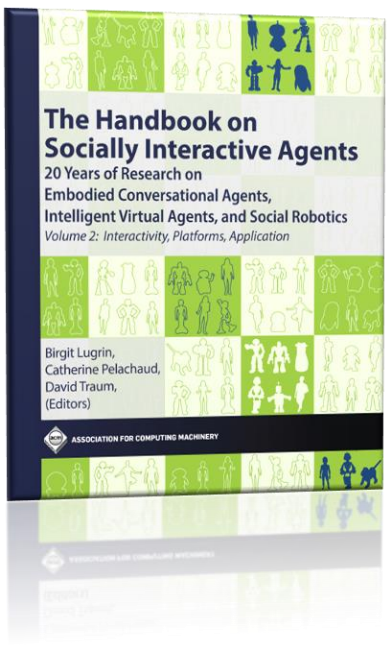
# Interaction in Social Space

Hannes Högni Vilhjálmsson




**Author note:**

This is a preprint. The final article is published in "The Handbook on Socially Interactive Agents" by ACM.

**Citation information:**

H. Vilhjálmsson (2022). Interaction in Social Space . In B. Lugrin, C. Pelachaud, D. Traum (Eds.), *The Handbook on Socially Interactive Agents – 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics*, Volume 2: Interactivity, Platforms, Application (pp. 3-44). ACM.

DOI of the final chapter: 10.1145/3563659.3563662

DOI of volume 2 of the handbook: 10.1145/3563659

Correspondence concerning this article should be addressed to Hannes Högni Vilhjálmsson, hannes@ru.is

# 14 Interaction in Social Space

Hannes Högni Vilhjálmsson

## 14.1 Motivation

> Most important of all, space is one of *the* basic, underlying organizational systems for all living things—particularly for people. [Hall 1966, xii]

A person occupies space and is surrounded by space, which may be occupied by other people. Our mere presence in the social space of others makes us participants in a social performance. Our actions are observed, interpreted, and reacted to. Some of those actions may lead to further interaction, while others may not. Through careful coordination of actions and reactions, we are capable of negotiating the terms of our co-presence, without even uttering a single word.

Humans are capable of managing their co-presence without much effort, their bodies giving off cues of position, orientation, and posture, as they traverse social spaces. There are situations though, that require more effort and strategic thinking, such as when an easily recognizable person of fame or notoriety needs to cross a public hall without causing commotion. For some, social spaces may even induce forms of social anxiety. Regardless of the ease or difficulty in dealing with social space, its existence is of paramount importance to human life as it envelopes face-to-face interaction with family, friends, colleagues, teachers, service personnel, and, in fact, all those who simply happen to live and work around us.

To interact with humans, SIAs need to occupy and manage these spaces as well. However, coming up with a general computational solution for a dynamic social space is not trivial, and that is perhaps one reason designers of SIA systems often choose to constrain the space heavily, for example by planting the SIA into a fixed formation with a human user who is ready to interact.

In order to set the SIAs free, and let them roam physical and virtual spaces where they can choose to interact or not to interact with those who share the space with them, we need to understand and appreciate the structure of that space and how that knowledge can be exploited to successfully manage human co-presence.

## 14.2 Models and Approaches

First, we look at how to describe social space and the social elements it contains. We then examine how the body fits into that space and by what means it can interface with it. This will

give us the necessary vocabulary to review the different social functions that people carry out with spatial behavior to achieve their social goals.

The theoretical models in this section represent the works of some of the pioneers of human public behavior studies, including Albert E. Scheflen, Adam Kendon, Erving Goffman and Edward Hall. Their models are frequently cited by those dealing with social space, but here their key concepts are organized and presented holistically to form a single framework. Instead of citing them at every turn in the text, tables are provided at the end of each section, summarizing the main theoretical concepts reviewed and from what source they were brought (see Tables 14.1 - 14.3). The accompanying figures in this section were manually created in 3D character modeling and animation tools[1], and are used to illustrate the concepts discussed. The figures therefore do not depict virtual agents being simulated in real time.

### 14.2.1 The Structure of Social Space

The term *social space* refers to any spatial environment where people have access to other people through their embodied presence. They have an opportunity to perceive each other using multiple senses and affect each other over non-verbal and verbal channels. The space could be anything from a city square to a classroom. It is not given that all occupants of a social space are socially involved or engaged with one another, but they all need to exhibit minimal social awareness and manage their co-presence.

Within the social space (see Figure 14.1), there may be *gatherings*, where two or more people are in each other's immediate presence, essentially forming collections of people of various sizes and shapes. A *social situation* is a social space or a section of it, where upon entering it a person immediately participates in a gathering. Naming a *social occasion* can explain the reason for the social situation, such as a staff meeting or a funeral. Gatherings can contain multiple social engagements, organized into formations as described below, but also individuals that are not fully engaged, for example *bystanders*.

While the term gathering describes a relatively loose collection of people, *formations* are tighter clusters where more coordinated activities occur. Formations are important because their structure represents a pattern of social convention that organizes and facilitates interaction. A formation is made up of individual *locations* that each can comfortably hold an individual body and provide room for the individual's actions. Typically, a location is larger than the size of the body, often extending half of the person's width in each direction (see Section 14.2.2). Locations can be pre-allocated, such as with furniture, and are claimed by participants during a social event. The simplest formation is the *dyad*, which is created by pairs of people who are committed to one another and cluster closely. Another fairly simple formation arises when people arrange themselves into an array and take a common orienta-

[1] The tools used were Character Creator® and iClone® from Reallusion®

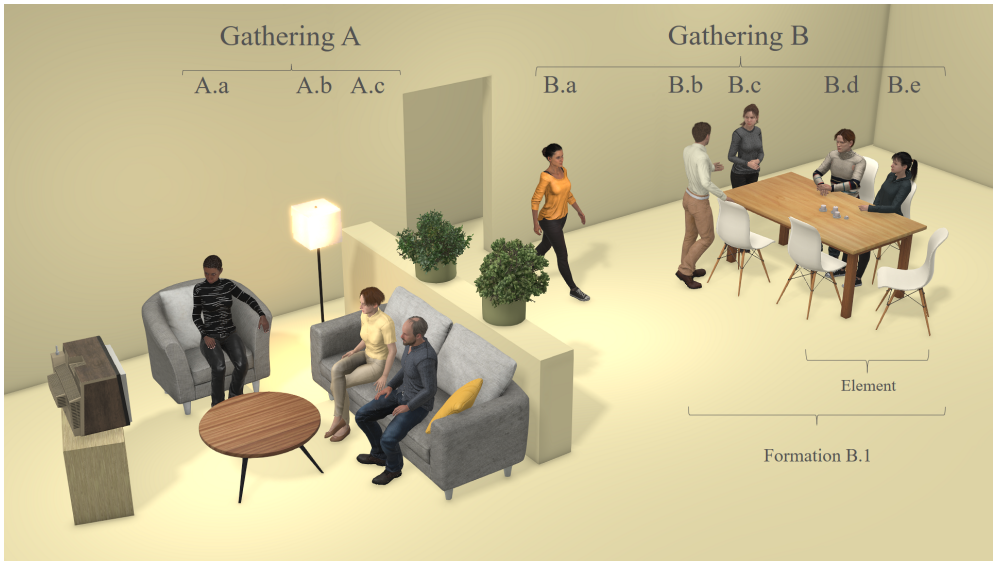



Figure 14.1: A relatively large indoor *social space* that includes two *gatherings*, A and B. Gathering A contains a single sitting *formation*, within which A.b and A.c form an *element*. Gathering B is a particular *social situation*, where the *occasion* is a staff meeting. The people around the table belong to *formation* B.1, inside of which B.d and B.e. form an *element*. Person B.a becomes a part of gathering B simply by walking through the door, due to the situation, even though she does not belong to a formation yet. In fact, she may just be passing by, in which case she may temporarily hold the role of a *bystander*. Each person occupies a *location*, and the empty chairs denote unassigned *locations*.

tion, such as when sitting together on a sofa to watch TV or standing side-by-side to watch an event. Such a formation has been called an *element*.

A more complex formation is the so-called *face-formation* or *F-formation*, which occurs when people take locations facing one another (see Figure 14.2). The shape of such a formation is influenced by the number of participants, typically growing from triangles for three people, to squares for four, and to expanding circles for more people. These often become unstable around 10-12 people, at which point they may break into smaller clusters.

The distance between participants, and thus the size of the face-formation, can vary according to many things like how involved or committed the participants are with each other, how noisy or crowded the environment is, or how furniture is arranged.

The space occupied by and surrounding a formation has been characterized as consisting of different *zones* or *regions*, each with a particular role. The innermost zone is the space formed by the overlapping orientations of participants, essentially the center. This space has


Figure 14.2: The space occupied by and surrounding a face-formation, or *F-formation*, has been characterized as consisting of three concentric zones. The *o-zone* is the unoccupied space in the center, the *p-zone* is occupied by the participants and the *r-zone* surrounds the formation.

been termed the *o-zone*. Surrounding that space, the participants themselves occupy a zone that is large enough to comfortably hold their bodies and accommodate the required distance between them. This has been termed the *p-zone* (participant zone). Further out, behind the participants, lies a region of space approximately double the width of a person. This region has been termed the *r-zone* and serves as a sort of a transition zone between the formation and the rest of the social space.

Formations may need to accommodate the movement patterns of other people in the space by leaving certain passageways open for others to pass through. These are termed *intersected formations*. Sometimes intersections are formed by physical passageways and barriers, such as isles through auditoriums or room dividers. While certain types of formations hold when people are in motion, zones are not preserved.

Physical structures can directly influence gatherings and how they break into formations, and sometimes they are deliberately set up for facilitating such social occasions. The divider

| Concept | Description | Sources |
|---|---|---|
| *social space* | Any space with mutual embodied access | EG |
| *gathering* | Collection of two or more people | EG |
| *social situation* | Space that adds people immediately to a gathering | EG |
| *social occasion* | Reason for a social situation | EG |
| *formation* | Tight cluster of coordinated activity within a gathering | AS, AK |
| *dyad* | Formation of two people | AS |
| *element* | Formation of people with common orientation | AS |
| *F-formation* | Formation of people facing each other | AS, AK |
| *bystander* | Participant in gathering, but not in formation | EG |
| *formation zones* | Space occupied by and surrounding formations | AS, AK |
| *o-zone* | Innermost zone covered by overlapping orientations | AS, AK |
| *p-zone* | Zone occupied by formation participants | AS, AK |
| *r-zone* | Outermost zone serving as formation transition area | AS, AK |
| *intersection* | Passageway or structure splitting formations | AS |

Table 14.1: Some useful concepts that describe the structure of social space and their sources (EG = [Goffman 1963]; AS = [Scheflen 1976]; AK = [Kendon 1990])

and furnishing in Figure 14.1 facilitate two simultaneous gatherings and certain formations within them. A seat is a location for seating that can then be arranged into larger modules that together circumscribe an o-zone for face-formations (such as the chairs in Figure 14.1), or they can arrange into elements of common orientation facing a target, such as a TV (such as the couch in Figure 14.1), stage, or a theatre screen in an auditorium. Furniture can also provide work surfaces that serve as focus points for participants during interaction.

### 14.2.2 The Body in Social Space

By being present in a social space, a person will occupy a certain physical volume. While there is a large variation in body shapes, an approximation to the space occupied by a single person will suffice for most models of social space. One useful approximation of this volume is four cubic *cubits*, as shown in Figure 14.3. The cubit measure was originally defined as the distance from a full-grown person's elbow to the end of their middle finger[2]. The best way to understand this distance is to bring both elbows to your sides and bend them to 90°, sticking your forearms out in front, which then creates a perfect square cubit. While exact cubit lengths have varied slightly through the ages, 20in, or roughly 0.5m, is a good round approximation. Four cubic cubits divide the body into four *body regions*, which correspond to the feet and

[2] Encyclopedia Britannica, https://www.britannica.com/science/cubit. Accessed August 7, 2021.


