipulations.**

Exp. #	Factors (Algorithm)	Factors (Experts)
15	Police Strength Too Low	Police Strength Too Low [51]
16	Police Strength Too Low, Number of participants > 100	Police Strength Too Low, Too Late [81]
17	Hard Core of Demonstrators, Low Perceived Chance of Punishment	Police Strength Too Much [74]

Table 8.3: Important Factors in Test Cases Exp. 15–17.

cient strength, but the expert also pointed out that its intervention occurred too late. The algorithm, in contrast, reported the number of participants as a significant factor in the violence. This is a factor that cannot typically be changed dynamically, and of course the algorithm cannot differentiate static from dynamically-controllable factors (we leave such extension for future work). Note, however, that the algorithm does recognize that the eruption of violence occurs when, in addition to police responding too weakly, they also respond too late. Nonetheless, the algorithm does not report this fact as a key factor in the eruption of violence. In this result the algorithm differs from the expert opinion.

Finally, disagreement between the expert and the algorithm is apparent for case Exp. 17. The expert believes that the main factor accounting for the violence is that the police acted too harshly, while the algorithm points out the existence of a core of demonstrators and a low perceived chance of punishment, as the key factors. Note, of course, that the algorithm and experts do not provide contradictory results. Perhaps both are correct. Our algorithm's goal was to discover opportunities for intervention, and perhaps the expert's analysis accounted for a large portion of the state space, where no intervention is possible (since there, the police acted too harshly, but this cannot be reversed).

Indeed, almost as a side-effect of this analysis, not only do we discover which factors are important, but also under what circumstances to act on them. These circumstances are easily determined by examining the state of the qualitative behavior, as denoted by the node in question. Take for example Exp. 15, where the highest-entropy state (where the algorithm recommends increasing police response) has the following attributes: Moderate weather, high cohesion of the demonstrators, emotional event, a hard core of demonstrators is present, between 100 and 1000 participants, weak police strength applied, too quickly, lack of a spokesman or representative for the demonstrators, evening hours and dark, property damage already caused by the demonstrators (i.e., low violence already erupted). Under these specific settings, the corrective action to take would be to immediately increase the strength of the police response, in hopes of preventing the violence from escalating.

Chapter 9

Future Directions and Final Remarks

We summarize the key contributions of this thesis in Section 9.1 and in Section 9.2 we discuss future directions for this research.

9.1 Summary of the Key Contributions

In the first part of this dissertation we used an agent-based approach for modeling crowd behaviors. Our contribution from this part of the work is as follows:

- We presented validation of the SCT model (and competing models) against human crowd behavior. We evaluated the SCT on pedestrian phenomena and showed that the SCT model generated pedestrian behavior more in tune to human pedestrian behavior. The results support the general applicability of the SCT model.

- We evaluated the SCT model in the evacuation domain. The use of SCT in evacuation leads to increased grouping of the agent (herding), as we show in the experiments and this herding has been reported to occur in real-world emergencies [2, 36].

- We have shown that it is possible to have a task-neutral, architecture-level, social comparison process capable of working well across multiple tasks. We explored alternative action-selection mechanisms that enable continuous comparison (required for such processes), and provided evidence that one of the mechanisms, based on threshold selection, was superior.

- We revised the underlying comparison process itself to account for the comparison group size, a key issue neglected in previous work. We have shown that in some cases the revised model offers superior performance to that of the previous model. Nonetheless it is important to note that our results are less conclusive than we would have preferred.

- We explored the impact of micro-level, individual agent, cultural parameters on macro-level crowd behavior. Building on existing literature which investigated culture in human crowds, we identified important cultural parameters in two physically crowded domains (pedestrian movement and evacuation). We implemented these parameters in established agent-based simulations for these domains, and used the simulations to measure their impact on crowd dynamics. We thus go beyond existing work, which has focused on describing cultural parameters of individuals, without investigating their crowd-level effects.

- Exploring cultural differences in the pedestrian motion domain, we conducted three sets of experiments. The first explored the effect of each parameter by itself, in mixed crowd settings (mixed, in the sense that the parameter in question were varied among the agents). The second explored mixing agents, each with a pre-set bundle of the parameters (i.e., values of each of the parameters matched recorded videos from different countries and cultures. Finally, the results of the simulation were quantitatively validated against data extracted from videos of crowds in five different countries.

