Blending Embodied Interactions Across Heterogeneous Contexts

THE UNIVERSITY OF
SYDNEY

Emily Wong

Supervisor: Prof. Eduardo Velloso

School of Computer Science
Faculty of Engineering and IT
University of Sydney
Australia

A thesis submitted to fulfil the requirements of the degree of
Master of Philosophy in Human-Computer Interaction

January 2026

Statement of Original Authorship

This is to certify that the content of this thesis is my own work. This thesis has not been submitted for any other degree or purpose.

I certify that the intellectual content of this thesis is the product of my own work, and that all assistance received in preparing this thesis and all sources have been acknowledged.

____

Emily Wong
Date:

Acknowledgements

Although written between 2024 and 2025, this thesis draws on over three years of brain-bending conversations about what it means to ‘blend’ distributed spaces. It has been a great privilege to take part in these discussions and contribute to the emerging field of blended realities. This journey is thanks to the encouragement and inspiration of the many people around me, to whom I owe my gratitude.

A special thank you to my supervisor Eduardo Velloso, for his inspiring approach to the rigorous pursuit of knowledge, and for fanning the flames of my weird ideas. I always leave our conversations feeling excited to be a researcher. Without his leadership, willingness to engage in philosophical discussions and constant belief in me, this research would not exist.

Thank you to Jens Emil Grønbæk for the many thought-provoking conversations spanning space, collaboration, and computing. I am immensely grateful for his generosity of time and knowledge, for showing me around Aarhus, his infectious enthusiasm that drives blended realities forward, and involving me in this mission.

Thank you to all the members of the Blended Realities Collective for their contribution to the foundational ideas in this thesis. In particular Germán Levia, for the creativity and energy he brought to designing studies, both in this thesis and earlier work. To Adélaïde Genay, for her insight that strengthened the Spatial Heterogeneity Framework and contributions across various blended realities projects. To Brandon Syiem, for his support running the Sydney Lab and organising a number of influential workshops. To Juan Sánchez Esquivel, Tinghui Li, Jarrod Knibbe, Andrew Irlitti, Marvin Bai, and Tiare Feuchtner, for kindly sharing their time and knowledge in various workshops that seeded the ideas in this thesis.

To my peers and colleagues at the University of Sydney and The University of Melbourne for their continued encouragement; to Frank Vetere for the engaging conversations on design in Human-Computer Interaction; and to Wafa Johal and Jorge Goncalves for welcoming me into the lab and looking out for me.

To Ava Scott and Nina Rajcic, for the late-night conversations on the cognitive dissonance of being captivated by technology’s possibilities while wary of its application, what good research looks like, and where Human-Computer Interaction fits into everything.

Thank you to Lewis Nicolson, Olivia Stagg, Oliver Breedon and Edward Howard, for your friendship, house dinners, and making the last year scribe’s honour (not horror).

A heartfelt thank you to Iwan Williams, my resident Philosopher of Mind and Cheer Captain. Thank you for teaching me the art of arm-chairing, introducing me to the world of externalism, and expanding my outlook on life in the best way possible.

Finally, thank you to Rhona, Kim and Andrew Wong for all your love and support. You give me the motivation and ability to follow my Ngāruawāhia Moth.

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

Authorship Attribution Statement

This thesis contains material that was previously published in: Wong, E., Genay, A., Grønbæk, J. E. S. & Velloso, E. (2025), Spatial heterogeneity in distributed mixed reality collaboration, in 'Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems', CHI '25, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3706598.3714033.

This material comprises of Chapter 4, Section 2.2.2 and Section 6.2. I co-designed the framework with Adelaïde Genay, Jens Emil Grønbæk and Eduardo Velloso, I performed the literature review, designed the figures and wrote the drafts of the manuscript.

The Association for Computing Machinery, Conference on Human Factors in Computing Systems requires that the lead author must be the corresponding author.

During the preparation of the thesis ChatGPT was used for editorial purposes. The use of this generative AI tool includes: making sections more concise and improving sentence structures. I confirm that where text was modified by generative AI, the content was reviewed for possible errors, inaccuracies, and bias. I take full responsibility for the submitted thesis and ensuring the work is my own and I have used generative AI within the parameters of use.

In addition to the authorship attribution statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Emily Wong
Date:

As the supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Professor Eduardo Velloso
Date:

Abstract

Blended realities, an emerging subset of mixed reality (MR) systems, offers new possibilities for supporting embodied, non-verbal interactions across physically distributed environments. These systems deliberately integrate physical and digital spaces to enable a sense of co-presence and shared activity. However, blending coherent embodied interactions across dissimilar physical spaces remains challenging. Differences in spatial layout, scale, and social norms can introduce perceptual distortions that lead to miscommunication. Although existing approaches focus on reducing spatial discrepancies, these often overlook the role of place-making and how people use space to construct roles, identities, and expectations.

This thesis investigates how embodied actions are transmitted across heterogeneous contexts in blended realities, and the design trade-offs this entails. Importantly, a focus on *context* includes both spatial and sociocultural aspects of collaboration.

To examine these aspects in concert, I introduce a taxonomy of three interaction transmission modes: (1) transcription, where interactions are relayed without modification; (2) transformation, where spatial mappings are adjusted to align disparate environments; and (3) translation, where actions are adapted by the system to preserve sociocultural norms.

In the transformation mode, spatial heterogeneity—the degree of difference between the spatial layouts of distributed spaces—is a foundational challenge. However, blended realities currently lacks a unifying language and theoretical construct to identify, discuss, and compare different blended realities solutions. To overcome this challenge, I present the Spatial Heterogeneity Framework, developed through analysis of 14 canonical systems, academic workshops, and a systematic review of 32 blended reality systems from 1995–2024. This framework characterises how different systems support blended proxemics and where design interventions are required as spatial heterogeneity increases.

Although spatial heterogeneity is an important aspect of blending heterogeneous contexts, few studies investigate how transforming spatial information is enabled together with translating sociocultural norms. In this thesis I demonstrate how these different transmission modes can contradict each other in complex blended realities environments. To explore how people weigh up the trade-offs between these different transmission strategies, I conducted a qualitative user study with 20 participants. By presenting speculative interaction scenarios across different social and task-based contexts, I investigate how people weigh priorities such as spatial coherence, social appropriateness, and task efficiency when engaging with blending techniques. These

results demonstrate how transformation and translation trade-offs are perceived, including when participants expect systems to preserve spatial information versus sociocultural norms.

I make three overarching contributions in this thesis:

1. **The more dissimilar the spaces are, the more potential there is for perceptual distortions.**
   The Spatial Heterogeneity Framework shows that to sustain spatially coherent interactions across distributed environments, blended realities systems must assume a degree of spatial similarity or difference. As heterogeneity increases, fewer embodied actions are afforded "for free" by the physical layout, demanding more complex system engineering and raising the likelihood of perceptual distortions.
2. **Sociocultural norm preferences are mediated by the blended realities system.**
   Beyond spatial heterogeneity, coherent communication requires accounting for sociocultural norms. I extend Judee Burgoon's (1993) Expectancy Violations Theory to explain how people evaluate trade-offs between maintaining accurate spatial information and upholding sociocultural expectations. This extension highlights how divergent perspectives and awareness of system mediation can obscure the way people judge behaviours as intentional or as system glitches.
3. **There are ethical implications for adapting embodied actions in blended realities.**
   Adapting embodied actions in blended realities introduces trade-offs with transparency, raising ethical questions about when systems should reveal whether actions stem from human intent or machine mediation. There are other options for navigating heterogeneous contexts, involving user-driven strategies, such as adjusting the physical layout or modifying activity requirements. However, these place the burden of adaptation on the people using the system. Future research should consider the relationship between transparency and reducing cognitive load.

Together, these contributions advance a theoretical foundation for designing context-sensitive blended realities. It goes beyond aligning geometries, aiming to answer a need for systems that account for how spaces become socially-informed places across heterogeneous settings.

Table of contents

List of figures

Glossary of terms

active zone

A functional area in The Spatial Heterogeneity Framework that identifies the selected area inside the physical space where the entire collaborative activity happens.

activity zones

A component of the Spatial Heterogeneity Framework that identifies the different functional areas of the distributed spaces that need to be blended.

actual proxemics

The proxemics that are enacted by a collaborator in their local environment.

affine heterogeneity

The third level of spatial heterogeneity. The blended zones at this level are similar but have different dimensions, which requires stretching or reflection of one or more dimensions. If such a transformation was used to map the two zones directly, collaborators' actions might appear distorted (e.g. a step in one space might be equivalent to two steps in the other or writing on a surface might appear backward if not properly reflected).

appropriate proxemics

A mental model of the proxemic rules that a collaborator expects to be followed in a given collaborative context.

blended proxemics

The social interactions between people, objects and space supported across the distributed locations. The corresponding component in the Spatial Heterogeneity Framework identifies embodied interactions that are afforded by the spatial layout or through the use of a blending technique.

blended realities

A subset of distributed mixed reality experiences, where the remote collaborator's actions are presented according to the context in which the local collaborator views them.

blended zone

The area of the local collaborator’s public zone that should overlap or ‘blend’ with their remote collaborator’s public zone.

blending technique

A technical solution used to adapt the local collaborator’s actions so that they make sense in context of the remote collaborator’s space, particularly when blended proxemics are not afforded by the spatial layout “for free”. Examples of a blending technique include deixis warping (Sousa et al. 2019) or redirected walking (Wang et al. 2022).

categorical heterogeneity

The highest level of spatial heterogeneity. The blended zones are so different at this level of spatial heterogeneity that only the objects can be mapped to each other, but not the space between them. Spatial references are meaningless, though collaborators might still be able to make verbal references to the objects themselves.

causal heterogeneity

The lowest level of spatial heterogeneity. The idealised gold standard degree of spatial heterogeneity, equivalent to meeting face-to-face, where every physical element in one space is perfectly mapped to its counterpart in the other space, and actions in one space have consequences in the other. If the spaces are blended at this level, spatial heterogeneity is not a challenge. However, they require substantial infrastructure to guarantee such bidirectional causation.