Figure 14.3: A person occupies a space that is roughly four cubic cubits, dividing the body into four *regions*. To be able to comfortably act, a person will need an area that is roughly $2 \times 2$ square cubits, which corresponds to what is called a *location*. A *cubit* unit is approximately 0.5m long.

lower part of the legs, the upper part of the legs and pelvis, the torso, and finally the head (see Figure 14.3). The exact arrangement of the cubic cubits can vary a bit, for example, between a person standing and sitting down.

Naturally, if the four cubic cubits were the only space a person could peruse in a social gathering, they would feel quite constrained and be unable to perform most actions. By doubling the footprint of this space, however, we get something that corresponds to a person's

*location*, which is a unit or cell that allows us to place a person comfortably in a social environment (see Section 14.2.1 and Figure 14.3).

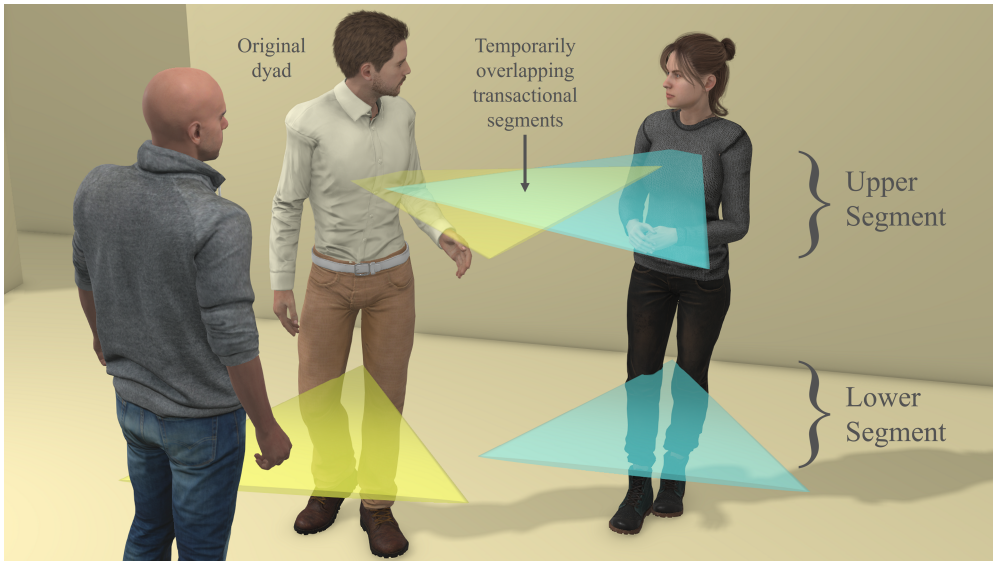



Figure 14.4: Orienting different body regions can claim different spaces, or *segments*, around a person. By combining the orientations of the upper two regions (head and torso), a single upper segment can be commanded, and likewise, by combining the orientations of the lower two regions (legs and pelvis), a lower segment can be claimed. This makes it possible for instance to engage in a side involvement while still committing to the original dyadic formation.

While the physical body is bound to its exact location, the four different body regions also claim space beyond that location through independent outward orientation. That is, by orienting the head, torso, pelvis, and legs/feet, four different *segments* of space further away can potentially be commanded socially by the person (see Figure 14.4). Rather than commanding all four potential segments, it is more common for people to claim only two segments by combining the orientations of the two upper body regions (claiming an upper segment) and the orientations of the two lower body regions (claiming a lower segment). A segment that coincides with an activity sustained by the person, such as watching TV or interacting with someone, is termed a *transactional segment*. This segment is usually respected by others, and typically will not get needlessly invaded. Taken together, the space occupied by the person and their claimed segments is termed their *territory*. The size of a claimed segment will vary by context, but the extent of personal influence is naturally constrained by sensing distances, which serve as a foundation for the interpersonal distances described by the *proxemic* classification system [Hall 1966].

Originally conceived as part of the study of animal behavior and how animals react to the proximity of others, the classification of distances from a human body has resulted in four useful distances: *Intimate*, *personal*, *social*, and *public* (see Figure 14.5). The exact distances are not fixed numbers, but they roughly define spaces around a person that correspond to different interaction opportunities. These are largely influenced, on the one hand, by relatively universal physical limitations of the human species such as visual acuity and hearing, but on the other hand, they are influenced by contextual variables such as demographic and cultural norms.

The distances summarized below represent the results of Hall's study performed on middle-class healthy adults from the northeastern seaboard of the United States [Hall 1966]. Hall cautions that these should not be taken as a representation of human behavior in general [Hall 1966], but it is worth noting the reference to general human capacity for perceiving and acting at different distances.

*Intimate distance* is closer than one cubit, or 0.5m, away (see Figure 14.5). One could therefore say that two people at an intimate distance are more or less sharing the same *location* (see above). At this distance people can easily touch and embrace each other, and whispers can be heard. They have a good, detailed view of each other's face, but the rest of the body is not readily seen and is visually distorted. *Personal distance* corresponds to the comfortable distance one typically maintains from others, which ranges from 0.5m to 1.5m. By placing people at separate locations, for example, with chairs, this distance is ensured. People can easily hear speech at moderate voice levels and they can perceive each other's full bodies without visual distortion. It is possible to reach and grasp, at the closer end of the range, but slightly further out people can barely touch. *Social distance* is kept with those one intends to conduct relatively impersonal interactions with, ranging from about 1.5m to 3.5m. At this distance touch is not possible without effort and high visual detail in faces are not easily perceived, but normal voices can be heard. At the far end of the range, it may become harder to maintain visual contact and disengagements become less of an effort. *Public distance* is anything further than 3.5m and would be considered a relatively safe distance from anyone, for example, a person you would want to avoid. At this distance only exaggerated or amplified voices can be heard, and gestures become more prominent than facial expressions.

Bringing social influence beyond the closer distances people can engage in pointing, also called a deictic, a reference, or simply just a *point*. It doesn't have to be a classic pointing hand gesture, with an extended index finger, it could be something more subtle like a brief flick of the hand toward a distant target. It doesn't even have to be a hand gesture at all, it could also be accomplished with gaze, head movement, or even with a foot. This ability to project a clear, yet invisible, ray into the entire social space can serve many important functions (see Section 14.2.3) and is just another example of how people apply their bodies' flexible articulation to exploit the space around them.

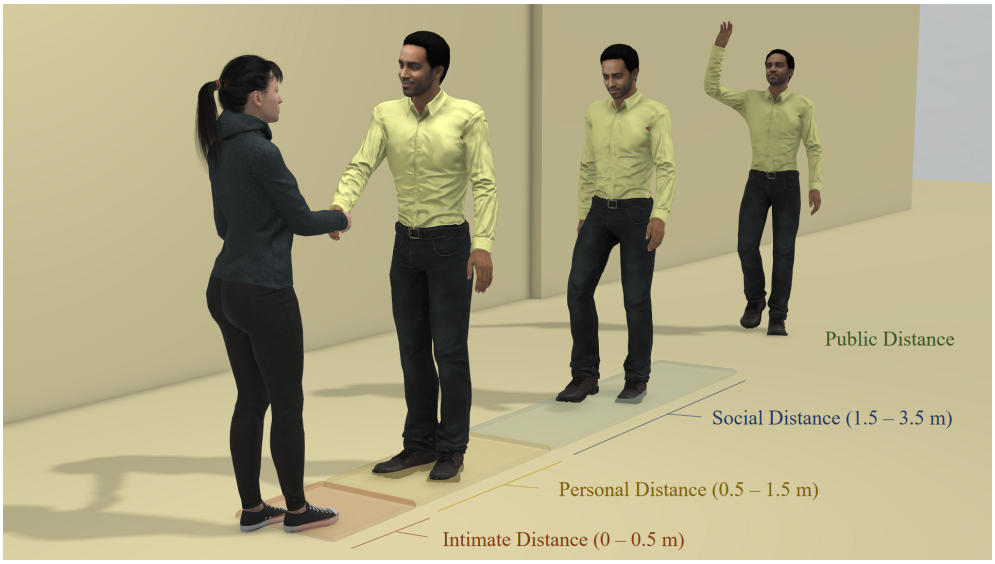



Figure 14.5: Proxemics describe opportunities for interaction at different interpersonal distances based on human sensory capabilities as well as cultural norms. Four ranges are identified by Hall [1966]: *Intimate*, *personal*, *social*, and *public*. As a person approaches, the exact interaction configuration is negotiated through a salutation sequence. During the approach it is important to break mutual gaze in order to minimize perceived threat.

### 14.2.3 Interaction Functions in Social Space

Now that we have an idea of how social space is structured, and how the human body occupies that space, we turn to the interaction that occurs within that space, essentially how people accomplish their social goals.

There is a fundamental principle that underlies this discussion, and that is that *people generally adhere to social norms in social spaces*. People strive to ”fit in,” that is, they do not want to attract undue attention. This requires a balance, where people do not want to thrust themselves on others but they simultaneously do not withdraw themselves from their presence. The interaction functions observed in social spaces emerge from this closed natural system of social order [Goffman 1963]. These functions can be classified as belonging either to *unfocused interaction* or *focused interaction*. We look at each in turn.

***Unfocused Interaction*** Unfocused interaction is what takes place when people are merely managing their co-presence in the social space without meaningful mutual engagement, such as conversations. An example would be a person walking through a crowded mall and casually browsing for good deals at one of the stores. At a very basic level, the person needs to *navigate* the space without causing others distress by physically colliding with them. Where possible,

| Concept | Description | Source |
|---|---|---|
| *cubit* | Elbow to fingertip, approximately 0.5m (20in) | AS |
| *region* | One of four cubic cubit sized portions of the body | AS |
| *location* | Area needed by person for action (approx. $2 \times 2$ cubits) | AS |
| *segment* | Space further out claimed by orientation of a body region | AS |
| *territory* | The space both occupied and claimed by a person | AS |
| *intimate distance* | Closer than one cubit away, sharing location | EH |
| *personal distance* | Comfortable distance, between 1 to 3 cubits away | EH |
| *social distance* | Impersonal interaction distance, between 3 and 7 cubits away | EH |
| *public distance* | Safe distance from anyone, over 7 cubits away | EH |
| *point* | Extending or orienting a body part to call out a target | AS |

Table 14.2: Some useful concepts that describe the spatial structure bodies and their sources (AS = [Scheflen 1975, 1976]; EH = [Hall 1966])

the person should try to stay at least a personal distance away from others, but this can be hard in cramped spaces such as elevators. Navigation will often follow certain patterns to make the job less cognitively demanding, for example, pedestrians tend to route along paths that are already established, such as by keeping to the same side of the walkway or simply by *following* the people ahead of them, sometimes forming *streams* or *lanes* where people string along, one following another. Where bottlenecks occur, people resort to common routing techniques such as *step-and-slide*, where they momentarily turn sideways while stepping past an oncoming person.

But being present in a social space is about much more than establishing a collision-free path. Another very important skill that people need to rely on is to signal an expected level of social awareness and at least a minimal availability to the social setting. If these signals are not sent, the person is likely to stand out in some way, for example as being hostile or judgmental, as being ill or at best lost in their own thoughts (see Figure 14.6a). That kind of attention one should give to others by briefly looking at them, as they pass by, has been termed *civil inattention*. This is just enough eye contact to properly perceive the people around you and acknowledging their presence, but not long enough to constitute an invitation for a longer engagement (see Figure 14.6b). Typically, mutual gaze that lasts for 2s or less, is not interpreted as a meaningful attempt to engage [Goffman 1963].