- Exploring cultural differences in the evacuation domain, we presented a subset of results which demonstrate how cultural parameters (such as the seriousness with which evacuees treat indications of the need to evacuate) affect evacuation time and panic levels. For these experiments, we additionally examined the effect that authority figures have on the evacuation measures. We found that in some cultures (in particular where agents treat evacuations seriously), authority figures do not speed up evacuations. In others (in particular where agents do not take evacuations seriously), the authority figures have a calming effect (lower the panic level), while they still increase the rate of evacuation.

In the second part of this dissertation we used a qualitative reasoning approach to modeling crowd behavior. Our contribution from this part of the work is as follows:

- We proposed the use of qualitative reasoning (QR) method in social sciences. Its use in modeling demonstrations, as presented in this dissertation, is only one of the possible domains of application. The nature of qualitative modeling seems to fit well with the nature of social science knowledge. This proposed technique can therefore be used to validate and test theories against qualitative data.

- We presented and compared three QR models for predicting the level of violence in demonstrations: the Base model, the Police model and the BIU model.We evaluated these models on twenty four real life test case scenarios. The results show that the BIU model makes better predictions than the compared models and it also was found to be sensitive to changes.

- We compared the performance of our QR models with the machine learning method, i.e., a decision tree. While, the machine learning method made accurate predictions, it failed in the sensitivity analysis. Thus, the BIU model built with the QR approach can account for *what if* scenarios as opposed to the decision tree. Furthermore it is preferable for reasoning about the potential violence level in order to improve the police decision making process.

- We developed an algorithm which analyzes the qualitative simulation graph of each test-case to determine the factors that are most important in influencing the outcomes of the specific case under consideration.

9.2 Future Directions

Indeed, there are many questions still open. We believe that there are several important directions for future work in reference to the agent-based approach:

- Preliminary experiments with human subjects' evaluation of synthetic situations have consistently yielded replies that indicate expectations not only of reduced attraction in some cases, but of actual avoidance. In other words, in some settings, human subjects sometimes expect agents to move away from a group, rather than simply not move towards it. This avoidance is also discussed in more modern elaborations of social comparison theory, but is not yet accounted for in our models. We plan to extend the SCT model to include the repelling forces. Thus, each agent should not only be attracted to the similar but also should avoid the dissimilar.

- A possible direction of research which we hope to pursue would be to take advantage of the highly accurate predictions made by the models we have developed. Building on their fidelity, it may now be possible to begin investigating their use for tasks other than simulations. For instance the use of these models to identify suspicious behaviors (e.g., a person posing as a pedestrian, but really not belonging to the crowd) should be considered.

- We plan to continue our work on collecting and analyzing movies of crowds in different cultures. We believe that it would be possible to form an international collaboration with colleagues in different countries in order to collect data and make it available to crowd researchers world-wide. We believe that the availability of annotated, analyzed data is a true stumbling block in this field.

There are also several points which have been left for future work in reference to the qualitative reasoning approach:

- We plan to expand our model to account for bidirectional influences (feedback loops). For example, in the BIU model the hostility toward the police increases the violence level. However, increasing the violence level has no impact on hostility. We believe that such expansion is necessary to provide a more accurate prediction.

- We plan to tackle the next logical step in the use of QR for social simulation, which is to advance beyond determining the important factors, to determining plans of action that utilize them.

Bibliography

[1] F. H. Allport. *Social Psychology.* Boston: Houghton Mifflin, 1924.

[2] K. Andrée and B. Eriksson. Cultural differences in an evacuation scenario - a study comparing australian and swedish responses. Technical report, Lund University, 2008.

[3] Answers.com. What is the average man's walking stride length? http://wiki.answers.com/Q/What is the average mans walking stride length, 2011.

[4] T. Balch. *Behavioral Diversity in Learning Robot Teams.* PhD thesis, Georgia Institute of Technology, 1998.

[5] C. Beaulieu. Intercultural study of personal space: A case study. *Journal of applied social psychology*, 34(4):794–805, 2004.

[6] R. Berk. A gaming approach to crowd behavior. *American Sociological Review*, 1974.

[7] W. Berkowitz. A cross-national comparison of some social patterns of urban pedestrians. *Journal of cross-cultural psychology*, 2(2):129, 1971.