communicator characteristics

The salient features of an interaction partner, such as gender, age, personality and communication style (White 2008, p.191).

communicator reward valence

The assessment of positive or negative attributes of a interaction partner that moderate the evaluation of an expectancies violation. When a well regarded communicator interacts at a closer distance than expected, for example, people are likely to evaluate the violation positively, but when the same distance is adopted by a communicator who is not well regarded, then people are likely to evaluate the violation negatively (White 2008, p.192).

congruent body position and trajectory

A type of blended proxemics where if the local collaborator stands, sits, walks or turns to face a particular object, every step of this trajectory has a corresponding point in the remote collaborator’s space.

consistent vicinity

A type of blended proxemics where the interactions on and around particular landmarks are kept consistent.

context characteristics

Aspects of the environment that might define how individuals should communicate in a particular situation, such as the formality of the setting or nature of the task (White 2008, p.191).

deictic cues

Pointing, gaze shifts, and head or body orientation (Levinson 2006), often enabled in digital systems to help people reference shared objects and establish mutual understanding during collaboration.

embodied actions

The creation and interpretation of meaning through people's bodily engagement with the surrounding environment, where actions derive significance from their spatial, social, and cultural context (Dourish 2001). In this thesis embodied actions refers to the movements carried out by the collaborators in their local environments.

embodied interactions

The creation, manipulation, and sharing of meaning through engaged interaction with artefacts (Dourish 2001, p.126). In this thesis, embodied interactions refer to the overall way a blended realities system transmits embodied actions across the distributed environments.

expectancies

The behaviours that are appropriate for a situation or particular group, or they may reflect what we know to be typical behaviour of a specific individual. Expectancies can refer to what we anticipated will occur (predictive expectancies) or to what is desired or preferred (prescriptive expectancies) (White 2008, p.191).

expectancy violations theory

Explains how individuals respond when others behave in ways that deviate from expected norms during social interactions. Expectations are shaped by cultural conventions, personal experiences, and situational contexts. When they are violated, people evaluate the behaviour as positive or negative, depending on factors such as relational dynamics and the communicator's perceived reward value (Burgoon 1993).

externalism

A position in philosophy of mind, which argues that cognition is not (entirely) determined by what happens inside the brain, but also depends on a person's interaction with and relation to the external world.

facing formations (f-formations)

A theory describing how groups position their bodies in space to signal shared focus and social engagement. F-formations typically involve 2–5 participants and vary in shape depending on the task. The central o-space marks the group's shared focus, the surrounding p-space defines membership, and the outer r-space separates the group from its environment (Kendon 1990).

heterogeneity ladder

A component of the Spatial Heterogeneity Framework that identifies six different levels of spatial heterogeneity that blended realities systems generally assume.

heterogeneous contexts

The degree of spatial and sociocultural disparities across distributed environments.

inactive zone

The zone inside the physical space that the collaborator is neither using nor planning to use during the collaborative activity.

independent zone

The area inside the public zone that is visible to other collaborators but is inaccessible or does not guarantee a mapping with other the collaborators' public zones. This is different to the private zone since remote collaborators *can still see interactions* in this area but are not able to act in this area themselves.

mixed reality

A class of systems that merge computer generated virtual environments with real physical environments (Ens et al. 2019). These sit along a reality-virtuality continuum (Milgram & Kishino 1994), which can include augmented reality on one end, and fully immersed experiences like substitutional reality on the other (Simeone et al. 2015).

modes of transmitting embodied actions

The different ways in which embodied actions are encoded by blended realities systems and sent to the remote environment, where they are then decoded in context of the remote space and received by the remote participant (Shannon 1948).

non-linear continuous heterogeneity

The fourth level of spatial heterogeneity. The blended zones at this level share features, but they are not easily aligned. For example, if trying to map a straight corridor onto a curved one, the points in one space are warped to fit the remote space.

non-linear discontinuous heterogeneity

The fifth level of spatial heterogeneity. The blended zones at this level might have discontinuities, amplifying the problems of the previous rung. For example, a person walking from one blended area of their space to another separated by private space may appear as suddenly "pausing" their walk in the remote partner's space and then resuming it as they finish traversing the private space. Here, the local space is spliced into sections and these smaller homogeneous spaces are mapped to each other.

one-to-one scale

A type of blended proxemics where the avatar's movements and interactions in their local space are at the same scale as their actions in the remote space.

perceived proxemics

The proxemics, represented by the remote collaborator's avatar that the local collaborator observes in their local environment. Importantly these proxemics may have been adapted by the blended realities system.

physical manipulation

A type of blended proxemics where moving an object in the local space, it also moves the equivalent object in the remote user's space.

physical space

The fixed (e.g. walls), semi-fixed e.g.(tables) and mobile features (e.g. laptops) of the collaborators' environments, as well as the space between them.

private zone

the area within the local collaborator's active zone that is not visible to their remote collaborator.

proxemics

The study of how people use space in social interaction, particularly the physical distances they maintain between themselves and others, and how these spatial arrangements communicate meaning.

public zone

The selected area inside the local collaborator's active zone that is visible to their remote collaborator.

relational characteristics

The degree of familiarity between partners or the equality of status between interaction partners (White 2008, p.191).

rigid heterogeneity

The second level of spatial heterogeneity. The blended zones at this level have identical fixed and semi-fixed structures, requiring only a rotational transformation for the mapping. This level requires carefully arranging the space so that the physical objects and the space around them perfectly align.

semantic mapping

The process of identifying and matching spatial features based on their function or meaning, regardless of their exact physical form or location, to enable blending techniques like redirected pointing.

social noise

When adaptations in the transform and translate mode come into conflict and the intended meaning of an action is distorted.

solutions matrix

A component of the Spatial Heterogeneity Framework that shows the assumptions different systems make about the distributed spaces' level of spatial heterogeneity. Systems that support the highest level of spatial heterogeneity will support all the levels below.

spatial deictic cues

A type of blended proxemics where if the collaborator points or verbally references the position of an object in their local space, their communication method points to the corresponding object in the remote space.

spatial heterogeneity

The degree of difference between the spatial layouts of distributed spaces. These layouts include objects, which may be fixed (e.g. walls), semi-fixed (e.g. tables), or mobile (e.g. laptops), as well as the space between them.

theory of mind

The ability to attribute mental states (such as beliefs, intentions, desires, emotions, and knowledge) to others, understanding that these states may differ from one's own.

tight and loose coupling

The degree to which semantically mapped objects are similar or dissimilar in functionality e.g. mapping a whiteboard in the local space to a whiteboard in the remote space is more tightly coupled than mapping a whiteboard to a desk. In the latter example, both these spatial features function as a drawing surface but are not identical objects.

transcribe mode

When embodied actions are carried directly from one space to another without any adaptations.

transform mode

When embodied actions are adapted by the blended realities system to overcome spatial differences.

translate mode

When embodied actions are adapted by the blended realities system to uphold specific sociocultural norms.

verbal referencing

A type of blended proxemics where participants know about the presence of objects common to both spaces and can verbally identify them. It should be noted that this blended proxemic does not include deictic cues, such as “go right" or “put that there", these sit in spatial deictics since they have a directionality associated with them.

violation valence

The positive or negative evaluation of a expectancies violation. Expectancy Violations Theory predicts that a violation that has a positive valence will typically lead to better interaction outcomes than a non-violation. A violation that has a negative valence will typically lead to worse interaction outcomes than simply meeting expectations. The valence of a violation determines whether it will be better to do what is expected or to deviate from the norm (White 2008, p.191).

Chapter 1

Introduction

> *Imagine a planet like Earth – call it "Twin Earth" – on which the liquid that fills the lakes and rivers that people drink is not H2O but a different compound XYZ, with similar superficial characteristics... upon learning that Twin Earth "water" does not consist of H2O at all, we Earthers would say "it isn't really water". The word "water" has a different meaning on Earth and on Twin Earth.*
>
> **Hillary Putnam's** Twin Earth Thought Experiment (1973).

While working as an event facilitator, I used to run an ice-breaker game called *categories*. The instruction was simple: "everyone get into groups according to how you travelled to work today." At first, participants would look puzzled, thinking, "how *did* I get to work today?" Then suddenly the room would erupt with voices calling out, "Bike! Car! Tram! Walk! Train!" Changing tactics, participants would quickly discover that spatial grouping was more effective than verbal communication. In some workshops, a participant might even grab a pen and a sheet of paper, scrawling TRAM in bold letters. Holding it aloft, tram-goers would swarm toward them. Within minutes, the room transformed into a map of commuters. Bikers in one corner, tram-goers gathered under their sign, walkers, train-goers, and drivers forming their own clusters. Laughter and light-hearted commentary usually followed; participants reflecting on the city's abundance of cyclists or the social patterns revealed by how people travelled.

This simple game demonstrates how cognition and coordination can be externalised through space, artefacts, and body movements. What began as a verbal exchange quickly evolved into spatial clustering, supported by physical props and gestures. By the end, the entire room could infer commuting patterns by observing the size and distribution of groups in space.

These embodied actions are an essential part of how people coordinate social activities and offload cognition in everyday tasks. The term *embodied actions* (Dourish 2001) refers to the way people communicate information through their physical engagement with the external world. The way people interact in space, through the use physical artefacts, gestures, or spatial arrangements, extends cognitive processes beyond "skin and skull" (Clark & Chalmers 1998) into the external environment. This philosophical stance positions external space as an active component in cognition. Dourish (2001) goes further to suggest that designed objects enable *embodied interactions*, "the creation, manipulation, and sharing of meaning through engaged interaction with artefacts" (2001, p.126). This was evident in the participant who scrawled "tram" on a piece of paper to encourage action, or similarly how drawing on a whiteboard might allow someone to externalise their ideas. In turn, this person's collaborator may respond with embodied gestures, nodding enthusiastically or leaning back with crossed arms to express agreement or hesitation. In each case, the physical space facilitates externalised actions that actively support cognitive and social processes.