There are times when people care to only maintain unfocused interaction, thus doing what they can to avoid being brought into greater social involvement. To this end they carry out different strategies. Perhaps the most effective strategy is to remove the opportunity for mutual observation altogether, through what has been termed an *involvement shield*. Such shields can be anything from complete shelters from the rest of the gathering, such as adjacent bathrooms,

(a) Ignoring (b) Attending (c) Inviting

Figure 14.6: A person who does not display signals of minimal attention toward others in a gathering may seem to be deliberately ignoring them (a). A brief sign of attention is civil, without commitment to engagement (b). Basic attention can be turned into an invitation to engage by adding full head orientation, raised eyebrows, and a smile (c).

to covering one's face with a newspaper. Today, cell phones provide convenient involvement shields as they are readily available in most people's pockets and offer a legitimate call for diverted attention. It is important to understand that these shields still need to be socially acceptable by the gathering or a person might risk getting called out for disrupting it or otherwise displaying contempt for it.

Instead of resorting to physical barriers or objects, one can also attempt to maintain a public distance from other people, where possible, and stay out of the r-zones of group formations. Where closer distances are required, full or partial orientation away from others, for example, by taking an outward facing seat, can help cut lines of mutual perception. Gaze can also be cast down on the ground, for example, in an elevator or while being forced to pass within someone's personal or social distance.

The point at which two people mutually react to one another, and possibly recognize each other (e.g., as acquaintances or friends), is a potential first step in a sequence of behaviors that often precede further engagement. If at this point mutual gaze is held for longer than would pass for civil inattention, moving to the next step in the so-called *salutation* sequence is very likely to occur. That step would be the *distance salutation*, where the typical behavior involves tossing the head upward, with the chin pointing toward the other person, while raising eyebrows and smiling (see Figure 14.6c). This could also involve waving (see Figure 14.5). If this salutation is mutual, and no avoidance takes place, the two parties will likely approach each other, with the intent to start further interaction (see Figure 14.5). During the approach, gaze is broken, as it would be considered too threatening or intense to maintain constant gaze throughout. When the people get into either the social zone or personal zone (often depending on the level of formality), they perform a *close salutation*, which grants them an opportunity

| Concept | Description | Source |
|---|---|---|
| *unfocused interaction* | Co-presence without meaningful interaction | EG |
| *focused interaction* | Commitment to social engagement or involvement | EG |
| *navigating* | Moving through space while avoiding collisions | Any |
| *following* | Navigating behind another person that picks the path | Any |
| *streaming* | People following one another along a path | PC |
| *lane formation* | Same as *streaming* | EG |
| *step-and-slide* | Stepping past an oncoming person by turning sideways | PC |
| *civil inattention* | Minimal eye contact, less than 2s, to signal awareness without engagement | EG |
| *involvement shield* | Avoiding social involvement by preventing mutual observation | EG |
| *recognition* | Signals to a person that they have been recognized by another | AK |
| *distance salutation* | Clear invitation to further social engagement from afar | AK |
| *approach* | Navigating toward another person with intent to engage | AK |
| *close salutation* | Greeting at end of approach, e.g., hand shake, establishes formation | AK |
| *commitment* | How interested a person is in interacting with another | AS |

Table 14.3: Some useful concepts that describe interaction functions in social space and their sources (EG = [Goffman 1963]; AS = [Scheflen 1976]; AK = [Kendon 1990, Kendon et al. 1981]; PC = [Collett and Marsh 1981])

to negotiate the spatial arrangement of their subsequent interaction (e.g. keeping them at arm's length in the social zone with a handshake or bringing them into the personal zone with a hug). The interaction has now become a *focused interaction*.

***Focused Interaction*** Focused interaction is what happens when people commit to some form of an extended social engagement or involvement such as a conversation or watching something together. Other chapters in this book dwell on what happens during focused interaction, especially around conversations. Therefore, this section will only briefly mention a few examples where spatial behavior helps in managing co-presence during focused interaction.

Clear visual reference through pointing behavior, as we have seen in other chapters in this book, can provide useful information during conversation, but it can also serve as a way to connect individuals who belong to the same gathering, without being directly linked to the ongoing conversation. Either the pointing behavior binds the people together because they are pointing toward the same target, for example, people gathering to watch something unusual

(a) Low mutual commitment

(b) High mutual commitment

Figure 14.7: Body configurations exhibiting different levels of commitment within a formation

happening across the street, or they could point toward each other to indicate or strengthen relations. That could happen when recognizing a friend in a crowd, or when talking to several people but wanting to demonstrate greater commitment to one of them (see below). The ability to execute pointing of various types and with various targets, both near and far, can therefore carry an important relational function.

The usefulness of dividing the body into separate orientation regions becomes clear when we consider multiple people orienting toward a common point of interest. How the rest of the body is oriented will signal whether the people are experiencing this together or are unaffiliated individuals who merely happen to attend to the same thing. While one region, such as the head, is oriented toward the interest, one or more of the other body regions would exhibit mutual orientation toward those individuals they are "with," but ensure minimal orientation toward others. Through the flexible orientation mechanisms of the body, people can establish and maintain membership in formations and even in formations within other formations (see Figure 14.4). Combined with position, this is how people signal and organize their social *affiliation* and *engagement*, which is essential to managing their participation in a social gathering.

Within established formations, we can talk about different levels of commitment between participants, as exhibited through the configuration of their bodies (see Figure 14.7). When commitment is low, we would see fewer body regions oriented toward the other person and more partial orientations (wide angles). We would see body regions covered (e.g., with crossed arms or legs) and kept still (to minimize noticeable behavior). The individuals would also

attempt to maintain maximum possible distance between them, given the constraints of the formation (see Figure 14.7a). When commitment is high, we would see multiple body regions orienting toward the other, and more full orientations (acute angles). Body regions would be exposed to the partner but possibly barred from others. We could also see shorter distances between the people and even tactile contact (see Figure 14.7b).

## 14.3 History / Overview

We now look at how social space has factored into the evolution of SIAs, and virtual agents in particular, considering how the above theoretical models have been incorporated and implemented so far. The history can be roughly divided into four phases, two of which started before the publication of Cassell et al. [2000b] and two which started later. The spans of years shown indicate the apex of each phase.

### 14.3.1 First Phase: Behavioral Animation (Approx. 1985 - 1995)

Probably the most influential work on modeling groups of virtual agents traversing three-dimensional space, while being under some sort of a social influence, is the *boids* (”bird-oids”) model proposed by Craig Reynolds [1987]. Inspired by particle systems, used to animate complex dynamic phenomena, the work extended individual particles by adding orientation and 3D shapes to them, as well as the ability to react to other surrounding particles. Essentially, this set up some of the basic components of social motion, that is position, orientation, and the ability to define interaction rules. But instead of implementing rules of human social behavior, the ”boids” incorporated behaviors of flocking birds and other animals, which included their desire to stick with the flock while avoiding crashing into its other members. This resulted in three motion rules that would directly modify the velocity of each boid super particle: Avoid imminent collisions with your closest flock mates (static avoidance), match their velocity (predictive avoidance), and attempt to get closer to the center of the flock. The resulting simulation produced animation that most would agree resembled flocking birds fairly well, but more importantly, Reynolds had introduced the powerful concept of *Behavioral Animation*, and rule-based group interaction, to a very large audience in the graphics and animation world.

Bringing a similar set of motion rules to the modeling of pedestrian behavior, Helbing and Molnár [1995] introduced the so-called *Social Force Model*. While the leap from animal behavior models to human models may seem very large, Helbing and Molnár argued that simple or standard situations that are well predictable could in fact depend on fully automatic reaction, suitably modeled as equations of motion. More complex or new situations, however, would still require deliberate action, better described, for example, with probabilistic models. Focusing then on the standard situations, Helbing and Molnár [1995] propose a so-called *social force* that has the power to alter the *preferred velocity* of a pedestrian trying to reach a certain destination. This modifying force is meant to sum up internal motivations of pedestrians to perform certain automated motions in response to their environment. One

such automation rule describes a repulsive territorial effect, citing Scheflen [1976], that ensures that agents keep a certain distance from others (assuming they are strangers). Another automation rule creates an *attractive effect* toward friends or performers, which can form groups (”comparable to molecules”). These two automation rules have the greatest effect for what is perceived in front of the agent, but much less for influences behind the agent. In addition to these two rules, the social force also includes slight random fluctuation to represent variation in behavior.

Simulations that run this model seem to produce a number of behaviors that can be observed in human behavior data, including *lane formation*. By focusing on the reactive aspect of pedestrian motion, complementing more deliberate motion planning, this approach brought Reynolds' animal-focused models, and more generally Behavioral Animation, squarely into the realm of human social behavior modeling.

As if the challenge of modeling human velocity was not enough, animating the walking motion, while also interacting with the environment through gaze and gesture, was a tremendous challenge for early social agent researchers. Putting this all together into a comprehensive and complex virtual human software framework called *Jack*, Badler [1997] describes a two-level architecture where one level optimized reactivity to the environment and another that executed scripts or planned complex tasks [Badler 1997]. The reactive level could establish and maintain a seamless and almost tangible bond between the virtual person and the space that they occupied, for example, by placing the feet correctly on the ground during locomotion, gazing toward sensed people and objects according to an attention model, and grasping objects with the appropriate hand shape. These are all things that require little conscious planning, but without them the body would literally lose its grip on the environment. Jack was able to perform many of these movements using *inverse kinematics*, a mathematical method for configuring a chain of joints so that it can reach a given point in space [Zhao and Badler 1994], a crucial component for bringing behavioral animation to fully articulated human figures.

While the Jack system ran one of the earliest embodied conversational agent demos [Cassell et al. 1994], showcasing two virtual agents engaged in fully automated face-to-face conversation, it was also capable of supporting interactions among multiple virtual agents in a furnished virtual house, as demonstrated in *JackMOO* [Badler et al. 2000, Shi et al. 1999]. Here Jack models were in fact avatars of remote human users that could command their virtual avatar bodies around the house using natural language instructions. The Jack agent's awareness of the environment made it capable of breaking high-level commands down into meaningful spatial actions, such as walking up to another avatar before speaking to it. It was also capable of picking appropriate behavior realization, such as the style of non-verbal greeting, based on social context. Here the social smarts mainly arose from plan execution of deliberate actions and sub-actions, whereas the reactive level did not seem to be carrying out any social motion or orientation based on a theoretical social model.

Similar to JackMOO, BodyChat [Vilhjalmsson 1996, Vilhjálmsson and Cassell 1998] was a shared virtual environment where users were represented by smart avatars that automated some of the non-verbal cues of social co-presence and interaction in social space. Unlike JackMOO, BodyChat focused completely on reactive behavior, triggered autonomously by simple socially related sensory events such as other avatars entering or leaving social range, or others signaling interest in further engagement. BodyChat essentially took several existing social theories and implemented them in specific behaviors and rules that triggered them. Since the primary goal with BodyChat was to facilitate managing co-presence and establishing contact between users, these theoretical models included *civil inattention* [Goffman 1963], *openness to engagement* [Cary 1978, Scheflen 1976], and a model of human greetings, which included *distance salutations* and *close salutations* [Kendon 1990]. It may have been somewhat controversial to let the avatars generate social signals autonomously while their users drove them around the virtual environment, but since high-level decisions were still under users' control, such as whether they were open to social contact, the automation ended up positively affecting the social experience [Cassell and Vilhjalmsson 1999].

### 14.3.2 Second Phase: Embodied Conversation (Approx. 1995 - 2005)

Around this time, some of the first successful Embodied Conversational Agents were being born. They generally delved deep into the rich coffers of existing social behavior models to endow fully autonomous virtual agents with social face-to-face interaction skills. However, most of these agents were designed and implemented for a relatively simple configuration of the social space, where the user and agent were in a dyadic formation with no one else around. By placing the agent on a stationary display, the agent would not be expected to move, and thus it was typically the user's job to approach and engage the agent [Cassell et al. 2000b], if they were not already in conversation from the start. Representing, navigating and exploiting the social space was therefore secondary for most of these systems.

Even though they could not move around, limited spatial awareness was built into some of the ECAs, allowing them to detect the presence of a single physical person at the appropriate social range. Sometimes they would also detect whether that person oriented and/or gazed toward the agent [Cassell et al. 1999, Thorisson 1997]. A combination of distance and orientation would then trigger an engagement with the agent, who would then proceed to greet and have a conversation with the person. The remainder of the interaction would then take place at a single, usually well defined, location - in a fixed formation.