[8] V. J. Blue and J. L. Adler. Cellular automata microsimulation of bidirectional pedestrian flows. *Transportation Research Record*, pages 135–141, 2000.

[9] H. G. Blumer. Collective behavior. *Principles of Sociology*, 1939.

[10] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. WEKA–experiences with a java open-source project. *Journal of Machine Learning Research*, 11:2533–2541, 2010.

[11] G. Brajnic and M. Lines. Qualitative modeling and simulation of socio-economic phenomena. *Journal of Artificial Societies and Social Simulation*, 1998.

[12] A. Braun, S. R. Musse, L. P. L. de Oliveira, and B. E. J. Bodmann. Modeling individual behaviors in crowd simulation. In *Computer Animation and Social Agents*, pages 143–148, 2003.

[13] A. Braun, S. R. Musse, L. P. L. de Oliveira, and B. E. J. Bodmann. Modeling individual behaviors in crowd simulation. *CASA*, 2003.

[14] B. Bredeweg, F. Linnebank, A. Bouwer, and J. Liem. Garp3 - workbench for qualitative modelling and simulation. *Ecological Informatics*, 4(5–6):263–281, 2009. http://hcs.science.uva.nl/QRM/software/.

[15] B. Bredeweg and P. Salles. Mediating conceptual knowledge using qualitative reasoning. In S. Jrgensen, T.-S. Chon, and F. E. Recknagel, editors, *Handbook of Ecological Modelling and Informatics*, pages 351–398. Wit Press, Southampton, UK, 2009.

[16] J. L. Bryan. Cultural variations in the behavior of people in fire situations. In *International Conference on Fire Safety*, 1978.

[17] K. Carley and A. Newell. The nature of the social agent. *Journal of Mathematical Sociology*, 1994.

[18] A. Carmeli and I. Ravid-Yamin. Research report on the subject of crowd events and public order. Technical report, Ministry of public security, Bureau of the chief scientist, 2006.

[19] U. Chattaraj, A. Seyfried, and P. Chakroborty. Comparison of pedestrian fundamental diagram across cultures. *Advances in complex systems*, 2009.

[20] W. Daamen and S. P. Hoogendoorn. Experimental research of pedestrian walking behavior. *Transportation Research Record*, pages 20–30, 2003.

[21] L. Festinger. A theory of social comparison processes. *Human Relations*, pages 117–140, 1954.

[22] K. D. Forbus. Qualitative reasoning. In *CRC Handbook of Computer Science and Eng.,*. CRC Press, 1996.

[23] K. D. Forbus and S. E. Kuehne. Towards a qualitative model of everyday political reasoning. In *Proceedings of the Nineteenth International Qualitative Reasoning Workshop*, 2005.

[24] N. Fridman and G. A. Kaminka. Towards a cognitive model of crowd behavior based on social comparison theory. In *AAAI-07*, 2007.

[25] N. Fridman, G. A. Kaminka, and M. Traub. First steps towards a social comparison model of crowds. In *International Conference on Cognitive Modeling (ICCM-09)*, 2009.

[26] N. Gilbert and K. G. Troitzsch. *Simulation for the social scientist*. Open University Press, 2005.

[27] G. R. Goethals and J. M. Darley. Social comparison theory: An attributional approach. In *Social comparison processes: Theoretical and empirical perspectives*. Washington, DC: Hemisphere, 1977.

[28] G. R. Goethals and W. M. P. Klein. Interpreting and inventing social reality: Attributional and constructive elements in social comparison. In *Handbook of social comparison: Theory and research*. New York: Plenum, 2000.

[29] K. L. Hakmiller. Threat as a determinant of downward comparison. *Journal of experimental social psychology*, 2:32–39, 1966.

[30] E. Hall. A system for the notation of proxemic behavior. *American anthropologist*, 65(5):1003–1026, 1963.

[31] E. Hall. *The silent language*. Oxford, England: Anchor, 1973.

[32] E. Hall. *Beyond culture*. Anchor books, 1977.

[33] E. Hall and M. Hall. *Understanding cultural differences*. Intercultural Press, 1990.

[34] D. Helbing. Boltzmann-like and boltzmann-fokker-planck equations as a foundation of behavioral models. *Physica A*, 196:546–573, 1993.

[35] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews on Modern Physics*, 73, 2001.

[36] D. Helbing, I. J. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407:487–490, 2000.