As geographically distributed teams have become more common, collaborators increasingly find themselves working across different physical environments. This spatial separation has motivated a wide range of computer-supported approaches for enabling embodied interactions across distributed locations. Systems that enable these embodied interactions range from video conferencing systems that preserve spatial sound and gestures (Baldis 2001), to media spaces (Buxton 2009), shared virtual environments like collaborative virtual environments (CVE) (Benford et al. 2001) or mixed reality (MR) (Ishii & Kobayashi 1992, Orts-Escolano et al. 2016, Suzuki et al. 2020), and even telepresence robotics (Tsui et al. 2011). Each of these approaches carry different trade-offs in terms of how well they sustain the spatial and social dynamics that make embodied interactions meaningful.

In particular, MR environments promise to merge the benefits of physical space with the flexibility of digital augmentation, enabling new forms of collaboration across distributed spaces (Ens et al. 2019, Benford et al. 1998). However, to coherently transmit embodied interactions across distributed locations, MR systems often require the spaces to have similar physical layouts (Grønbæk et al. 2023). This presents a practical challenge for adoption, since collaboration spaces often differ in size and layout; a situation that has become increasingly common with the rise of home offices and work-from-anywhere culture.

Blended realities is an emerging subset of distributed MR experiences, where the remote collaborator's actions are presented according to the context in which the local collaborator views them. Current solutions enable this experience by remapping features in the local participant's space to correspond with features in their remote collaborator's space (Grønbæk et al. 2024). In doing so, it is possible to to adapt the embodied actions carried out in one space to the spatial layout of the other. For example, in **Figure 1.1** the professor has a large office with a desk and whiteboard to the left of the desk. Whereas their student has a smaller space with a desk and a whiteboard directly behind them. With blended realities, it is possible for the professor to walk over to their whiteboard and the student will see the professor's avatar appear at the whiteboard behind them. If the student stands up and joins them, from the professor's perspective, the student's avatar will appear beside them at their whiteboard so the two can use embodied cues to take turns sketching ideas. As such, blended realities enable working across spatially dissimilar environments.

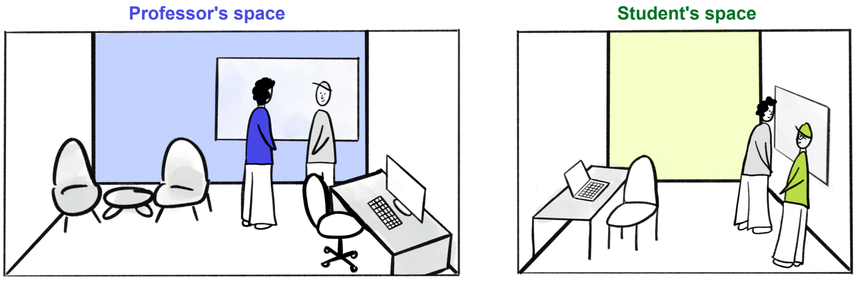


Fig. 1.1 Current blended realities systems overcome the spatial differences of distributed locations by mapping the physical features of the spaces to each other. While the professor and student's whiteboards are in different locations, the system will render their avatars so they appear in front of the same semantically mapped object. This helps them preserve embodied actions across the heterogeneous environments.

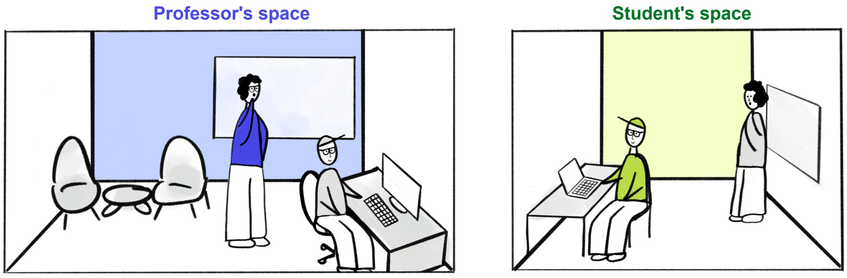


Fig. 1.2 This figure demonstrates the cultural nuances of heterogeneous contexts. While the student sitting behind the desk is spatially accurate, from the professor's perspective, it might seem strange or rude that the student is sitting in a space traditionally reserved for them.

In this thesis, I use embodied actions to describe how people move and act within their local space, and embodied interactions to refer to the overall way that blended realities systems adapt and mediate actions across distributed environments.

To enable blended realities, it is not enough to focus on the spatial mapping alone. The system must also account for sociocultural aspects of place-making enabled by the way people situate themselves in different spaces (Harrison & Dourish 1996, Lawson 2007). Consider once again a professor and student collaborating in a blended realities system (**Figure 1.2**). A purely spatial mapping might render the student's avatar in a chair directly behind the professor's desk, since that is the corresponding location in their environment. While spatially accurate, such a configuration would appear socially inappropriate, as students would rarely sit behind the professor's desk in a real-world setting. A more contextually appropriate design would render the student in the guest chair, preserving both spatial alignment and sociocultural expectations. This example demonstrates the complexity of context, which involves both the denotation of entities and their positions in space (Dey 2001), as well as the sociocultural norms that influence how these positions are interpreted and enacted in different social situations (Dourish 2004). In this thesis, I

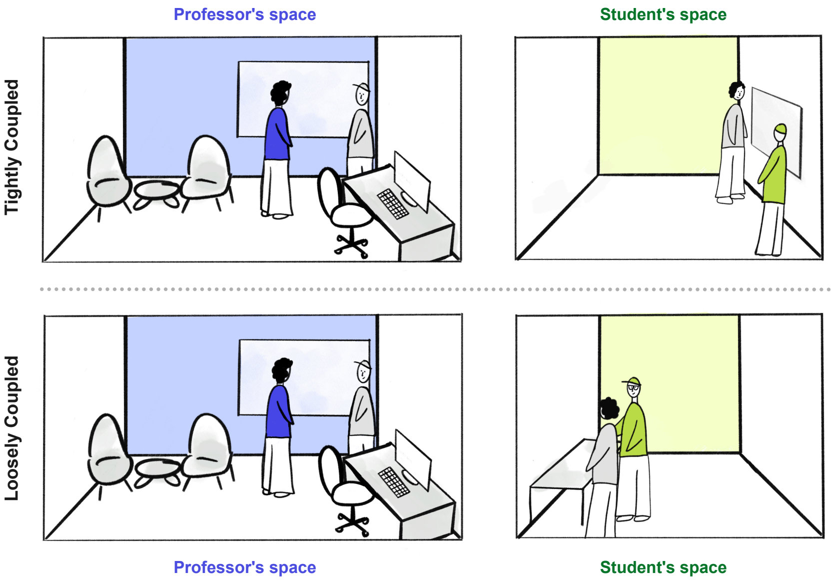


Fig. 1.3 In the top scenario, the professor's whiteboard is tightly coupled with the student's whiteboard. Whereas, in the bottom scenario, the professor's whiteboard is loosely coupled with the student's desk: while in both locations the professor and student are standing at their designated drawing surfaces, the objects differ significantly, creating a 'twin world' (Putnam 1975).

group spatial and sociocultural disparities across distributed locations under the umbrella term, *heterogeneous contexts*.

A unique aspect of blending heterogeneous contexts is that the shared illusion of co-location consists of different ground truths (Wong, Grønbæk & Velloso 2024). The previously exemplified blended realities system (**Figure 1.1**) adapts the professor's actions so they make sense in the context of the student's environment, while simultaneously doing the same for the professor. In this way, a kind of "twin world" (Putnam 1975) is instantiated, where the spatial and sociocultural features of one environment can be more tightly or loosely coupled with another. For example, in **Figure 1.3** spatial interactions might be tightly coupled if the professor and student's spaces are identical, but only loosely coupled if their environments differ significantly. In a tight coupling, the student might have the same desk and whiteboard layout as the professor, but in a loose coupling, the student may use a desk as both a tabletop and a drawing surface. In this case, when the professor walks toward the whiteboard, the student sees them remain at their desk. Within the activity, both participants are in front of their drawing surfaces, yet spatially they occupy very different positions. In one environment, the professor is standing at the whiteboard, while in the other, both are seated at the desk. This "twin world" feature of blended realities makes it distinct from other types of collaborative systems.

As distributed contexts become more heterogeneous and embodied actions are loosely coupled with remote objects, maintaining a shared reality grows increasingly precarious (Wong, Grønbæk & Velloso 2024). The complexity of technical solutions required to transmit actions coherently

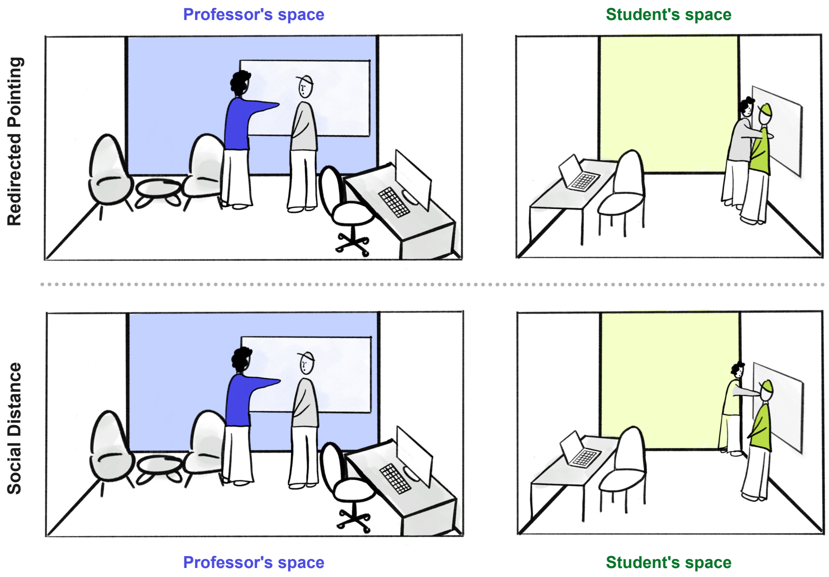


Fig. 1.4 Technical solutions in blended realities can conflict. In the top scenario, redirecting the professor's gesture preserves spatial alignment but invades the student's personal space. In the bottom scenario, offsetting the avatar avoids this, yet causes the professor's avatar to be rendered as though they are drawing inside the student's wall.

from one context to another makes it challenging to avoid contradicting interaction rules (Benford et al. 1998). For instance, in **Figure 1.4**, one rule tells the system to redirect the Professor's body posture so that the student sees them writing on their smaller whiteboard surface, yet another sociocultural rule could dictate that the professor should be placed at a socially acceptable distance, which offsets them from the whiteboard so they end up drawing inside the student's wall. These contradictions force the designer to compare the spatial and sociocultural interactions in the context of the collaborative scenario: what benefits do coherent embodied interactions around a shared surface provide? Does social distance matter if the collaborators are best friends, rather than a student and a professor? What if they are under severe time constraints, with real-life consequences?