The most notable exception to this among the early ECAs was the STEVE agent [Rickel and Johnson 1999, 2000], whom the user joined aboard a virtual naval vessel using a head-mounted virtual reality display. STEVE was a tutoring agent capable of instructing the user on how to use the equipment on board, both through demonstration and by providing feedback. STEVE possessed dyadic conversation skills and could also guide the user around the vessel by navigating its passageways. But a particularly advanced spatial behavior involved STEVE's

ability to command two separate spatial segments at once, by deliberately orienting the two lower body regions toward the equipment, to claim a task-related segment, while orienting the two upper regions toward the user at appropriate times during the interaction to maintain social engagement. Claiming the task segment, or *transactional segment*, provided the user with a strong suggestion for mutual orientation toward a common point of interest. It also ensured unhindered access to the equipment being operated on. Had other people been there, they would not have crossed the space between STEVE and the equipment due to this strong social signal.

Another spatial behavior exhibited by some early ECAs was to refer to shared objects in the environment through deictic gestures or pointing [André et al. 2000, Cassell et al. 2000a, Lester et al. 1999, Thorisson 1997]. The generation of co-verbal pointing gestures is well covered elsewhere in this handbook, but taking pointing further, the Cosmo agent was also capable of navigating closer to the object being referenced to avoid potential referential ambiguity [Lester et al. 1999, 2000]. Unlike STEVE, Cosmo did not occupy the same 3D space as the user, and thus the more nuanced management of co-presence was not required. In some sense, the culmination of the deictic demonstration agent occurred with the whole-body motion planning and synthesis framework of Huang and Kallmann [2016]. Their framework used data extracted from human experiments to generate lifelike locomotion, body positioning, and demonstration actions for arbitrarily located objects and observers, taking into account obstacles and visual occluders in the space. However, little interaction took place beyond the demonstration itself.

Arguably, one of the most spatially aware of the classic ECAs was MAX [Kopp and Wachsmuth 2004], which was built to engage with a human in a face-to-face collaborative activity. Even though the collaboration took place in a relatively static spatial configuration, standing around a fixed worktable, the upper body had to be carefully and continuously managed [Nguyen and Wachsmuth 2011]. By analyzing its own reaching space, or *peripersonal space*, and projecting it onto its partner, it could model an *interaction space*, which essentially covers the space reachable by both agent and human. This space sits in the intimate to personal range, according to the proxemics theory [Hall 1966], and extends the visual attention space, covered by the o-space of an F-formation [Kendon 1990] - thus refining existing spatial social theories with a model of physical collaboration space [Nguyen and Wachsmuth 2011].

### 14.3.3 Third Phase: Gatherings and Groups (Approx. 2005 - 2010)

Having conversations with agents in larger and more complex spaces got pushed by 3D applications such as social training environments and games, further propelled by rapid 3D game engine advances. The Tactical Language and Culture Training System, built using the Unreal Engine, was one such system that simulated various social scenarios involving multiple virtual agents and an avatar controlled by a user [Johnson et al. 2004]. All of them could freely move around the virtual environments. A modular social component attached to each

virtual agent, called *Social Puppet* [Vilhjalmsson et al. 2007], would use a set of rules to generate appropriate reactive social behavior as they approached each other and the avatar, allowing them to engage each other naturally in conversation. During conversations, the social component would also animate a range of conversational behavior synchronized with speech. However, the agents were not able to establish and maintain dynamic group formations. Once a certain proximity to a single target agent or avatar was reached, they would simply stay in place and get ready for conversation. This was a limitation shared by many game-based applications.

Scaling up the complexity of social scenes, the pedestrian simulation by Shao and Terzopoulos [2005] was capable of filling a replica of New York's Penn Station with hundreds of people going about their everyday commuter business, while ensuring a certain level of behavioral realism. For the most part, this realism arose from reactive behavior rules, extending earlier steering frameworks [Helbing and Molnár 1995, Reynolds 1987, 1999], that truly exhibited awareness, and to some extent prediction, of the behavior of surrounding pedestrians across a number of different environmental configurations such as wide-open spaces and narrow stairs. This reactive local navigation control was coupled with higher-level cognitive control that provided the pedestrians both with navigational goals as well as non-navigational activities such as sitting to rest, watching performers, chatting with friends, and queuing up at vending and ticketing machines.

This work successfully demonstrated an impressive crowd in a complex environment, but on closer inspection, in spite of being aware of each other, individuals exchanged no reactive social signals. They either hurried through "lost in their own thoughts" or executed specific behavior routines when engaged in an activity. The models of motion, while impressive in dealing with complex navigation, were not drawn from any social theories describing the intricacies of social exchange. Yet, the sheer spectacle and premise of the simulation inspired a wide range of crowd simulations that followed, many of which started relying more on social theories to improve upon the social realism.

The HiDAC (High-Density Autonomous Crowds) system [Pelechano et al. 2007] took further steps toward social realism by deepening the model of the individual virtual agent, including a description of personality and mood. Different kinds of rules, including psychological, physiological and geometrical, would then influence the final physical forces acting upon the individually simulated bodies. Finally, the rules were context dependent, that is, they would only get applied during situations where they were relevant, such as dropping politeness when a person would enter a state of panic and try to evacuate a place as fast as possible. This overall approach lent itself well to controlling greater behavioral detail, especially relevant to creating believable first-person, or egocentric, experiences as opposed to only focusing on achieving reasonable overall crowd movement [Pelechano Gómez et al. 2007].

Creating dynamic conversation formations of virtual agents that would accommodate participants joining and leaving was a relatively clear need in interactive social virtual environ-

ments, which both Jan and Traum [2007] and Pedica and Vilhjálmsson [2008] addressed in their work around the same time. The former built on a multimodal turn-taking simulation of a small group discussion [Padilha and Carletta 2002], which was applied to a group of background characters in a virtual training scenario [Jan and Traum 2005]. Encouraged by the increased believability achieved by simulating these conversations in the background and making them visible through appropriate non-verbal behavior, their next step was to simulate dynamic positioning and orientation for these characters based on who participated in each conversation [Jan and Traum 2007]. This work incorporated the distances from proxemics and proposes a specific Social Force model that builds up a motivation to move by adding up an attractive force toward the speaker, a repelling force from outside noise, a repelling force from getting too close to other agents, and a force toward a circular formation for the group in the social range [Jan and Traum 2007].

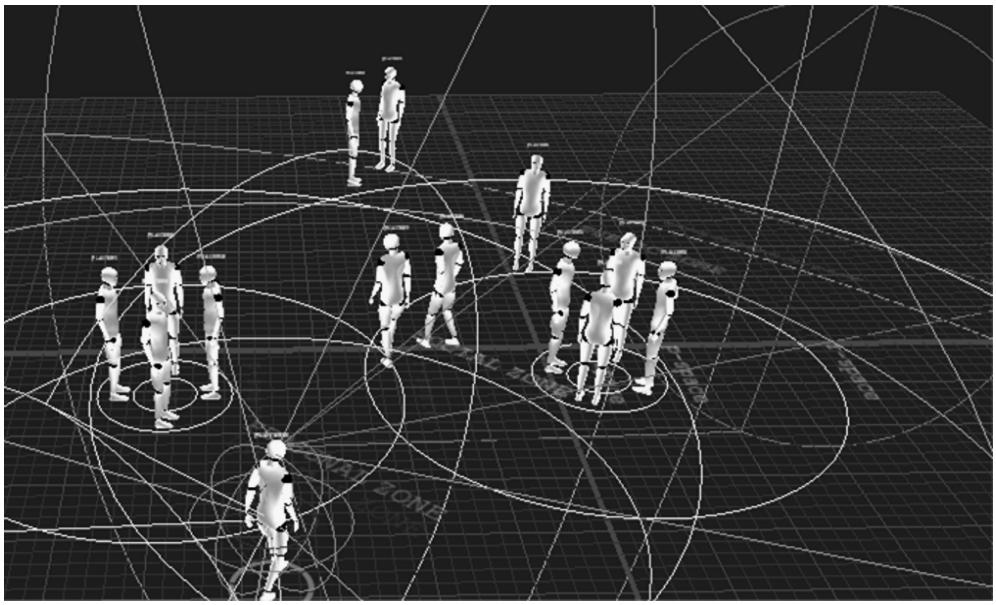


Figure 14.8: CADIA Populus was a modular and dynamic social simulation framework that extended the social force model through theories of human territoriality and proxemics.

A similar idea grew out of the Social Puppet work that extended the foundation of the Social Force model [Helbing and Molnár 1995] through theories of human territoriality [Scheflen 1976], as well as proxemics [Hall 1966], to arrive at a modular and dynamic social simulation framework that included interaction with a player-controlled *avatar* [Pedica and Vilhjálmsson 2008, 2012, Pedica et al. 2010].

The original implementation of this framework, called *Populus* (see Figure 14.8), gave every character, including the player's avatar, perceptual capabilities through an easily extend-

able sensory system. In addition to basic senses such vision and hearing, a higher level ”social perception” was implemented, which would generate sensory events when others would cross interpersonal distance boundaries within the field of view or get within personal distance regardless of angle [Pedica et al. 2010]. Another important feature that set this work apart from its predecessors was the modeling of multiple body regions, making it possible to let motivating forces adjust the orientation of eyes, head, and torso separately from direction of locomotion. This provided a much finer control over the social body and the management of affiliation and engagement within the social space.

Motion rules could be tied to various conditions, including sensory events. Populus focused first on steering rules that would establish and maintain an F-formation. More specifically, all participants in a conversation group performed a series of reactive positioning and re-orienting behaviors that were aimed at defending the group's o-space. In the case of a disruption, for example, another participant joining the group or even just passing close enough to trigger a brief defense of personal space, a series of compensating movements by group members would help reach a stable formation again. Thus an equilibrium was maintained within the F-formation.

Another important social concept embraced by Populus was the concept of a *frame*, which is the set of behavior rules and social norms that participants engaged in interaction, such as a group conversation, silently accept [Goffman 1974]. By making the frame and its rules computationally explicit, specific behavior rules could be triggered based on what interaction was taking place. This led to a series of Populus extensions that dealt with particular social scenarios, such as non-focused interaction at bus stops and sidewalks [Cafaro et al. 2009], business transactions affected by social interruptions [Thrainsson et al. 2011], finding and joining tables at a restaurant [Carstensdottir et al. 2011], and social navigation through tight spaces [Oliva and Vilhjálmsson 2014]. A number of additional unpublished scenarios were built and explored by students at the Center for Analysis and Design of Intelligent Agents (CADIA) at Reykjavik University, where the framework was used for both teaching and research.

Populus, which was scripted in Python on top of the Panda 3D game platform[3], was later re-designed and re-implemented as the *Impulsion* plug-in for the Unity 3D® game engine, where behaviors were implemented using Behavior Trees [Pedica and Vilhjálmsson 2012]. This integration with a major game engine facilitated migration into several larger scale projects, including the Virtual Reykjavik language training platform [Vilhjalmsson et al. 2014], where additional attention was given to behaviors around starting a conversation with strangers [Ólafsson et al. 2015]. It also made an integration possible with one of the major Embodied Conversational Agent platforms, VIB (Virtual Interactive Behavior System), also formerly known as Greta [Niewiadomski et al. 2009], creating social agents that could both

[3] See https://www.panda3d.org/

manage their co-presence in social space and generate expressive co-verbal behavior during conversations [Cafaro et al. 2016].