[37] D. Helbing and P. Molnar. Self-organization phenomena in pedestrian crowds. In F. Schweitzer, editor, *Self-organization of Complex Structures: From Individual to Collective Dynamics*, pages 569–577. Gordon and Breach, 1997.

[38] D. Helbing, P. Molnar, I. J. Farkas, and K. Bolay. Self-organizing pedestrian movement. *Environment and Planning B*, 28:361–384, 2001.

[39] D. Helbing and T. Vicsek. Optimal self-organization. *New Journal of Physics*, 1999.

[40] H. Helson. *Adaptation level theory.* New York: Harper and Row, 1964.

[41] L. F. Henderson. The statistics of crowd fluids. *Nature*, 229:381–383, 1971.

[42] L. F. Henderson. On the fluid mechanics of human crowd motion. *Transportation research*, 8:505–515, 1974.

[43] H. A. Hornstein, E. Fisch, and M. Holmes. Influence of a model's feeling about his behavior and his relevance as a comparison other on observers' helping behavior. *Journal of Personality and Social Psychology*, 1968.

[44] W. Jager, R. Popping, and H. van de Sande. Clustering and fighting in two-party crowds: simulating the approach-avoidance conflict. *Journal of Artificial Societies and Social Simulation*, 4(3), 2001.

[45] J. Kamps and G. Péli. Qualitative reasoning beyond the physics domain: The density dependence theory of organizational ecology. In *Proceedings of QR95*, 1995.

[46] T. Kretz. *Pedestrian traffic: simulation and experiments.* PhD thesis, Universität Duisburg-Essen, 2007.

[47] B. Kuipers. Qualitative reasoning: modeling and simulation with incomplete knowledge. *Automatica*, 25(4), 1989.

[48] J. A. Kulik and H. I. M. Mahler. Social comparison, affiliation, and emotional contagion under threat. In *Handbook of social comparison: Theory and research*. New York: Plenum, 2000.

[49] G. Le Bon. *The crowd: A study of the popular mind*. Dunwoody, Ga., N.S. Berg, 1968 edition, 1895.

[50] R. Levine and A. Norenzayan. The pace of life in 31 countries. *Journal of cross-cultural psychology*, 30(2):178, 1999.

[51] J. M. Lewis. A value-added analysis of the heysel stadium soccer riot. *Current psychology*, 1989.

[52] M. J. Matarić. Designing and understanding adaptive group behavior. *Adaptive Behavior*, 4(1):50–81, December 1995.

[53] C. McPhail. *The myth of the madding crowd*. Aldine de gruyter, 1991.

[54] S. Milgram, L. Bickman, and L. Berkowitz. Note on the drawing power of crowds of different size. *Journal of Personality and Social Psychology*, 13(2):79–82, 1969.

[55] T. M. Mitchell. *Machine learning*. McGraw-Hill, 1997.

[56] M. Moussaïd, S. Garnier, G. Theraulaz, and D. Helbing. Collective information processing and pattern formation in swarms, flocks, and crowds. *Topics in ocgnitive science*, 1(3):469–497, 2009.

[57] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz. The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PLoS One*, 5(4):e10047, 2010.

[58] M. Mozina, J. Demsar, M. W. Kattan, and B. Zupan. Nomograms for visualization of naive bayesian classifier. In *Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2004)*, pages 337—348, 2004.

[59] A. Newell. *Unified Theories of Cognition*. Harvard University Press, Cambridge, Massachusetts, 1990.

[60] T. Osaragi. Modeling of pedestrian behavior and its applications to spatial evaluation. In *Autonomous Agents and Multiagent Systems (AAMAS)*, pages 836–843, 2004.

[61] A. Parducci. *Happiness, pleasure and judgment: The contextual theory and its applications*. Mahwah, NJ: Lawrence Erlbaum, 1995.

[62] S. Patrick, P. M. Dorman, and R. L. Marsh. Simulating correctional disturbances: the application of organization control theory to correctional organizations via computer simulation. *Journal of Artificial Societies and Social Simulation*, 2(1), 1999.

[63] M. Patterson, Y. Iizuka, M. Tubbs, J. Ansel, M. Tsutsumi, and J. Anson. Passing encounters east and west: Comparing japanese and american pedestrian interactions. *Journal of nonverbal behavior*, 31(3):155–166, 2007.