While the ability to work across heterogeneous contexts creates exciting new opportunities for remote collaboration, the aforementioned design trade-offs and the existence of different ground truths raise ethical considerations. When systems are designed to adapt certain actions across heterogeneous contexts, participants may not realise that their collaborator experiences the action differently. This creates the potential for unintentional deception or misrepresentation and raises important questions about how interaction adaptations are designed and who or what controls this. It is important to address these issues early in the development of new technologies, and by raising such considerations, I provide the conceptual groundwork for future research.

In this thesis, I investigate what it means to blend spaces and sociocultural situations to transmit embodied actions across heterogeneous contexts. The ultimate goal of blended realities is to allow the perceiver to access their physical environment and view actions relative to their context, regardless of where the actions are carried out. I present a taxonomy that explains the different modes of transmitting embodied actions between contexts in blended realities. Focusing first on the *transform mode*, I examine how current solutions remap heterogeneous spaces to overcome their spatial differences. In doing so, I present the Spatial Heterogeneity Framework that can be used to identify and compare different strategies for blending heterogeneous spaces. Second, I argue that future solutions should consider a *translate mode* that adapts embodied actions and turn spaces into contextualised places, with specific norms, goals, and expectations. I examine the design trade-offs that arise when enabling the transformation and translation modes together across heterogeneous contexts. Finally, I discuss the implications of these interactions for the user experience of bended realities. Specifically, I raise the ethical issue of deception when transform and translate interactions cause ground truths to diverge, and participants are represented to their remote collaborators in a way that differs from what they experience in their local environment.

1.1 Motivation

As technologies evolve, they create new opportunities to understand how people collaborate across physical distances. For decades, researchers have debated whether remote collaboration technologies should strive to enable systems like Sutherland's (1965) "ultimate display", which seeks to replicate physical space dynamics across distributed locations, or instead move "beyond being there" to create experiences that surpass face-to-face interaction and are uniquely digital (Hollan & Stornetta 1992).

A primary critique of replication is that any imitation of in-person interaction will not measure up to the real deal. This has motivated designers to reimagine collaborative spaces in novel ways (Ishii & Kobayashi 1992, Dourish & Bellotti 1992, Gutwin & Greenberg 2002). Tools like Miro[1], which incorporates Gutwin and Greenberg's (2002) concept of workspace awareness, demonstrates how abstracting space within 2D surfaces can effectively support real-world collaboration.

Yet, despite arguments for abstraction, researchers continue to explore technologies that recreate immersive, three-dimensional spaces. These technologies offer new opportunities to reconstruct physical environments across distributed locations and continue to challenge previous assessments of the technical costs and interaction benefits for replicating spatial affordances in collaborative technologies (Benford et al. 1998). This ongoing evolution of technical capabilities means distributed collaboration remains an open challenge, one that invites deeper inquiry into the human experience of connecting across physical distance.

Blended realities is a particular solution area with new possibilities for transmitting embodied actions from one physical space to another. This emerging area of research has produced a rich

---

[1] www.miro.com

set of solutions that create blended physical-digital spaces, where embodied interactions across distributed spaces take place. These approaches include redirecting deictic gestures (Yoon et al. 2021, Fink et al. 2022), providing different task-based models of shared space (Herskovitz et al. 2022), mapping floor spaces (Lehment et al. 2014), and mapping functional interest points across spaces in a tightly coupled (Jo et al. 2015, Congdon et al. 2018) or loosely coupled manner (Grønbæk et al. 2023, Johnson et al. 2023, Grønbæk et al. 2024). While these works provide a growing set of interaction techniques and system designs, three key problem spaces remain unaddressed:

Problem Space 1) **Although spatial mapping is essential for blended realities, there is currently a lack of shared terminology and foundational theory to describe the trade-offs between different mapping strategies.** Without a common language or framework, it is difficult to compare solutions, understand their implications, or predict when one approach may be more effective than another. This limits researchers' ability to build on each other's work and hinders the development of generalisable design knowledge.
Opportunity: Develop a structured framework and shared vocabulary to discuss and compare how blended realities overcome spatial differences between distributed environments.

Problem Space 2) **Most existing approaches to blended realities focus on overcoming geometric differences between physical locations but could be extended to consider sociocultural aspects of context.** Importantly, places for collaboration are created through the accumulation of social meaning, cultural norms and patterns of activity (Harrison & Dourish 1996). Without attending to these sociocultural aspects of context, blended realities risk feeling disjointed or socially incoherent, even when geometric alignment is achieved.
Opportunity: explore how sociocultural interactions, such as personal space or social etiquette, should be supported across contextually distinct spaces.

Problem Space 3) **As distributed contexts become increasingly heterogeneous, the recontextualisation of interactions grows more complex, introducing new design trade-offs.** In this complexity, interaction "rules" embedded within systems may conflict. For example, one rule may prioritise spatial alignment for action coherence, while another enforces appropriate interpersonal distance for social comfort. Understanding how such trade-offs emerge and how they should be managed is crucial for guiding future design decisions that shape the user experience of blended realities.
Opportunity: understand the trade-offs designers must navigate when working across both spatial and sociocultural boundaries in MR systems.

By pursuing the opportunities presented, I envision a future for blended realities where systems can strategically recontextualise embodied actions across heterogeneous environments. Crucially, this vision does not argue that technologies for remote collaboration should all aim to mimic face-to-face experiences. Instead, it lays the groundwork for identifying when attributes of co-located embodied actions should be preserved through a solution like blended realities, and when alternative options should be selected to best support remote collaboration.

1.2 Aim and scope

This thesis investigates how embodied actions are transmitted across heterogeneous contexts in blended realities and the design trade-offs this involves.

To structure these investigations (**Figure 1.5**), in Chapter 3 I introduce a taxonomy with three modes of transmitting embodied actions in: (1) *transcription*, where actions are carried directly from one space to another without any adaptations; (2) *transformation*, where the system constructs models of each space and adapts embodied actions to overcome physical mismatches; and (3) *translation*, where embodied actions are adapted by the blended realities system to uphold specific sociocultural norms. Chapter 3 addresses Problem Space 2, aiming to demonstrate how sociocultural aspects of collaboration might be considered when blending heterogeneous contexts and the trade-offs between these different modes.

In Chapter 4, the *transformation* mode serves as the foundation for my development of the Spatial Heterogeneity Framework. This responds to Problem Space 1, conceptualising how blended realities map distributed spaces. By developing this framework, I aim to help researchers identify and compare mapping strategies, offering insight into their respective assumptions and limitations.

In Chapter 5, I investigate *transformation* and *translation* modes in unison, extending spatial heterogeneity to include sociocultural dimensions of place-making (Harrison & Dourish 1996). In doing so, I aim to investigate how the social and cultural aspects of a collaborative context change the way people prefer their embodied actions to be adapted across distributed spaces. Specifically, I respond to Problem Space 3, examining the trade-offs between upholding spatial information and sociocultural norms when the rules that uphold transformation and translation modes of interaction conflict.

The following research questions guide these investigations:

- **(RQ1)** How do current canonical blended realities systems overcome spatial heterogeneity challenges?
- **(RQ2)** How might current approaches for overcoming spatial heterogeneity be characterised in a theoretical framework?
- **(RQ3)** How do people weigh the trade-offs between upholding spatial information and sociocultural norms during remote collaboration?
- **(RQ4)** What should designers and developers consider when incorporating transformation and translation modes into blended realities systems?