Other work has also modeled personal space mathematically to calculate optimal placement of simulated people with respect to one another in a social gathering, arriving at a natural equilibrium in standing formations. In Laga and Amaoka [2009], the mathematical model not only considers the distance but also the facing direction by shaping the personal space as an oval stretching out in front of each person. The work of Karimaghalou et al. [2014] further extends the updated social force model from Pedica and Vilhjálmsson [2008] by activating the repulsive force before personal distance is violated (they term this a *predictive* repulsive force), in order to avoid overshooting, which could result in unnatural oscillation. In addition, this work added a model of interest, which drove participants to leave and join groups dynamically, creating a relatively dynamic social gathering [Karimaghalou et al. 2014].

Also concerned with generating believable interactions between people in a gathering, the work of Sun et al. [2012] introduces a three-step model: triggering, initiating and animating. The triggering step relied on several contextual parameters and designer input to produce a reasonable number of conversations. The initiating step was responsible for bringing the two participants into a chosen formation, taking proxemics into account, and starting a parameterized conversation of a chosen type (several *dyadic conversation archetypes* were proposed). These were selected and adjusted based on agent attributes as well as environmental context. Finally, a Behavior Tree was used to animate the conversation itself, from start to finish, taking into account the supplied parameters.

While these works helped make a gathering look more social by generating engagement, they mostly focused on social interaction while the virtual agents were standing or sitting still. One of the greatest challenges in simulating behavior in social space is to model social interaction while moving through that space, for example, between couples in a moving dyadic formation.

Bringing social interaction into motion, the seminal work of Peters and Ennis [2009] used a video corpus to analyze both standing and walking groups of pedestrians in an urban campus, to be simulated by their Metropolis visualization engine. As the behavior differed between areas, a specialized tool MetroPed was developed to place individuals and groups into different behavior zones, such as open areas and corridors for walking. Using formation templates, based on observed walking formation forms, and cohesion scores between individual members within formations, the simulation was capable of generating relatively realistic looking groups of people walking together while adapting to path width and oncoming pedestrian traffic. The authors claim that the key to advancing social simulations, such as this one, is the iterative process of corpus analysis, model enhancement, and evaluation. In particular, *perceptual evaluation*, where independent viewers judge the realism of the results, is proposed as an important evaluation method.

The work of Popelová et al. [2011] extends classic steering frameworks, and the frequently used *leader-follower* steering paradigm [Reynolds 1999], with steering rules and parameters that define a steering partnership between two agents aiming for the same destination. With special behaviors for *giving way* and *waiting for partner*, these rules applied to a number of pedestrian scenarios produced higher social believability scores in a perceptual evaluation than a baseline leader-follower model [Popelová et al. 2011].

One of the most elaborate frameworks for simulating social groups of pedestrians traveling together was originally presented in Rojas and Yang [2013] and further elaborated on in Rojas et al. [2016, 2014]. In this work, pedestrian agents can be assigned to slots attached to an invisible group agent that takes care of global navigation. Slot positions are defined by formation templates, which include ”abreast,” ”u-shape,” and ”river” while moving and a ”huddle” when stopping (essentially an F-formation). These slot positions represent desired locations for the pedestrian agents, but actual success at staying in formation may be affected by other factors such as individual collision avoidance. Furthermore, the choice of formation template changes dynamically based on available space. Finally, perhaps the most interesting aspect of the framework is the notion of subgroups within the group. These can be created by ”locking” slots together so that they stay side-by-side, even when the overall formation folds, for example, from a wider formation to a narrower one. Pedestrians in such locked subgroups can hold hands or place hands on each other’s back. They will also occasionally gaze toward each other, further indicating social engagement [Rojas et al. 2016].

Similarly, Ren et al. [2017] extend the well-known Reciprocal Velocity Obstacle algorithm for stable and collision-free local navigation [van den Berg et al. 2008] with the notion of *connection-agents*, a subset of neighbors to stay close to. By adjusting a few parameters, such as the size of such subsets and upper bounds of distances between members of subsets, a number of emergent social grouping behaviors have been observed, ranging from couples walking together to larger groups following a leader [Ren et al. 2017]. While an important step for turning collision avoidance into something more socially believable, this work did not attempt to model any finer aspects of social spatial interaction such as upper body orientation or gaze.

### 14.3.4 Fourth Phase: Immersive Studies (Approx. 2010 - 2020)

Virtual Reality lets human users fully enter the social space of the virtual agents, opening the door for studying and exploiting their social co-presence in ways that were not possible within desktop simulations and 3D games. While virtual agents, such as STEVE [Rickel and Johnson 1999], already existed in VR at the turn of the century, expensive equipment and technical challenges limited such research to a handful of labs. Two of the most influential ones, the Virtual Environments and Computer Graphics lab at University College London and the Research Center for Virtual Environments and Behavior (ReCVEB) at University of Santa Barbara, provided early evidence that the shared virtual space obeyed the same

theoretical social principles as physical space. For instance, people would stay at least an intimate distance away from virtual agents while getting a closer look at them - and further away if the virtual agents started at them [Bailenson et al. 2003]. And even when conscious of the artificiality of the virtual agents, unconscious reactions would indicate respect for social norms and be in line with scores for social anxiety [Garau et al. 2005]. These, and similar, results helped propel the idea that significant social interaction with virtual agents could be had in VR, potentially benefiting applications ranging from anxiety treatment to social skill training. When a new generation of high fidelity, but low cost, virtual reality equipment emerged a decade later, virtual agents and their human counterparts flocked into the shared social space, leading to a number of new insights.

Studies of co-presence in VR tend to focus on two explicit aspects of social behavior: proxemics and gaze patterns. The two are quite related, as for instance indicated by the need to break mutual gaze during an approach to avoid appearing too aggressive (see Figure 14.5). Supporting the earlier research on the measurable human subjective and behavioral response to the proximity of virtual agents in VR [Bailenson et al. 2003, Garau et al. 2005], a physiological response has also been confirmed [Llobera et al. 2010]. Individual agents or groups of four agents, would approach the subject and stop at intimate, personal, or social distance. Both the number of agents and closer distances would correlate with measurements of higher arousal, based on skin conductance [Llobera et al. 2010]. Surprisingly, the results did not seem to differ much between virtual agents that were human-like in appearance and agents that were simple geometrical cylinders, in part explained by the potential threat of getting hit by one [Llobera et al. 2010].

By letting participants signal with a button press when virtual agents have reached a comfortable and then an uncomfortable distance, the significance of the boundary between the personal distance and social distance has been confirmed in VR [Bönsch et al. 2018]. Discomfort was always reported before the 1.5m outer limit of the personal distance was crossed, with the reported comfort distance placing all of the agents within the social distance range of 1.5m - 3.5m. However, the study also showed that the angle of approach, the emotion displayed, and the number of agents would further modulate the distance. For instance, single agents could come closer when expressing a happy emotion compared to an angry emotion, and approaching from the front would require greater distance than when being approached from the sides [Bönsch et al. 2018]. Perhaps surprisingly, there was some indication that groups of all smiling agents had to be kept further away than even groups of all angry agents [Bönsch et al. 2018], but since mutual gaze was not broken during the approach, the smile may have been interpreted as sign of confident aggression.

Examining a more nuanced interplay between interpersonal distance and gaze, Kolkmeier et al. [2016] conceived a study based on the equilibrium theory [Argyle and Dean 1965], that states that the balance between the two can be used to regulate a perceived level of intimacy. A pilot established that an agent switching between gazing toward and away from the human

at random intervals between 2s and 5s was perceived neutral, while always responding with mutual gaze upon being looked at, and letting it linger for 1.5 s afterwards, was perceived intimate (a constant stare was considered too ”creepy,” as one would expect). Furthermore, an agent staying at 0.75m was perceived as neutral, while stepping closer than 40cm was considered intimate - corresponding well with Hall's personal and intimate distances, respectively. During the full study, these were then manipulated in a group conversation between a human subject and two virtual agents, and the gaze and proxemic response of the human measured. The strongest response was received when an equilibrium state was broken by changing both distance and gaze at the same time, while changing only gaze elicited a weaker response than changing only the distance [Kolkmeier et al. 2016].

Bringing in a cultural dimension, Obaid et al. [2012] devised a pilot study where human subjects, belonging to either a high contact (Arabic) or a low contact (German) culture, would join four couples of virtual agents. Two couples would exhibit the gaze and proxemic behavior that correlated with the Arabic and German culture, respectively, but two couples would adopt inconsistent gaze and proxemic behavior. While more data is needed, tendencies in heart rate indicated that the subjects, irrespective of culture, remained more relaxed when the gaze and proxemic behavior of the virtual agents were consistent [Obaid et al. 2012].

Studying immersive interaction with larger gatherings of virtual agents in VR has been technically difficult because of the heavy computational demands, both from simulating a larger number of agents and from rendering the scene for each eye at high frame rates for the Head-Mounted Display. However, better hardware has made this attainable in recent years and several platforms for crowd interaction in VR have emerged.

Crowd simulations that support a first-person immersive experience need to support a relatively convincing animation and rendering of the virtual agents when viewed up-close. The PedVR platform [Narang et al. 2016] addresses this by combining classic 2D global and local navigation approaches, including RVO [van den Berg et al. 2008] and the social force model [Karimaghalou et al. 2014], with the SmartBody character animation system [Shapiro 2011], which is capable of both synthesizing realistic locomotion behavior and fine-grained and synchronized non-verbal behavior such as gesture, gaze, and facial expressions. By using a Behavioral Finite State Machine for directing the agent behavior, a number of different kinds of interactive social gatherings could be produced and tested. The results of a study that compared PedVR with and without gaze behavior, that reacted to the presence of a human subject, suggest that such behavior has substantial impact on participants' sense of social presence [Narang et al. 2016]. Anecdotal evidence, including subjects that apologized to gazing virtual agents they collided with, supports the existing theories about the strength of social norms that govern social space and the non-verbal cues that signal adherence to them. In perfect line with these results Kyriakou et al. [2017] reported that subjects navigating through an outdoor mall with oncoming pedestrians would report the highest level of presence and realism when the pedestrians exhibited basic social behavior such as verbal salutations,

gaze, and other gestures toward the subjects. And similarly, the subjects would sometimes feel compelled to return salutations, even though they were in the middle of a difficult task chasing a target [Kyriakou et al. 2017].

Another VR crowd platform suitable for immersive studies is the F2FCrowds system that incorporates a novel navigation algorithm, Interaction Velocity Prediction, which predicts whether the avatar of a human user is trying to approach a virtual agent for face-to-face interaction [Randhavane et al. 2017]. When such an approach is detected at a public distance, a virtual agent will slow down and gaze at the human and help close the distance until within social range, at which point it will stop and attempt communication. During communication, the virtual agent can exhibit listening head movements but will not engage verbally. After the human starts diverting attention, the agent breaks away again to end the engagement [Randhavane et al. 2017]. A study that compared this full implementation to a baseline PedVR approach without gaze [Narang et al. 2016] and to a version without head movement, found it to significantly increase the sense of social presence as well as elicit a stronger reaction from the human users [Randhavane et al. 2017]. It should be noted that users moved around the immersive environment using a joystick rather than literally walking.

A similar system for engaging virtual agents in a crowd, but using room-scale VR to let users walk around a small segment of a virtual mall, was implemented for the social skill training game SoCueVR [Thordarson and Vilhjálmsson 2019]. Here, a human player was tasked with approaching strangers that are passing by to collect money for charity. After the player makes eye contact with someone, the virtual agent would either exhibit inviting behavior (see Figure 14.6c) or appear annoyed and attempt to ignore the user (see Figure 14.6a). At this point, the user could address the virtual agent verbally, such as by saying ”Excuse me” to initiate contact [Ólafsson et al. 2015], or look for someone else. By correctly identifying those that exhibit inviting behavior, the player could raise more money over the duration of the game instead of wasting precious time interacting with agents who would rather be on their way to somewhere else. Initial usability testing showed that this natural interaction paradigm, rooted in the social theories discussed in this chapter, worked very well for the game, leading to targeted and smooth social engagement with the virtual agents [Thordarson and Vilhjálmsson 2019].