[64] N. Pelechano, C. Stocker, J. Allbeck, and N. Badler. Being a part of the crowd: towards validating vr crowds using presence. In *Autonomous Agents and Multiagent Systems (AAMAS)*, 2008.

[65] C. Reynolds. Opensteer: Steering behaviors for autonomous characters. *Website.[Available: http://opensteer. sourceforge. net]*, 2004.

[66] C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques (SIGGRAPH-87)*, pages 25–34, New York, NY, USA, 1987. ACM Press.

[67] C. W. Reynolds. Steering behavior for autonomous character. In *Proceedings of the Game Developers Conference*, pages 763–782, 1999.

[68] S. J. Rymill and N. A. Dodgson. A psychologically-based simulation of human behaviour. In *Theory and Practice of Computer Graphics*, pages 35–42. 2005.

[69] P. Salles and B. Bredeweg. Modelling population and community dynamics with qualitative reasoning. *Ecological Modelling*, 195:114–128, 2006.

[70] D. Schweingruber and C. McPhail. A method for systematically observing and recording collective action. *Sociological methods & research*, 27(4):451, 1999.

[71] J. E. Singer. Social comparison: progress and issues. *Journal of experimental social psychology*, 2:103–110, 1966.

[72] W. P. Smith and G. B. Arnkelsson. Stability of related attributes and the inference of ability through social comparison. In *Handbook of social comparison: Theory and research*. New York: Plenum, 2000.

[73] B. Steffen and A. Seyfried. Methods for measuring pedestrian density, flow, speed and direction with minimal scatter. *Physica A: Statistical mechanics and its applications*, 389(9):1902–1910, 2010.

[74] C. Stott and J. Drury. Crowds, context and identity: Dynamic categorization process in the ṕoll tax riot. *Human Relations*, 2000.

[75] D. Thalmann. The foundations to build a virtual human society. In *Proceedings of Intelligent Virtual Actors (IVA-2001)*, pages 1–14. Springer-Verlag, 2001.

[76] P. C. Tissera, M. Printista, and M. L. Errecalde. Evacuation simulations using cellular automata. *Journal of Computer Science and Technology*, 2007.

[77] M. C. Toyama, A. L. C. Bazzan, and R. da Silva. An agent-based simulation of pedestrian dynamics: from lane formation to auditorium evacuation. In *Autonomous Agents and Multiagent Systems (AAMAS)*, 2006.

[78] J. Tsai, N. Fridman, E. Bowring, M. Brown, S. Epstein, G. Kaminka, S. Marsella, A. Ogden, I. Rika, A. Sheel, et al. Escapes-evacuation simulation with children, authorities, parents, emotions, and social comparison. In *Proceedings of the 10th international conference on autonomous agents and multiagent systems (AAMAS 2011), Innovative applications track*.

[79] X. Tu and D. Terzopoulos. Artificial fishes: physics, locomotion, perception, behavior. In *SIGGRAPH '94': Proceedings of the 21st annual*

conference on Computer graphics and interactive techniques, pages 43–50, New York, NY, USA, 1994. ACM Press.

[80] R. Turner and L. M. Killian. *Collective Behavior*. Prentice Hall, 1993 edition, 1972.

[81] B. Useem. The state and collective disorders: The los angeles riot/protest of april 1992. *Social Forces*, 76, 1997.

[82] J. Volkmann. Scales of judgment and their implications for social psychology. In *Social psychology at the crossroads*. New York: Harper and Row, 1951.

[83] J. Wander. *Modelling consumer behavior*. PhD thesis, University of Groningen, 2000.

[84] L. Wheeler, J. Suls, and L. Wheeler. Individual differences in social comparison. *Handbook of social comparison: Theory and research*, pages 141–158, 2000.

[85] Wikipedia: The Free Encyclopeida. Category:demonstrations (hebrew). http://he.wikipedia.org/w/index.php?title=, 2010.

[86] U. Wilensky. NetLogo. Center for Connected Learning and Computer-Based Modeling—Northwestern University; http://ccl.northwestern.edu/netlogo/, 1999.

[87] M. Wolff. Notes on the behaviour of pedestrians. *In People in Places: The Sociology of the Familiar*, pages 35–48, 1973.

[88] K. Yamashita and A. Umemura. Lattice gas simulation of crowd behavior. In *Proceedings of the International Symposium on Micromechatronics and Human Science*, pages 343–348, 2003.