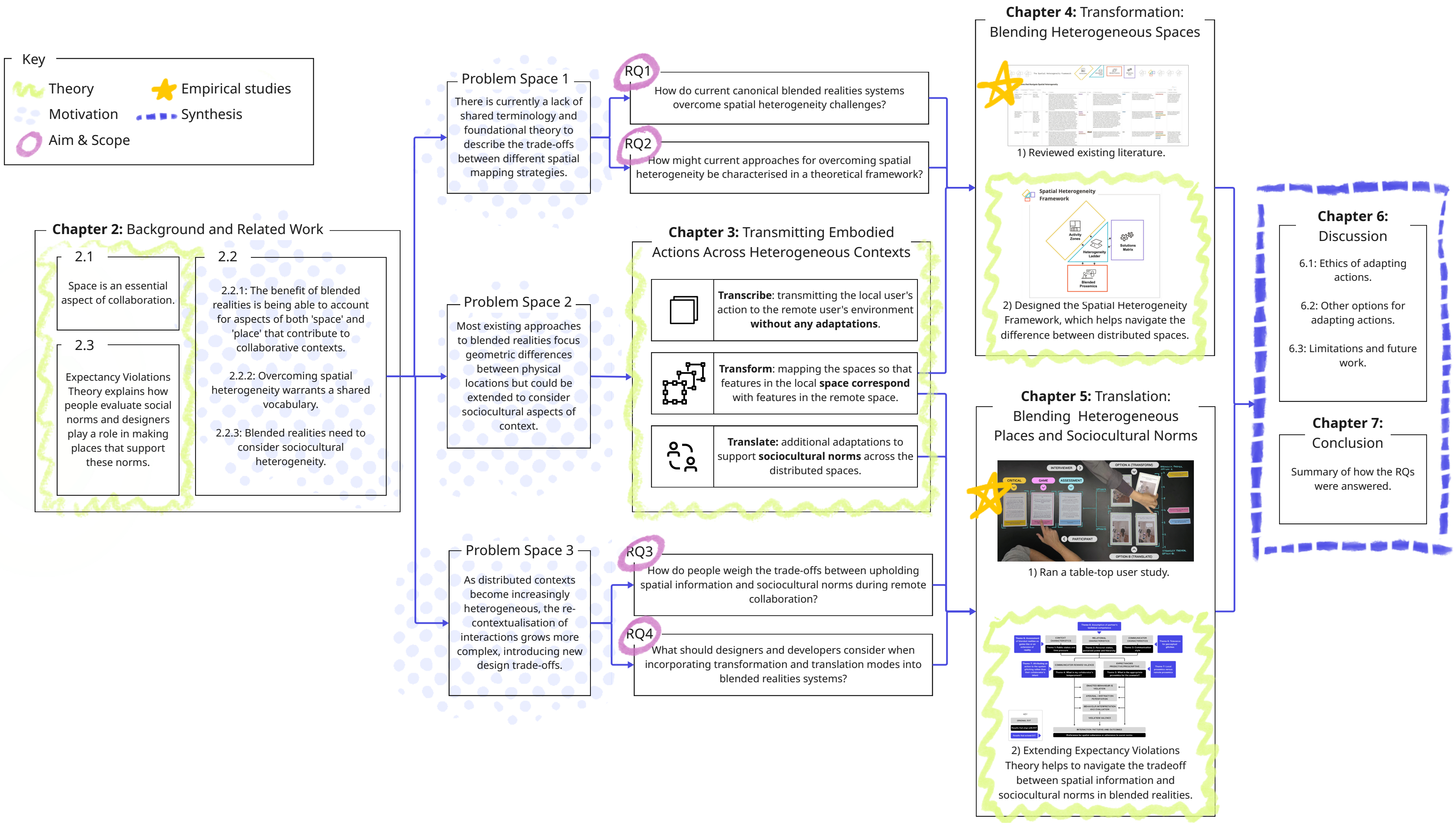


Fig. 1.5 Thesis structure: Chapter 2 gives an overview of the theoretical ideas supporting this thesis, as well as gaps in the current literature that motivate the problem spaces. In response to Problem Space 2, Chapter 3 introduces a taxonomy of three modes for transmitting embodied actions across distributed spaces (transcription, transformation and translation), further explicating the sociocultural element of heterogeneous contexts. This taxonomy structures the empirical studies in Chapter 4 and Chapter 5 that answer the main research questions posed by Problem Space 1 and 3. Chapter 6 and Chapter 7 synthesises the investigation findings, answering the research questions. Chapter 2 reviews prior theoretical work, while Chapter 3, Chapter 4, and Chapter 5 extend these ideas and introduce new theoretical contributions.

This thesis draws on emerging research in blended realities to answer these research questions and build foundational theory for the design of these systems in the future. Its scope focuses on technologies for blended realities across the reality-virtuality continuum (Milgram & Kishino 1994, Speicher et al. 2019, Skarbez et al. 2021) that intentionally blend physical environments to support collaborative interactions across remote spaces. It excludes AR systems where physical space is not actively involved in the collaborative task, but also includes VR systems that anchor users in physical environments, such as substitutional reality (Simeone et al. 2015). By focusing on blended realities, this research contributes to ongoing debates about the role of tangibility and realism in remote collaboration, especially as new tools challenge conventional distinctions of what it means to be co-located.

1.3 Methodology

To answer the identified research questions, I carried out two investigations. The first investigation addresses (RQ1) and (RQ2), and the second focuses on (RQ3) and (RQ4). Each study takes a critical realist stance and assumes there is a real world that exists independently of people's perception, but that an understanding of it is mediated by individual experiences or particular theories being employed (Gorski 2013). I outline this philosophical stance, as well as my methodological strategy, the theoretical constructs I employed, and my positionality, to actively inform the reader how the findings of this thesis should be interpreted.

Across both studies, my interpretation of the results was informed by theories of *embodied interaction*, which proposes that meaning emerges through people's practical engagement with technologies and environments (Dourish 2001); *theory of mind*, which provides a framework for understanding how people put themselves in others' shoes to understand their perspective and experiences (Wimmer & Perner 1983); theories that view culturally informed place-making and practical space-design as interconnected aspects (Dourish 2006, Lawson 2007, Harrison & Dourish 1996); and *externalism*, which views cognition as an active dialogue with external resources, artefacts, and spatial arrangements and goes beyond "skin and skull" (Clark & Chalmers 1998, Newen et al. 2018). Alternatively, I acknowledge my industry experience as a user experience designer and workshop facilitator, which informs my point of view of what "good collaboration" looks like. I have also been researching mixed reality (MR) technologies for two years, which means my understanding of these technologies is likely different from the average research participant. In line with a critical realist perspective, I outline these theoretical concepts and my positionality to clearly articulate the lens I viewed the gathered data through.

In my first study, I developed the Spatial Heterogeneity Framework, based on an analysis of 14 canonical blended realities systems and a workshop with academics specialised in MR. I then validated the framework against 32 other blended realities systems developed between 1995 and 2024, and iterated the framework accordingly. My analysis of these different papers, was informed by a variety of theories in MR, including: interaction proxemics (Mentis et al. 2012), blended interaction spaces (O'Hara et al. 2011), spatiality in MR (Benford et al. 1998), and the reality-virtuality continuum (Milgram & Kishino 1994). I synthesised the information in the identified papers, based on the ontological assumption that the blended realities systems presented by these papers are real. Meaning that, while the theoretical lenses I applied might influence my interpretation of their value, it is possible to reach a shared epistemological view of how these real systems work. Therefore, the resulting Spatial Heterogeneity Framework is grounded in an analysis of how real blended realities systems operate.

In the second study, I carried out a user study with 20 participants, examining the design tension between spatial information and sociocultural norms through a qualitative card-sorting activity. This activity was informed by Comparative Structured Observation (Mackay & McGrenere 2025), counterfactual reasoning (Oulasvirta & Hornbæk 2016) and card sorting methods (Fincher & Tenenberg 2005). These cards presented speculative interaction scenarios involving trade-offs between spatial information and sociocultural norms. In a semi-structured interview, I probed participants to understand how they weighed up these trade-offs in heterogeneous collaboration contexts. I opted for a low-fidelity qualitative approach that gave me flexibility to test specific interactions and a wide scope to probe unexpected factors that contributed to participants' preferences.

Taking a critical realist stance, ontologically, I do not take what was said during the interviews, nor my interpretation of these narratives, as a ground truth. However, I believe that this data can act as an indicator of real phenomena that exist in the world. Although the narratives shared by participants are constructed, they are not random and stem from real cognitive and experiential mechanisms. Epistemologically, there is a need to interpret the data shared by participants and attempt to recognise how this data relates to the real interactions demonstrated by the cards in the study. To do so, I applied Braun and Clarke's (2006) *Thematic Analysis* to systematically sort the data into themes. My approach to the thematic analysis and synthesis of the final results was guided by theories of non-verbal communication and social cognition, including *expectancy violations theory* (Burgoon 1993), *proxemics* (Hall 1966), and Theory of Mind (Wimmer & Perner 1983). I recognise that these theoretical lenses and my own personal experiences influence how I interpreted the data. However, the data is itself real and connected to interactions presented by the real counterfactual cards. While a degree of scepticism toward the intersubjectivity of the participants' narratives and my own interpretation of these is healthy, I believe the presented results point toward a reality of how people interpret and navigate interactions enabled by blended realities in different social contexts.

While a world exists where real actions take place, there can be multiple subjective interpretations of those actions, depending on a person's role, experience, or accepted social norms. I do not assume there is one singular or objective interpretation of embodied interactions in space. Rather, I treat knowledge as shaped by both the observer and the observed. While my philosophical stance acknowledges the materiality of computer programs, actions, and spaces, it also recognises the interpretive, situated nature of the knowledge generated in this research.

1.4 Argument Overview

This thesis presents an overarching argument that to negotiate heterogeneous contexts in blended realities, these systems need to consider both spatial *and* sociocultural aspects of the collaborative environment. However, as the spatial and sociocultural contexts of these environments become more heterogeneous, they require looser coupling, which makes it increasingly difficult to render a coherent shared reality. Addressing these challenges requires careful consideration of when embodied actions, the activity set up, or spatial layout should be adapted. I construct this argument across four parts, shown in **Figure 1.6**.

Part I: Blended realities must carefully consider both spatial information and sociocultural norms when transmitting actions across distributed spaces. In Chapter 3, I demonstrate how blended realities transmit embodied actions across distributed spaces using three modes: *transcribing* actions without any adaptation, *transforming* spatially dependent actions (e.g. redirected pointing), and *translating* actions to preserve sociocultural norms (e.g. rendering avatars within appropriate social distances).