## 14.4 Similarities and Differences in IVAs and SRs

With regard to social space, the largest difference between virtual agents and social robots is that the latter exist in the same natural social space as humans, by default. They therefore need to deal with managing their co-presence with real human bodies, and each other, from the moment they are switched on. In the early days of robotic research and development, designers were mainly concerned with ensuring human safety around robots going about their robotic tasks. This led to a number of solutions for operating in and safely navigating through complex environments, including advances in visual perception and obstacle avoid-

ance techniques. However, social robots actually need to make contact with humans, rather than avoiding them. The seminal work of Satake et al. [2009] clearly demonstrated the value of a behavior model that followed social norms and theories of human social space. Robots that simply went directly up to people and started speaking would frequently fail to initiate an interaction, whereas a more nuanced negotiation of mutual engagement across the different interpersonal distances (as in Figure 14.5) would prove much more successful [Satake et al. 2009]. Following this work, a number of similar approaches have been studied and the importance of applying social theories to social robotic research has only grown [Charalampous et al. 2017, Mavrogiannis et al. 2021, Rios-Martinez et al. 2015], however, going deeper into those methods is outside of the scope of this chapter.

Compared to robots, it is not as straightforward for virtual agents to share physical social space with humans. Early agents would either exist completely within a virtual environment, with limited human interaction (e.g., in crowd simulation), or they would exist on a fixed boundary between the physical and virtual social space. That boundary would then have been framed by a large monitor or a projection surface. One could argue that such a setup could mimic a stationary social robot, but the lack of physical presence may become a factor that is hard to overcome in the general sense [Li 2015], and also for specific interaction requiring a common frame of spatial reference or manipulation of physical objects, such as discussed in Holthaus and Wachsmuth [2012].

However, with advances in XR technologies, both virtual and augmented reality, humans are increasingly sharing full-scale social space with virtual agents in ways that are closer to the spatial interaction with social robots than before. In virtual reality, the human steps into the world of the agent (see Section 14.3.4), and in augmented reality, the agent steps into the physical world of the human as a graphical overlay. The latter may well generate the appearance of a physically instantiated virtual agent, and the management of social co-presence may look identical for both virtual agent and social robot in the same environment, but any modification of the environment becomes a real challenge for the virtual agent. How would it, for instance, re-arrange a chair to join a sitting formation of humans? A space of creative solutions will need to be explored to fully bring the virtual agents out into our world - could they, for example, be allowed to bring their own virtual chairs?

## 14.5 Current Challenges

### 14.5.1 Formalizing Continuous Spatial Behavior

The Embodied Conversational Behavior community came together and formalized the specification of multimodal communicative behavior as the Behavior Markup Language (BML) [Kopp et al. 2006]. It was relatively challenging work, but largely successful, resulting in a number of BML compatible components that researchers are already sharing, such as SmartBody [Shapiro 2011]. BML works particularly well for precisely specifying co-verbal behavior, where the non-verbal behavior can be scheduled relative to uttered words. BML is

fundamentally broken down into blocks for animated performances, within which everything is synchronized, which works well for conversations organized by dialogue planners. However, specifying how a person traverses social space, in the absence of spoken communication, has never been a strong point for BML.

The behavior that has to be specified in that case is a multimodal behavior that fundamentally involves locomotion. Locomotion is a continuous behavior without a known finishing time, which typically needs to be adjusted in reaction to a changing environment. It cannot therefore be specified as a predetermined performance, belonging to a finite block of behavior. What is instead needed is a set of arguments that can be sent to a locomotion engine that specify its behavior until it is next updated or some criteria is met. One has to think about these commands as a way to configure an engine that is continuously running. This is somewhat in line with more recent SAIBA discussions about the so-called contextual markup [Cafaro et al. 2014], which is meant to set up the environment in which the multimodal behaviors run, influencing things such as their manner of motion.

At the higher level of social function or intent, the Function Markup Language (FML) can serve behavior in general social spaces quite well. For instance, one could specify that a person intends to initiate contact with one person while intending to avoid another person. It is primarily a matter of populating FML with more social concepts and constructs, but as with BML the temporal aspect needs some consideration. FML commands are also gathered into discrete blocks that are sent to BML planners, which in turn are meant to produce corresponding BML blocks that realize those functions. In light of the preceding discussion about setting configuration parameters or a context for a behavioral engine, the FML commands, even if discrete, would end up updating those parameters rather than producing immediate BML. It is therefore quite possible to turn explicit and discrete high-level controls into continuous and dynamic behavior and motion at lower levels if the behavior and motion engines maintain their own state and never stop behaving. We just have to decide what parameters provide the range necessary to accomplish the wide variety of social functions that rely on this behavior.

### 14.5.2 Animating Movement in Tight Spaces

The conventional way to animate character movement across space is to use blend trees where different premade locomotion animation clips are blended together according to dynamic weights. These clips often represent body motion in typical locomotion states, such as for walking, running and turning, and the weights are often tied to forward and angular velocities. When these animations are blended together, characters can be shown smoothly changing speeds and directions, such as when slowing down in front of an obstacle, turning away from it and picking up speed again after avoiding it. However, the blended movement will always retain the general characteristics of the typical animations provided. That works relatively well when the environments can be traversed in more or less the same fashion, for example,

by sometimes walking and sometimes running. However, once the environments become more cluttered with things like furniture and other people, and when every move needs to be considered in a dynamic social context in addition to the physical one, conventional character locomotion techniques often don't cut it.

The solution may both involve carefully crafting the locomotion animations that make up the blend tree, aiming specifically at fine-scale social movement, and to combine that with procedural approaches such as full- or upper-body inverse kinematics that makes it possible to maintain segment orientation and reach for support when needed. What animation clips are needed then to cover locomotion in social situations? Instead of focusing on the faster end of the motion spectrum, such as running animations (primarily useful for games), one needs to add much finer and slower lower body movement, which would be capable of representing adjustments in position and orientation without actually walking. These are essentially shuffling or sliding motions with the feet, along with subtle weight shifts. One should, for example, be able to rotate fully in one spot when a new orientation of the lower segments is required, and shuffle or step to one side when required to make room for someone in a formation without changing orientation.

Regarding maneuvering around and into furniture, it has been suggested that some of the required behavior can be stored with the furniture itself, providing the characters with suggestions for how to orient and place the limbs, either when passing or when using the furniture. As long as these instructions don't interfere too much with other ongoing behavior, this may in fact be a very reasonable approach, as seen, for example, in Veutgen et al. [2018].

### 14.5.3 Mixing Theoretical Models

Because human behavior is so complex, modeling and simulating it has required a divide-and-conquer approach. However, once divided, the question remains how the different models, and implementations of them, can be brought back together to form a well-rounded social agent. This is particularly challenging when each realization of a model chooses a fundamentally different technical paradigm, for example, for representing the environment and for producing the movement within it. This "fragmentation" may end up slowing progress [Diamanti and Vilhjalmsson 2021]. The SAIBA effort has attempted to address this by defining standard interfaces between stages of multimodal communicative behavior planning [Kopp et al. 2006]. That way, as long as the inputs and outputs of a model remain standard, what happens within it is of no concern. For example, one could construct a model that turns a standard representation of a particular intent, such as "take turn," into a standard specification of the behavior that is most likely to be effective in carrying out that intent in a particular situation, such as "raising arms." How it came to that decision does not matter to other models, for example, the model that produces co-verbal gesture.

This is a good start, but currently it deals only with a particular limited portion of the overall social behavior, and perhaps of more fundamental importance, it does not explain

how to deal with the situation where multiple models produce outputs that collide in some way, for example, all wanting to instruct the arms to perform an immediate motion. One approach to deal with this is to blend the outputs together, to essentially produce a new behavior that takes both inputs into account. This often turns out relatively fine for locomotion behavior, where velocity vectors from two different models can be blended using weights. For example, a model that suggests a path toward a destination and another that suggests how to maintain suitable distance from other people can easily be combined in this way, where the final velocity would normally be fully weighted toward the destination movement, but getting too close to people would temporarily push the direction of movement away from the impending collision. However, this method would not work for combining something like different suggestions for what to look at.

For example, if one model suggests a brief glance at another person as part of displaying expected civil inattention, and another model suggests avoiding any mutual eye contact by picking a cell phone as the gaze target, to produce a shield. In this case, blending the gaze directions would result in a gaze that was neither here nor there.

Instead, a couple of arbitration techniques could be applied. One technique would attempt to schedule both but ensure only one is running at any given time. Here it is important to understand what behavior could possibly be sliced up to allow for the other to run. Slicing a short glance would not make sense, but staring at the cell phone could potentially be interrupted to produce a glance. Another arbitration technique would be to pick only one model and disregard the result of the other. This could be done based on assigned priorities and/or on a set of rules, taking into account the current context. These techniques could be a reasonable place to start, but there is no simple magic solution.

## 14.6 Future Directions

### 14.6.1 Creating Accessible Social Behavior Toolkits

The realism of human rendering has been rapidly advancing to the point where unique real-time photorealistic virtual humans can be created in a matter of minutes with free software today. Not only do these bodies appear human-like down to the pores in the skin and strands of eyebrow hair, but they are rigged for expressive animation, with muscles that wrinkle the skin and eyes that twinkle. Graphical shader programming underlies these impressive advances, where the interaction of light and surface gets modeled in ingenuous ways, imitating the natural processes that make our own world real to us.

But we need these bodies to interact with more than light. While we can play realistic recorded, or even live motion capture, animations on them, sticking them into a dynamic social space quickly reveals how utterly unprepared these human facsimiles are for social co-presence. Their bodies do not interact, out-of-the-box, with anything in the social space, and yet, as we have seen in this chapter, a body in social space responds continuously to the social stimuli of its environment.

As with light making a surface visible by interacting with it, we need the social space to produce at least a minimum appearance of social awareness for us to believe the person is present. We need something akin to a social shader programming language, where basic models of social awareness can bring the person into the space. Instead of referring to the elements and parameters of illumination, it would refer to the elements and parameters of the social context. Handing these behavior shaders out with the 3D models, in a form that require minimal setup, will deliver something closer to what the pre-rendered videos are hinting at: Life itself.

### 14.6.2 Adding Social Signals to Autonomous Vehicles

With more and more autonomous physical actors entering public human space, including autonomous vehicles, we need to ensure that their intentions are clearly communicated to the people around them. Humans are very good at reading into even the most subtle social cues and are quick to adjust their own behavior in response. This can happen without any social involvement - we are merely coordinating our co-presence.

One example of such spontaneous social interaction that impacts our behavior in public, is the brief exchange that occurs between a pedestrian approaching a crosswalk and the driver of a car that is also approaching. The pedestrian will instinctively look at the driver before entering the street, even though pedestrians have the right of way. In a matter of milliseconds the pedestrian will assess whether the driver is aware of them or not. Mutual eye contact, optionally with a quick head nod or a flick of a hand from the driver, will make the pedestrian more confident about crossing the street. This activity is in a sense a cooperative activity that becomes more effectively completed with communication. The lack of any reaction from the driver will create doubt about being noticed and stop the pedestrian.

Fully autonomous vehicles that drive without any human passengers will likely cause pedestrians to hesitate more around crosswalks, which may not necessarily be a bad thing, only less efficient. The real problem may arise when autonomous vehicles carry humans, the kind of entities that pedestrians are used to coordinate with. In this case, ensuring that the humans in the vehicle are aware of the pedestrian, even humans sitting in the driver seat may in fact send the wrong signal. Being seen by the human, does not mean the vehicle is at all aware of the pedestrian.

It is likely that we will adapt our pedestrian behavior to this eventual new reality and stop trusting social coordination in public places, inhabited by autonomous vehicles and bots. To err on the right side, we will give way to the machines, regardless of their capability for averting accidents. When awareness and intent is not communicated, there is no coordination.

What if we were to give autonomous vehicles, such as cars, a very clear external indication of awareness and the ability to manage their co-presence? What if the cues they would give off could be immediately recognizable as social? Whether we accomplish this by sticking animatronics into the driver seat, puppeteered by the vehicle, mount large articulated eyes on

the front grille, or just rely on flashing headlights, the bottom line is to manage the social space on human terms - as long as it is ours to occupy.