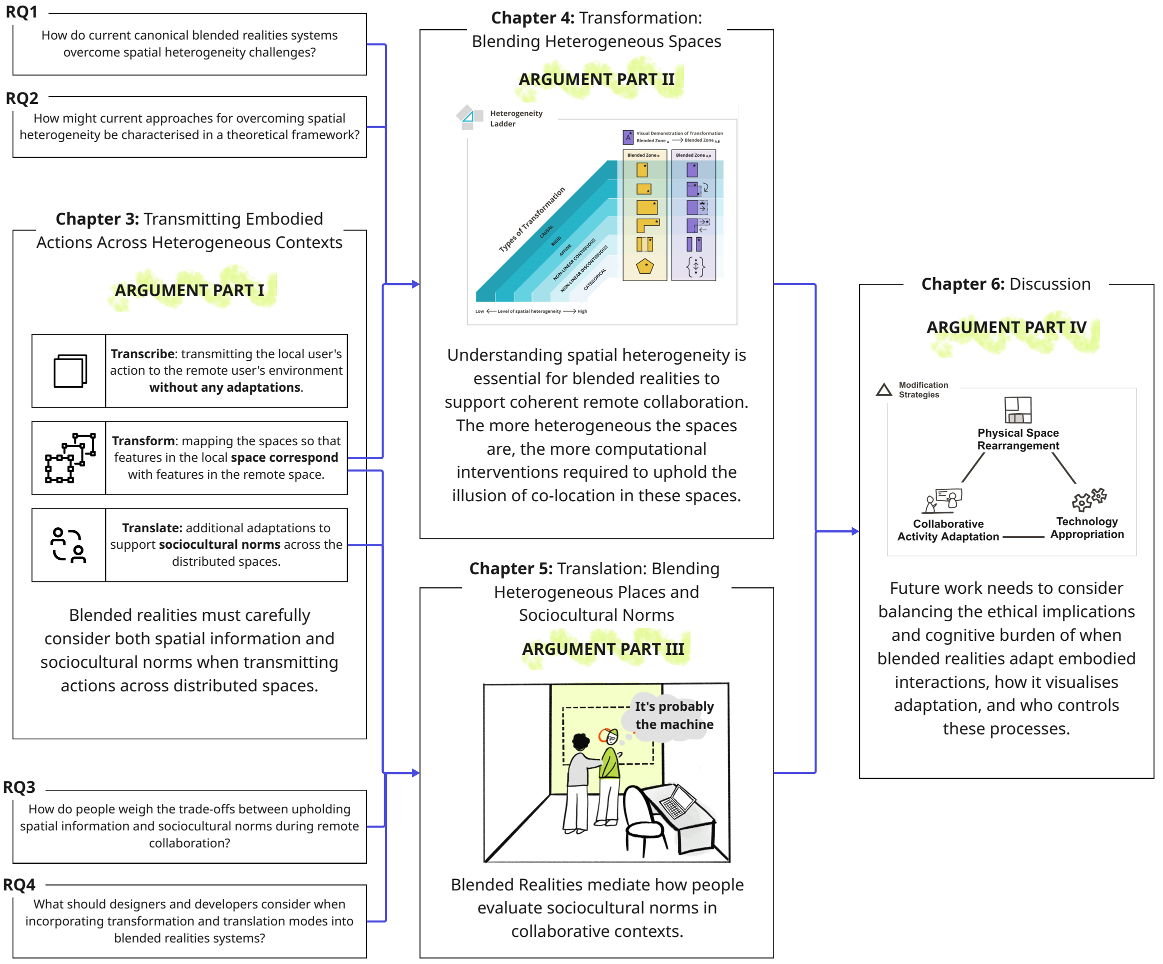
When the distributed locations are dissimilar, transcribing actions from one space to another no longer supports congruent spatial information *or* sociocultural norms. In such cases, collaborators' actions must be adapted using interaction techniques in either the transform or translate mode.

Inspired by the Shannon–Weaver model of communication, I show how "social noise" arises when transform and translate adaptations come into conflict and the intended meaning of an action is distorted. This "social noise" creates design trade-offs, where preserving spatial information (prefacing transformation adaptations) may compromise social norms, whereas upholding social norms (prefacing translation adaptations) may undermine accurate spatial information.

These design trade-offs highlight two epistemologies of context: one treats context as an external set of entities that can be denoted and mapped (Dey 2001), while the other views context as socially enacted and culturally situated (Dourish 2004). I argue that blended realities must navigate both perspectives to create cohesive environments for distributed collaboration.

Part II: Understanding spatial heterogeneity is essential for blended realities to support coherent remote collaboration. The more heterogeneous the spaces are, the more computational interventions required to uphold the illusion of co-location in these spaces. In Chapter 4, I show how spatial heterogeneity—the degree of difference between the spatial layouts of distributed spaces—is a foundational challenge for blended realities. As the degree of spatial heterogeneity increases, so too does the complexity of adaptations required to uphold spatially dependent interactions. However, the field currently lacks a unifying language and theoretical concepts to identify, discuss and compare different solutions for overcoming spatial heterogeneity.

In response, I present the Spatial Heterogeneity Framework, which describes four interrelated components. First *activity zones* define the functional roles of different areas in each collaborator's



RQ1: How do current canonical blended realities systems overcome spatial heterogeneity challenges?

RQ2: How might current approaches for overcoming spatial heterogeneity be characterised in a theoretical framework?

Transcribe: transmitting the local user's action to the remote user's environment **without any adaptations**.

Transform: mapping the spaces so that features in the local **space correspond** with features in the remote space.

Translate: additional adaptations to support **sociocultural norms** across the distributed spaces.

Blended realities must carefully consider both spatial information and sociocultural norms when transmitting actions across distributed spaces.

RQ3: How do people weigh the trade-offs between upholding spatial information and sociocultural norms during remote collaboration?

RQ4: What should designers and developers consider when incorporating transformation and translation modes into blended realities systems?

Understanding spatial heterogeneity is essential for blended realities to support coherent remote collaboration. The more heterogeneous the spaces are, the more computational interventions required to uphold the illusion of co-location in these spaces.

Blended Realities mediate how people evaluate sociocultural norms in collaborative contexts.

Future work needs to consider balancing the ethical implications and cognitive burden of when blended realities adapt embodied interactions, how it visualises adaptation, and who controls these processes.

Fig. 1.6 This thesis makes four main arguments.

physical environment, identifying the “blended zones” that should be mapped together across spaces. Second, the *heterogeneity ladder* characterises how similar or dissimilar these blended zones are, based on the arrangement of fixed, semi-fixed, and mobile features, with higher heterogeneity making it harder to sustain a shared sense of space. Third, *blended proxemics* uses the degree of heterogeneity to identify how proxemic cues (i.e. gestures, orientations, and other spatial interactions) are preserved “for free” by the spatial layout or adapted by the blended realities system to preserve the spatial meaning of these interactions across the dissimilar locations. Finally, the *solutions matrix* compares blended realities solutions that adapt interactions to preserve blended proxemics, demonstrating how blended realities will assume a certain degree of heterogeneity between the different locations’ blended zones.

I argue that the more similar the blended zones are, the more embodied interactions that rely on spatial information are afforded “for free”. Therefore, the more dissimilar the blended zones are, the more computational interventions are required to coherently transmit actions between the different spaces. As the number of blending techniques increases, the more likely it is for perceptual distortions to occur. For example, an avatar's body might be over-extended, crumpled or even break personal space norms, phasing through the local participant. I explicate and discuss these conflicts further in Chapter 5.

Part III: Blended Realities mediate how people evaluate sociocultural norms in collaborative contexts. In Chapter 5, I examine how people navigate the trade-offs between coherently transmitting spatial information and sociocultural norms. I focus on the tensions between transformation, where actions are adapted to preserve spatial accuracy, and translation, where actions are adapted to uphold sociocultural norms. I argue that people's preferences for these will change in particular collaborative situations.

Through a qualitative user study, I show that participants' evaluations of these trade-offs broadly align with Burgoon's Expectancy Violations Theory (EVT) (1993). Participants drew on factors such as their relationship with their collaborator, communication preferences, and the activity environment to assess if certain actions would be socially acceptable. However, their choices were influenced by their awareness of mediation from the blended realities system. In particular, participants often attributed social norm violations to technical limitations, choosing to overlook them as "system glitches".

Based on these findings, I advance two key arguments. First, EVT can be extended to account for how people navigate the trade-offs between transformation and translation adaptations in blended realities, offering a framework for understanding how sociocultural norms are evaluated in collaborative contexts. Second, awareness of system mediation alters how people apply Theory of Mind. Participants were more willing to dismiss sociocultural norm violations as system errors and often struggled to discern whether actions originated from their partner or the technology itself.

Part IV: Future work needs to consider balancing the ethical implications and cognitive burden of when blended realities adapt embodied actions, how it visualises adaptation, and who controls these processes. In Chapter 6, I examine the implications of my findings for designing blended realities that support collaboration across heterogeneous contexts. I argue that at higher levels of heterogeneity, where spaces are loosely coupled, ethical concerns emerge around the possibility of deception. Participants raised concerns about being unaware of how their actions are represented in their partner's space. Conversely, participants felt they were likely to excuse any inappropriate behaviour as technical glitches. This prompts a discussion about when systems adapt embodied actions, how these adaptations are visualised and who or what decides these processes.

I argue that the system adapting embodied actions is not the only way to maintain a sense of co-location. People themselves can adapt the collaborative environment by modifying their activity or altering their spatial layout to sustain coherent interactions. While this shifts some of the burden from the system to the user, it also reduces the need for increasingly complex rules that risk creating conflicts, glitchy experiences and potential deception.

Looking forward, I suggest that future systems could take a hybrid approach, where adaptation is shared between the system and the user. Rather than the system carrying the full responsibility for reconciling heterogeneous spaces, it could also prompt people when changes in layout or activity would improve collaboration. This dual strategy prompts new directions for designing ethically responsible and adaptable blended realities. This thesis lays the groundwork for these conversations, highlighting both the ethical stakes and the opportunities for rethinking how co-location is supported across distributed spaces.

1.5 Contributions

I make the following **theoretical claims**, supported by empirical and artefact contributions:

- To maintain spatially coherent interactions across distributed locations, blended realities systems must make assumptions about the degree of spatial heterogeneity. These assumptions can range across causal, rigid, affine, non-linear continuous, non-linear discontinuous, or categorical mappings.
- As spatial heterogeneity increases, the potential for perceptual distortions also grows, since fewer proxemic cues are afforded “for free” by the physical layout of objects.
- While prior work in blended realities has primarily focused on overcoming spatial heterogeneity, I show that coherent communication also depends on accounting for sociocultural norms across different collaborative contexts.
- I show that an extension of Burgoon’s (1993) Expectancy Violations Theory (EVT) can explain how people assess trade-offs between transformation (spatial information) and translation (sociocultural norms) in blended realities.
- An awareness of potentially divergent interaction perspectives (their own and their partner’s) can obscure Theory of Mind, leading participants to relax their judgments of whether certain behaviours are intentional or unintentional.
- Sustaining translation of sociocultural norms across heterogeneous contexts introduces trade-offs with transparency. Researchers must carefully consider how and when people should communicate the distinction between human intent and technological mediation, as this could influence users’ behaviour beyond the blended realities experience.
- When navigating heterogeneity, participants may resolve challenges in three ways: by relying on a technical solution, by adjusting their physical layout, or by modifying the activity requirements.

Empirical Contributions:

- Validation of the Spatial Heterogeneity Framework through an analysis of 32 papers (1994–2024) describing systems for blended realities.
- Identification of four components that determine how systems overcome spatial heterogeneity: (1) activity zones that different functions of space used in the collaborative activity, (2) assumptions made about spatial difference on the heterogeneity ladder, (3) blended proxemics either afforded by the physical space layout or enabled through blending techniques, and (4) how technical solutions enable these affordances using different blending techniques or by relying on the physical space layout.
- Identification of five themes from semi-structured interview data showing how participants' evaluations of interactions broadly aligned with Expectancy Violations Theory (EVT). These themes describe how participants considered: (1) public stakes and time pressure; (2) personal stakes, power, and hierarchy; (3) communication preferences; (4) predicted collaborator reactions; and (5) their model of appropriate versus actual proxemics.
- Identification of two themes showing how EVT should be extended to account for blended realities: (1) participants' familiarity with technology and their tolerance of glitches, and (2) how awareness of system mediation complicated their Theory of Mind.

Artefact Contributions:

- A taxonomy of action transmission that explains how embodied actions are conveyed across heterogeneous contexts in blended realities through three modes: transcription, transformation, and translation. This taxonomy structures the analysis of design trade-offs when instantiating blended realities in different contexts.
- The Spatial Heterogeneity Framework, a conceptual model grounded in proxemics theory, for identifying heterogeneity problems, mapping potential solutions, and analysing how current systems address them.
- A practical worksheet that guides researchers step by step through applying the Spatial Heterogeneity Framework.
- The concept of blended proxemics, which can be used to categorise the different proxemic-based interactions that support non-verbal communication and are enabled by blending heterogeneous spaces.
- A literature database that categorises 32 systems for blended realities, published between 1994 and 2024, according to the level of spatial heterogeneity they assume.

- An extension of the EVT flow diagram, which provides researchers with a tool to design blended realities systems that predict when people might desire to uphold spatial information or sociocultural norms across heterogeneous spaces.

1.5.1 Publications

The contents of this thesis have been published in (or submitted to) international peer-reviewed journals and conferences.

1. Chapter 3 and 5, including sections of Chapter 2 and 6 are under review at the ACM Conference on Human Factors in Computing Systems (CHI) 2026.
2. Chapter 4, including sections of Chapter 2 and 6 have been published at ACM CHI 2025 (Wong et al. 2025).

Chapter 2

Background and Related Work

In this chapter I argue that space is not a backdrop for collaborative activities, but an active medium involved in cognition and communication. In Section 2.1 I draw on theories of embodied interactions (Dourish 2001) and externalism (Newen et al. 2018) to position mixed reality (MR) as a compelling solution for remote collaboration. I show how transmitting non-verbal embodied actions that rely on space—such as proxemics (Hall 1966), facing formations (f-formations) (Kendon 1990), and deictic cues—MR provides the means for more intuitive remote collaboration (Ens et al. 2019). However, prior work cautions against mimicking co-located collaboration (Hollan & Stornetta 1992). Instead, in this section, I argue that researchers should investigate how MR interaction techniques can be strategically applied to preserve or reconfigure embodied actions in moments that matter. Identifying these moments and understanding how embodied actions should be adapted requires further investigation.

A core challenge for MR is coherently transmitting embodied actions across heterogeneous contexts. In Subsection 2.1.3 I explain this problem and position blended realities, an emerging subset of MR solutions, as a way to overcome differences between distributed environments. Following this, in Section 2.2 I demonstrate that while there has been advancements in blended realities, current systems lack a unifying language and conceptual framework, which makes it difficult to discuss and compare solutions.

Much of the research to date has focused on heterogeneity between environments as a spatial challenge. In Section 2.2 I also present research that supports an argument for heterogeneity being extended to consider how sociocultural norms are represented across the different distributed locations. However, considering both space and sociocultural norms across heterogeneous contexts creates trade-offs when rules that uphold these aspects conflict. I explicate the opportunity to understand how people weigh up these trade-offs in different collaboration contexts and the factors that influence these decisions. These problem spaces motivate the overatching aim to *investigate how embodied actions are transmitted across heterogeneous contexts in blended realities, and the design trade-offs this process entails.*

Finally, prior research raises both promising opportunities and ethical challenges for designing blended realities systems. In Section 2.3 I show that while these systems can preserve and adapt embodied actions, heterogeneity across contexts introduces tensions between spatial information and sociocultural norms. Judee Burgoon's (1993) expectancy violations theory (EVT) provides a framework to understand how people evaluate these trade-offs. However, it also exposes ethical concerns when system adaptations obscure intent or mask inappropriate behaviours. Prior work points to potential ethical responsibilities when making design decisions that inform how cognition and collaboration is supported across distributed spaces.

2.1 Space is a Medium for Embodied Actions and Cognition

This section demonstrates the importance of space during collaboration, as support for communication and cognition. Although immersive technologies reintroduce spatially dependent interactions, most systems rely on uniform layouts. There is an opportunity to address heterogeneous contexts, using blended realities to adapt embodied actions and maintain coherent non-verbal communication across dissimilar environments.

2.1.1 Mixed Reality Enables Embodied Interactions During Remote Collaboration

Space is an important medium for interpreting embodied interactions. In Human-Computer Interaction (HCI), theories of embodied interaction draw on phenomenological philosophy and cognitive science to propose that human perception and meaning-making arise through physical engagement with the world (Dourish 2001, Robertson 1997, Rogers 2004, Marshall & Necker 2013). This physical engagement is inherently spatial, and our understanding of others' actions is shaped by how bodies move in relation to each other and the surrounding environment. For instance, when a stranger enters a classroom, they can quickly infer that students seated in clusters likely know each other and that the person standing at the front of the room is the teacher (Buxton 2009). If the teacher gestures toward a student with a raised hand, the class knows it is that student's turn to speak, and they should direct their attention accordingly. In such scenarios, spatial arrangements and embodied cues provide context for interpreting social roles, intent and appropriate responses.

Although many scholars agree that embodied interactions and space are important aspects of communication (Dourish 2001, Marshall & Necker 2013), there are varying perspectives on how "space" should be supported during remote work (O'Hara et al. 2011, Johnson et al. 2021). Most notably, Hollan and Stornetta's (1992) influential work on *Beyond Being There* argues that instead of striving to mimic face-to-face interaction, remote systems should utilise the novel forms of connection that digital media can provide. However, a variety of scholars maintain the importance of preserving "spatial geometry" to facilitate embodied gestures (Heath & Luff 1993, O'Hara et al. 2011, Benford et al. 1998, Buxton 2009). In particular, Benford et al. (1998) suggest researchers should evaluate the cost and benefit of spatial information in terms of what it enables people to perceive and do. This position reinforces Heath and Luff's (1993) observational work, which showed how spatial needs vary according to task demands. These perspectives suggest that supporting spatiality in remote systems is not a binary choice between replication and abstraction. Rather, it requires a careful evaluation of the communicative benefits spatial representations offer, relative to the design effort and cognitive demands they impose.

Further emphasising the importance of space, externalist perspectives of cognition argue that familiar spatial affordances are important for supporting cognitive processes. For example, Clark (2010), building on Dourish's (2001) work on embodied interaction, refers to the phenomenological notion of treating space and tools as "transparent equipment". In familiar environments, such as classrooms or offices, people can "see through" learnt interactions with objects like whiteboards, desks, and chairs to focus directly on the collaborative task, rather than the mechanics of the interaction. This is why in virtual reality (VR) the absence of intuitive embodied interactions for manipulating the familiar environment often makes participants feel like *visitors* rather than *inhabitants* of the space (Clark 2010). In contrast, observations of co-located work reveal how the interplay of objects, people and space operates as external resources that participants can rely on to coordinate action and sustain mutual understanding (Mentis et al. 2012). This resonates with Suchman's (1987) theory of *situated action*, which shows that meaning-making is not pre-scripted but emerges through dynamic engagement with the spatial resources available at the moment of interaction. However, when collaborators are distributed, they lose shared access to many spatial affordances and familiar embodied actions, disrupting the external supports for collaboration.

Mixed reality (MR) provides a compelling opportunity to investigate design options for enabling non-verbal embodied cues across distributed spaces. Unlike other remote collaboration tools, such as VR, collaborative virtual environments or video conferencing, MR allows participants to access the natural affordances of their own space while simultaneously being embodied as an avatar or projected into their collaborators' remote locations. Using various degrees of virtuality (Milgram & Kishino 1994), with technologies like head-mounted displays (HMDs) (Orts-Escolano et al. 2016), projection mapping (Pejsa et al. 2016), or spatially configured displays (Buxton 2009), interactions can be remapped to correspond with the remote participants' physical spaces. This means spatially supported interactions, like moving toward a desk or other feature of the room to signal intent, can be coherently mapped across the different locations. The space is not treated as a backdrop but actively involved in external cognitive processes and social navigation (Klemmer et al. 2006). In this way, MR is particularly well-suited to explore how embodied interactions can be shared across physically separate spaces as a site for cognition.

2.1.2 Non-verbal Embodied Actions are Important Features of Collaborative Work

Theories from computer-supported collaborative work (CSCW) and interpersonal communication emphasise that effective collaboration requires more than just dialogue. It demands a *shared* *sense of space* where participants can engage in non-verbal embodied actions (Dourish 2001, Lawson 2007). Embodiment and mobility are critical components to enable these interactions and the dynamic flow of collaborative work (Zubek et al. 2022). When working together, individuals instinctively use their bodies and the physical environment around them, arranging themselves in various orientations and leveraging objects to support communication (Kendon 2010, Luff & Heath 1998). These non-verbal cues facilitate contextual understanding, rapid turn-taking, and interactions with both physical and virtual objects (Tian et al. 2023, Gerhard et al. 2004). To design these spatial dynamics, HCI researchers often draw on theories of non-verbal interaction such as proxemics, facing-formations (f-formations), and deictic gestures.