### 14.6.3 Increasing Multisensory Fidelity in XR

Rapid advances in XR technology are bringing virtual social agents face-to-face with people, full-scale and in high visual fidelity. However, occupying spaces with physical people brings us more than just the visual effect, it is a full multimodal experience that engages all of our senses. As a crowd gets denser, we start rubbing shoulders with those we pass, the air gets stuffier, and the sounds of chatter and shuffling feet get louder. It is well known that the more senses we can engage in VR, the stronger the presence in that environment. A classic use of social agents in VR is for treating anxieties and fears related to social situations, such as agoraphobia and enochlophobia. If we were able to reproduce some of the tactile, aural and perhaps olfactory dimensions of the crowd experience, such treatments could become even more powerful. Some work is being done in this area, and bringing it squarely into the domain of tight social spaces will open a new world of possibilities.

## 14.7 Summary

We have brought together some of the fundamental concepts and models for describing social space, and the social body within that space, drawn from the combined works of some of the pioneers in the study of public social behavior. At a higher level we talked about social functions that help manage and maintain co-presence through social order. We expect everyone to more or less conform to certain social norms, for everything to progress without incidents and embarrassing moments. Social agents that enter these spaces risk upsetting this balance with their lack of social finesse. We therefore should strive to give them the necessary awareness of the social space for them to meaningfully react and engage.

# Bibliography


E. André, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. 2000. The automated design of believable dialogues for animated presentation teams. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, eds., *Embodied Conversational Agents*, pp. 220–255. MIT Press, Cambridge, MA.

M. Argyle and J. Dean. 1965. Eye-contact, distance and affiliation. *Sociometry*, 28(3): 289–304. ISSN 0038-0431. DOI: 10.2307/2786027.

N. Badler. 1997. Real-time virtual humans. In *Proceedings The Fifth Pacific Conference on Computer Graphics and Applications*, pp. 4–13. IEEE. DOI: 10.1109/PCCGA.1997.626166.

N. Badler, R. Bindiganavale, J. Bourne, M. Palmer, J. Shi, and W. Schuler. 2000. A parameterized action representation for virtual human agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, eds., *Embodied Conversational Agents*, pp. 256–284. MIT Press, Cambridge, MA. https://repository.upenn.edu/hms/195.

J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. 2003. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7): 819–833. ISSN 0146-1672. https://doi.org/10.1177/0146167203029007002. DOI: 10.1177/0146167203029007002.

A. Bönsch, S. Radke, H. Overath, L. Asché, J. Ehret, T. Vierjahn, U. Habel, and T. Kuhlen. 2018. Social VR: How personal space is affected by virtual agents' emotions. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 199–206. IEEE. DOI: 10.1109/VR.2018.8446480.

A. Cafaro, R. Gaito, and H. H. Vilhjálmsson. 2009. Animating idle gaze in public places. In Z. Ruttkay, M. Kipp, A. Nijholt, and H. H. Vilhjálmsson, eds., *Intelligent Virtual Agents*, volume 5773 of *Lecture Notes in Computer Science*, pp. 250–256. Springer, Berlin, Heidelberg. ISBN 978-3-642-04380-2. DOI: 10.1007/978-3-642-04380-2_28.

A. Cafaro, H. Vilhjálmsson, T. Bickmore, D. Heylen, and C. Pelachaud. 2014. Representing communicative functions in SAIBA with a unified function markup language. In T. Bickmore, S. Marsella, and C. Sidner, eds., *Intelligent Virtual Agents*, volume 8637 of *Lecture Notes in Computer Science*, pp. 81–94. Springer, Cham. ISBN 978-3-319-09766-4. http://dx.doi.org/10.1007/978-3-319-09767-1_11.

A. Cafaro, B. Ravenet, M. Ochs, H. H. Vilhjálmsson, and C. Pelachaud. 2016. The effects of interpersonal attitude of a group of agents on user's presence and proxemics behavior. *ACM Transactions on Interactive Intelligent Systems*, 6(2): 12:1–12:33. ISSN 2160-6455. https://doi.org/10.1145/2914796. DOI: 10.1145/2914796.

E. Carstensdottir, K. Gudmundsdottir, G. Valgardsson, and H. Vilhjalmsson. 2011. Where to sit? The study and implementation of seat selection in public places. In H. H. Vilhjálmsson, S. Kopp, S. Marsella, and K. R. Thórisson, eds., *Intelligent Virtual Agents*, volume 6895 of *Lecture Notes in Computer Science*, pp. 48–54. Springer, Berlin, Heidelberg. ISBN 978-3-642-23974-8. DOI: 10.1007/978-3-642-23974-8_6.

M. S. Cary. 1978. The role of gaze in the iInitiation of conversation. *Social Psychology*, 41(3): 269–271.

J. Cassell and H. Vilhjalmsson. 1999. Fully embodied conversational avatars: Making communicative behaviors autonomous. *Autonomous Agents and Multi-Agent Systems*, 2(1): 45–64.

J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. 1994. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, pp. 413–420. ACM.

J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjalmsson, and H. Yan. 1999. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 520–527. ACM, Pittsburgh, PA. DOI: https://doi.org/10.1145/302979.303150.

J. Cassell, T. Bickmore, L. Campbell, H. Vilhjalmsson, and H. Yan. 2000a. Human conversation as a system framework: Designing embodied conversational agents. In *Embodied Conversational Agents*, pp. 29–63. MIT Press, Cambridge, MA.

J. Cassell, J. Sullivan, S. Prevost, and E. Churchill. 2000b. *Embodied Conversational Agents*. MIT Press, Cambridge, MA.

K. Charalampous, I. Kostavelis, and A. Gasteratos. 2017. Recent trends in social aware robot navigation: A survey. *Robotics and Autonomous Systems*, 93: 85–104. ISSN 0921-8890. https://www.sciencedirect.com/science/article/pii/S0921889016302287. DOI: 10.1016/j.robot.2017.03.002.

P. Collett and P. Marsh. 1981. Patterns of public cehaviour: Collision avoidance on a pedestrian crossing. In A. Kendon, T. A. Sebeok, and J. Umiker-Sebeok, eds., *Nonverbal Communication, Interaction, and Gesture*, Approaches to Semiotics, pp. 199–217. Mouton Publishers, The Hague. ISBN 90-279-3489-4.

M. Diamanti and H. H. Vilhjalmsson. 2021. Social crowd simulation: The challenge of fragmentation. In *Proceedings of the Workshop on Modeling and Animating Realistic Crowds and Humans (MARCH)*. IEEE, Online. DOI: https://doi.org/10.1109/AIVR52153.2021.00034.

M. Garau, M. Slater, D.-P. Pertaub, and S. Razzaque. 2005. The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 14(1): 104–116. https://doi.org/10.1162/1054746053890242. DOI: 10.1162/1054746053890242.

E. Goffman. 1963. *Behavior in Public Places; Notes on the Social Organization of Gatherings*. The Free Press, New York, NY. ISBN 978-0-02-911940-2.

E. Goffman. 1974. *Frame Analyses: An Essay on the Organization of Experience*. Harvard University Press, Cambridge, MA.

E. T. Hall. 1966. *The Hidden Dimension*. Doubleday, New York, NY.

D. Helbing and P. Molnár. 1995. Social force model for pedestrian dynamics. *Physical Review E*, 51(5): 4282–4286. https://link.aps.org/doi/10.1103/PhysRevE.51.4282. DOI: 10.1103/PhysRevE.51.4282.

P. Holthaus and S. Wachsmuth. 2012. Active peripersonal space for more intuitive HRI. In *Proceedings of the 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, pp. 508–513. DOI: 10.1109/HUMANOIDS.2012.6651567.

Y. Huang and M. Kallmann. 2016. Planning motions and placements for virtual demonstrators. *IEEE Transactions on Visualization and Computer Graphics*, 22(5): 1568–1579. ISSN 1941-0506. DOI: 10.1109/TVCG.2015.2446494.

D. Jan and D. R. Traum. 2005. Dialog simulation for background characters. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist, eds., *Intelligent Virtual Agents*, Lecture Notes in Computer Science, pp. 65–74. Springer, Berlin. ISBN 978-3-540-28739-1. DOI: 10.1007/11550617_6.

D. Jan and D. R. Traum. 2007. Dynamic movement and positioning of embodied agents in multiparty conversations. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 1–3. ACM, New York, NY, USA. ISBN 978-81-904262-7-5. https://doi.org/10.1145/1329125.1329142. DOI: 10.1145/1329125.1329142.

W. L. Johnson, S. Marsella, and H. Vilhjalmsson. 2004. The DARWARS Tactical Language Training System. In *Proceedings of the 26th Intersservice/Industry Training, Simulation, and Education Conference*. SSA.

N. Karimaghalou, U. Bernardet, and S. DiPaola. 2014. A model for social spatial behavior in virtual characters. *Computer Animation and Virtual Worlds*, 25(3-4): 505–517. ISSN 15464261. http://doi.wiley.com/10.1002/cav.1600. DOI: 10.1002/cav.1600.

A. Kendon. 1990. *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Studies in International Sociolinguistics. Cambridge University Press, Cambridge, England. ISBN 0-521-38938-0.

A. Kendon, T. A. Sebeok, and J. Umiker-Sebeok, eds. 1981. *Nonverbal Communication, Interaction, and Gesture*. Approaches to Semiotics. Mouton Publishers, The Hague. ISBN 90-279-3489-4.

J. Kolkmeier, J. Vroon, and D. Heylen. 2016. Interacting with virtual agents in shared space: Single and joint effects of gaze and proxemics. In D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer, and A. Leuski, eds., *Intelligent Virtual Agents*, volume 10011 of *Lecture Notes in Computer Science*, pp. 1–14. Springer, Cham. ISBN 978-3-319-47665-0. DOI: 10.1007/978-3-319-47665-0_1.

S. Kopp and I. Wachsmuth. 2004. Synthesizing multimodal utterances for conversational agents. *Computer Animation and Virtual Worlds*, 15(1): 39–52. DOI: https://doi.org/10.1002/cav.6.

S. Kopp, B. Krenn, S. Marsella, A. N. Marshall, C. Pelachaud, H. Pirker, K. Thorisson, and H. Vilhjalmsson. 2006. Towards a common framework for multimodal generation in ECAs: The Behavior Markup Language. In *Intelligent Virtual Agents*, volume 4133 of *Lecture Notes in Computer Science*, pp. 205–217. Springer, Berlin.

M. Kyriakou, X. Pan, and Y. Chrysanthou. 2017. Interaction with virtual crowd in immersive and semi-immersive virtual reality systems. *Computer Animation and Virtual Worlds*, 28(5): e1729. ISSN 1546-427X. https://onlinelibrary.wiley.com/doi/abs/10.1002/cav.1729. DOI: https://doi.org/10.1002/cav.1729.

H. Laga and T. Amaoka. 2009. Modeling the spatial behavior of virtual agents in groups for non-verbal communication in virtual worlds. In *Proceedings of the 3rd International Universal Communication Symposium on*, pp. 154–159. ACM Press, Tokyo, Japan. ISBN 978-1-60558-641-0. DOI: https://doi.org/10.1145/1667780.1667811.

J. C. Lester, J. L. Voerman, S. G. Towns, and C. B. Callaway. 1999. Deictic believability: Coordinated gesture, locomotion, and speech in lifelike pedagogical agents. *Applied Artificial Intelligence*, 13(4-5): 383–414. DOI: https://doi.org/10.1080/088395199117324.

J. C. Lester, S. G. Towns, C. B. Callaway, J. L. Voerman, and P. J. FitzGerald. 2000. Deictic and emotive communication in animated pedagogical agents. In J. Cassell, J. Sullivan, S. Prevost, and

E. Churchill, eds., *Embodied Conversational Agents*, pp. 123–154. MIT Press, Cambridge.

J. Li. 2015. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, 77: 23–37. ISSN 1071-5819. DOI: 10.1016/j.ijhcs.2015.01.001.

J. Llobera, B. Spanlang, G. Ruffini, and M. Slater. 2010. Proxemics with multiple dynamic characters in an immersive virtual environment. *ACM Transactions on Applied Perception*, 8(1): 3:1–3:12. ISSN 1544-3558. https://doi.org/10.1145/1857893.1857896. DOI: 10.1145/1857893.1857896.

C. Mavrogiannis, F. Baldini, A. Wang, D. Zhao, P. Trautman, A. Steinfeld, and J. Oh. 2021. Core Challenges of Social Robot Navigation: A Survey. *arXiv:2103.05668 [cs]*. http://arxiv.org/abs/2103.05668.

S. Narang, A. Best, T. Randhavane, A. Shapiro, and D. Manocha. 2016. PedVR: Simulating gaze-based interactions between a real user and virtual crowds. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, VRST '16, pp. 91–100. ACM, New York, NY. ISBN 978-1-4503-4491-3. https://doi.org/10.1145/2993369.2993378. DOI: 10.1145/2993369.2993378.

N. Nguyen and I. Wachsmuth. 2011. From body space to interaction space: modeling spatial cooperation for virtual humans. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 3*, pp. 1047–1054. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC. ISBN 978-0-9826571-7-1.

R. Niewiadomski, E. Bevacqua, M. Mancini, and C. Pelachaud. 2009. Greta: An interactive expressive ECA system. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pp. 1399–1400. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC.

M. Obaid, I. Damian, F. Kistler, B. Endrass, J. Wagner, and E. André. 2012. Cultural behaviors of virtual agents in an augmented reality environment. In *Proceedings of the 12th international conference on Intelligent Virtual Agents*, volume 7502 of *Lecture Notes in Computer Science*, pp. 412–418. Springer, Berlin. ISBN 978-3-642-33196-1. https://doi.org/10.1007/978-3-642-33197-8_42.

C. Oliva and H. H. Vilhjálmsson. 2014. Prediction in social path following. In *Proceedings of the Seventh International Conference on Motion in Games*, pp. 103–108. ACM, New York, NY, USA. ISBN 978-1-4503-2623-0. https://doi.org/10.1145/2668064.2668103. DOI: 10.1145/2668064.2668103.

E. Padilha and J. Carletta. 2002. A simulation of small group discussion. In *Proceedings of the 6th Workshop on the Semantics and Pragmatics of Dialogue*. https://www.research.ed.ac.uk/en/publications/a-simulation-of-small-group-discussion.

C. Pedica and H. H. Vilhjálmsson. 2008. Social Perception and Steering for Online Avatars. In H. Prendinger, J. C. Lester, and M. Ishizuka, eds., *Proceedings of the 8th International Conference on Intelligent Virtual Agents*, volume 5208 of *Lecture Notes in Computer Science*, pp. 104–116. Springer, Berlin. ISBN 978-3-540-85482-1. http://dblp.uni-trier.de/db/conf/iva/iva2008.html#PedicaV08.

C. Pedica and H. H. Vilhjálmsson. 2012. Lifelike Interactive Characters with Behavior Trees for Social Territorial Intelligence. In *ACM SIGGRAPH 2012 Posters*, pp. 32:1–32:1. ACM, New York, NY, USA. ISBN 978-1-4503-1682-8. https://doi.org/10.1145/2342896.2342938. DOI: 10.1145/2342896.2342938.

C. Pedica, H. H. Vilhjálmsson, and M. Lárusdóttir. 2010. Avatars in conversation: The importance of simulating territorial behavior. In *Proceedings of the International Conference on Intelligent Virtual*

*Agents*, volume 6356 of *Lecture Notes in Computer Science*, pp. 336–342. Springer, Berlin.

N. Pelechano, J. M. Allbeck, and N. I. Badler. 2007. Controlling individual agents in high-density crowd simulation. In *Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 99–108. Eurographics Association, Goslar, DEU. ISBN 978-1-59593-624-0.

N. Pelechano Gómez, C. Stocker, J. Allbeck, and N. Badler. 2007. Feeling crowded? Exploring presence in virtual crowds. In *Proceedings of PRESENCE 2007*, pp. 373–376. UPCommons. https://upcommons.upc.edu/handle/2117/15704.

C. Peters and C. Ennis. 2009. Modeling Groups of Plausible Virtual Pedestrians. *IEEE Computer Graphics and Applications*, 29(4): 54–63. ISSN 1558-1756. DOI: 10.1109/MCG.2009.69.

M. Popelová, M. Bída, C. Brom, J. Gemrot, and J. Tomek. 2011. When a couple goes together: Walk along steering. In J. M. Allbeck and P. Faloutsos, eds., *Motion in Games*, volume 7060 of *Lecture Notes in Computer Science*, pp. 278–289. Springer, Berlin.

T. Randhavane, A. Bera, and D. Manocha. 2017. F2FCrowds: Planning agent movements to enable face-to-face interactions. *Presence*, 26(2): 228–246. DOI: 10.1162/PRES_a_00294.

Z. Ren, P. Charalambous, J. Bruneau, Q. Peng, and J. Pettré. 2017. Group modeling: A unified velocity-based approach. *Computer Graphics Forum*, 36(8): 45–56. DOI: https://doi.org/10.1111/cgf.12993.

C. W. Reynolds. 1987. Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pp. 25–34. ACM, New York, NY. DOI: 10.1145/37401.37406.

C. W. Reynolds. 1999. Steering behaviors for autonomous characters. In *Proceedings of the Game Developers Conference*, volume San Jose, CA, pp. 763–782. Miller Freeman Game Group.

J. Rickel and W. L. Johnson. 1999. Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13(4-5): 343–382. DOI: https://doi.org/10.1080/088395199117315.

J. Rickel and W. L. Johnson. 2000. Task-oriented collaboration with embodied agents in virtual worlds. In *Embodied Conversational Agents*, p. 122. MIT Press, Cambride, MA.

J. Rios-Martinez, A. Spalanzani, and C. Laugier. 2015. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics*, 7(2): 137–153. DOI: 10.1007/s12369-014-0251-1.

F. Rojas, F. Tarnogol, and H. S. Yang. 2016. Dynamic social formations of pedestrian groups navigating and using public transportation in a virtual city. *The Visual Computer*, 32(3): 335–345. http://link.springer.com/10.1007/s00371-015-1187-7. DOI: 10.1007/s00371-015-1187-7.

F. A. Rojas and H. S. Yang. 2013. Immersive human-in-the-loop HMD evaluation of dynamic group behavior in a pedestrian crowd simulation that uses group agent-based steering. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, pp. 31–40. ACM, New York, NY, USA. DOI: 10.1145/2534329.2534336.

F. A. Rojas, H. S. Yang, and F. M. Tarnogol. 2014. Safe navigation of pedestrians in social groups in a virtual urban environment. In *Proceedings of the International Conference on Cyberworlds*, pp. 31–38. IEEE. DOI: 10.1109/CW.2014.13.

S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita. 2009. How to approach humans?: Strategies for social robots to initiate interaction. In *Proceedings of the 4th*

*ACM/IEEE International Conference on Human Robot Interaction*, pp. 109–116. ACM. DOI: https://doi.org/10.1145/1514095.1514117.

A. E. Scheflen. 1975. Micro-territories in human interaction. In A. Kendon, ed., *Organization of Behavior in Face-to-Face Interaction*, World Anthropology, pp. 159–174. Mouton Publishers. ISBN 978-3-11-090764-3.

A. E. Scheflen. 1976. *Human Territories: How we Behave in Space and Time*. Prentice-Hall, Boston, MA. ISBN 1-59593-364-6.

W. Shao and D. Terzopoulos. 2005. Autonomous pedestrians. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pp. 19–28. ACM, New York, NY. DOI: 10.1145/1073368.1073371.

A. Shapiro. 2011. Building a character animation system. In *Motion in Games, MIG 2011*, volume 7060 of *Lecture Notes in Computer Science*, pp. 98–109. Springer, Berlin.

J. Shi, T. Smith, J. Granieri, and N. Badler. 1999. Smart avatars in JackMOO. In *Proceedings of IEEE Virtual Reality*, pp. 156–163. IEEE. DOI: 10.1109/VR.1999.756946.

L. Sun, A. Shoulson, P. Huang, N. Nelson, W. Qin, A. Nenkova, and N. I. Badler. 2012. Animating synthetic dyadic conversations with variations based on context and agent attributes. *Computer Animation and Virtual Worlds*, 23(1): 17–32. DOI: https://doi.org/10.1002/cav.1421.

A. Thordarson and H. H. Vilhjálmsson. 2019. SoCueVR: Virtual reality game for social cue detection training. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pp. 46–48. ACM, New York, NY. DOI: https://doi.org/10.1145/3308532.3329440.

K. Thorisson. 1997. Gandalf: An embodied humanoid capable of real-time multimodal dialogue with people. In *The Proceedings of the 1st International Conference on Autonomous Agents*, pp. 536–537. ACM.

P. R. Thrainsson, A. L. Petursson, and H. H. Vilhjálmsson. 2011. Dynamic planning for agents in games using social norms and emotions. In H. H. Vilhjálmsson, S. Kopp, S. Marsella, and K. R. Thórisson, eds., *Proceedings of the International Conference on Intelligent Virtual Agents*, volume 6895 of *Lecture Notes in Computer Science*, pp. 473–474. Springer, Berlin.

J. van den Berg, M. Lin, and D. Manocha. 2008. Reciprocal velocity obstacles for real-time multi-agent navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1928–1935. IEEE. DOI: 10.1109/ROBOT.2008.4543489.

D. Veutgen, M. Massetti, J. Rossi, L. Veroli, A. Ásgeirsdóttir, G. Baldursdóttir, S. Gissurardóttir, G. Gudmundsson, T. Sigurdardóttir, V. Laenen, and H. H. Vilhjálmsson. 2018. Interpreting social commitment in a simulated theater. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pp. 289–294. ACM, New York, NY. http://doi.acm.org/10.1145/3267851.3267919. DOI: 10.1145/3267851.3267919.

H. Vilhjalmsson, C. Merchant, and P. Samtani. 2007. Social puppets: Towards modular social animation for agents and avatars. In D. Schuler, ed., *Online Communities and Social Computing*, volume 4564 of *Lecture Notes in Computer Science*, pp. 192–201. Springer, Berlin.

H. Vilhjalmsson, E. Björgvinsson, H. Helgadottir, K. V. Kristinsson, S. Ólafsson, A. Cafaro, N. Kramer, J. Braier, P. Wegge, S. Youn, C. Pedica, B. Arnbjörnsdóttir, and B. Bédi. 2014. Social Gatherings in Virtual Reykjavik. In *Poster and Demo at the 14th International Conference on Intelligent Virtual Agents*. Boston, MA.

H. H. Vilhjalmsson. 1996. Autonomous communicative behaviors in avatars. In *Proceedings of Lifelike Computer Characters '96*, p. 49. Snowbird, Utah.

H. H. Vilhjálmsson and J. Cassell. 1998. BodyChat: Autonomous communicative behaviors in avatars. In *Proceedings of the Second International Conference on Autonomous Agents*, pp. 269–276. ACM, New York, NY. DOI: 10.1145/280765.280843.

J. Zhao and N. I. Badler. 1994. Inverse kinematics positioning using nonlinear programming for highly articulated figures. *ACM Transactions on Graphics*, 13(4): 313–336. DOI: 10.1145/195826.195827.

S. Ólafsson, B. Bédi, H. E. Erla Helgdóttir, B. Arnbjörnsdóttir, and H. Högni Vilhjálmsson. 2015. Starting a conversation with strangers in virtual Reykjavik: Explicit announcement of presence. In *Proceedings from the 3rd European Symposium on Multimodal Communication*, pp. 62–68. Linköping University Electronic Press, Linköpings universitet. ISBN 1650-3740.