The term proxemics was first coined by anthropologist Hall (1966) to explain how personal space varies depending on people's relationships and interaction context. These zones—intimate, personal, social, and public—each carry distinct social connotations that shape non-verbal communication. For instance, one might infer that a couple standing closely together, at an intimate distance, is romantically involved. Mentis et al. (2012) apply the concept of proxemics to study the interplay of people, objects and space during co-located collaborative scenarios. Their research showed that people integrated digital objects into their *interaction proxemics*, the utility of these objects and collaborators' job roles informing their spatial relationships. Williamson et al. (2022) demonstrated the interconnectivity of proxemics and embodiment, showing that people are more likely to follow these zones of interpersonal distance when they are using an immersive VR headset rather than a desktop display. While proxemics is often applied in human-to-human interactions, Greenberg et al. (2011) also adapted the concept to propose *Proxemic Interactions*, where computer devices respond adaptively to the proximity of other devices and people. For example, triggering an action only when another device enters the "social zone". O'Hara et al. (2011) go further, applying concepts of proxemics and f-formations to develop Blended Interaction Spaces that thoughtfully combine physical spaces with specially configured digital displays. In these Blended Interaction Spaces, proxemics seamlessly transgress physical and digital environments.

Facing-formations (f-formations) is a related theory of non-verbal behaviour that has informed the study and design of embodied interactions in collaborative systems. Kendon's (1990) theory of f-formations shows how groups naturally organise their bodies in space to indicate shared focus and social involvement. These formations have been studied as a means to understand the way people gather socially (Marshall & Necker 2013, Williamson et al. 2022, Setti et al. 2015, Tong et al. 2016), design collaborative interactions with digital devices (Marquardt et al. 2012, O'Hara et al. 2011, Wong, Sánchez Esquivel, Leiva, Grønbæk & Velloso 2024, Grønbæk et al. 2017) or understand how people engage virtual agents Hedayati et al. (2020) and robotics (Barua et al. 2024) in collaborative systems. Goffman (2008) also studied f-formations, suggesting that body orientation between participants could communicate intent, attention and turn-taking interactions.

When considering non-verbal communication, an obvious attribute is the ability to point or gesture toward objects. Originating from linguistics, deixis refers to the way speakers identify items of interest, for example, using utterances like "that one, there" (Hanks 2009). This deictic language can be accompanied or substituted with embodied actions like pointing (Goodwin 2000), which plays a foundational role in communication and is established early in human development (Cochet & Vauclair 2010, Pechmann & Deutsch 1982). Deictic cues, such as pointing, gaze shifts, and head or body orientation (Levinson 2006), are often enabled in digital systems to help people reference shared objects and establish mutual understanding during collaboration (Gutwin & Greenberg 2002).

These non-verbal embodied actions free people from the need to construct sentences to identify an object of interest or convey social information (Pechmann & Deutsch 1982, Wong & Gutwin 2010), supporting group coordination and task completion (Kiyokawa et al. 2002, Bente et al. 2008, Fussell et al. 2000, Johnson et al. 2021, Abdullah et al. 2021). In distributed collaboration, people are physically separated, and many of these interactions must be artificially recreated through technology (Grønbæk et al. 2017, Wong, Sánchez Esquivel, Leiva, Grønbæk & Velloso 2024). The challenge lies in creating systems that adequately preserve these non-verbal cues and convey a sense of shared space.

2.1.3 Blended Realities Transmit Embodied Actions Across Heterogeneous Spaces

The potential of immersive technologies to revolutionise distributed collaboration lies in their ability to preserve embodiment and mobility across immersive spaces. In particular, Ens et al.'s (2019) recent review shows how MR research has progressed from “a focus on solving initial technical challenges in MR toward more meaningful investigations of collaboration" (2019, p. 89). At a broader level, an increasing number of studies have demonstrated the benefits of immersive technology for distributed collaboration, principally through the sense of social presence and embodiment provided (He et al. 2020, Sereno et al. 2022).

Initially limited by technical capabilities, avatars can now enable people to perform gestures, walk, make eye contact, and even experience haptic feedback (Oh Kruzic et al. 2020). Avatars that support dyadic interactions through bodily and facial expressions closely reach the quality of social exchanges in face-to-face interactions (Maloney et al. 2020, Smith & Neff 2018, Sanaei et al. 2023) and sometimes exceed that of non-immersive collaboration interfaces like Zoom (Oprean et al. 2018, Maloney et al. 2020, Steinicke et al. 2020, Sanaei et al. 2023, Abramczuk et al. 2023).

Social VR apps have been employed to conduct distributed workshops (Williamson et al. 2021), sparking enthusiasm about enhancing social interactions beyond merely replicating physical ones (McVeigh-Schultz & Isbister 2022). However, despite these advancements, VR shows significant drawbacks when used for distributed collaboration. While completely cut off from the real world, people are unable to interact with physical objects that may be useful for collaboration, such as physical models, prototypes, or notes (Grønbæk et al. 2017, Kiyokawa et al. 2002). Additionally, VR's teleportation-based navigation can disorient people, disrupting spatial continuity and mental mapping during discussions (Freiwald et al. 2021). This friction is understandable given the advancement in theories of cognition that show the physical world as an extension of the mind (Clark & Chalmers 1998). A detachment from the physical world in VR introduces an extra layer of interactions that people are required to learn in order to make full use of their external environment as a site for cognition (Clark 2010).

Various MR, AR, and media space research has investigated how to overcome the limitations of a fully virtual environment by integrating digital interactions with the physical world, enabling more natural and intuitive exchanges. Previous work, such as *Holoportation* (Orts-Escolano et al. 2016) or *Room2Room* (Pejsa et al. 2016), create the illusion that the remote user is present and can interact with local space by showing their hologram in a room that is identical to the remote user's one. Similarly, media spaces that utilise displays as "portals" like *Cisco Telepresence* or *HP Halo* (O'Hara et al. 2011) provide the illusion that two remote environments are seamlessly connected, blending them into a single extended space. Another approach has been to use asymmetric systems (McGill et al. 2015, Irlitti et al. 2023), letting the remote user join the local one via VR while the local user keeps sight of the real world and sees them in AR.

Although these examples blend physical-digital spaces and create a convincing sense of co-presence, many of them lack the versatility to accommodate collaboration settings that embrace the complexities inherent in real-world spaces. They either require collaborators to have identical physical environments (O'Hara et al. 2011, Irlitti et al. 2016, Orts-Escolano et al. 2016, Pejsa et al. 2016) or are asymmetrical and require one user to "leave" their physical space (McGill et al. 2015, Irlitti et al. 2023). While these immersive environments have valid use cases (Billinghurst & Kato 1999), they stop short of facilitating complex collaborative situations that involve the dynamic reconfiguration of people, objects and space (Ens et al. 2019, Wong, Sánchez Esquivel, Leiva, Grønbæk & Velloso 2024). Since distributed collaborators likely wish to leverage elements unique to their local environments, there is a pressing need for solutions that can adapt to the heterogeneity of these physical spaces while supporting collaboration.

Building on these possibilities, researchers have begun to explore how physically dissimilar spaces can be experienced as one, to support collaboration in MR. This emerging practice, referred to as *blended realities*, uses MR to transmit remote actions to the local collaborator's environment (Grønbæk et al. 2023, 2024). These could include technologies situated anywhere on the reality-virtuality continuum that deliberately blend physical and digital environments (Milgram & Kishino 1994). In this thesis I exclude augmented reality (AR) experiences that do not depend heavily on the physical environment but include virtual reality (VR) experiences that are anchored in physicality, such as substitutional reality (Simeone et al. 2015). Historically, MR systems relied on pre-configured, geometrically similar environments to maintain coherence across sites (Grønbæk et al. 2024). However, with advances in HMD display technology and spatial mapping, new work is investigating how to meaningfully connect spaces that were never designed to be compatible (Sra et al. 2018). The vision of blended realities being to allow people to work in and access their local physical space, as though they are co-located with their remote partner/s.

2.2 Blended Realities Systems Must Consider Heterogeneous Contexts

Heterogeneity between remote spaces challenges the illusion of co-location by breaking the common frame of reference that collaboration relies on (Heath & Luff 1993). Differences in spatial layouts can distort embodied cues such as deictic cues, proxemics, and f-formations, making familiar interactions feel strange and unfamiliar (Wong, Grønbæk & Velloso 2024). At the same time, differences in sociocultural norms mean that an adapted action can carry conflicting meanings across contexts, turning actions that seem appropriate in one space into an awkward situation that breaks sociocultural norms in another. Together, both spatial and sociocultural aspects of heterogeneity complicate how collaborators interpret each other's embodied actions and test the limits of systems designed to preserve a shared sense of space. This section examines how spatial and cultural heterogeneity disrupt co-location, explores the trade-offs they introduce, and surveys design strategies that attempt to mitigate these challenges.

2.2.1 Heterogeneous Contexts Involve Both Spatial and Sociocultural Challenges

The aim of many immersive systems for remote collaboration is to support the feeling of co-location across distributed spaces. When collaborators have access to a shared external world, they can focus on their joint tasks rather than on the mechanics of operating across different environments. This reflects phenomenological accounts of "transparent tools" (Clark 2010), where technologies recede into the background so that attention remains on the action itself. For example, when a person is using a pen, they are not thinking about the pen but rather what is being drawn. Similarly, by capturing measures like spatial presence (the feeling of "being there"), co-presence (the feeling of "being there together") and social presence (the sense of "being together with another") (Tran et al. 2024), empirical studies in collaborative immersive technology aim to understand what helps people go beyond noticing the system to be fully present in the simulation (Irlitti et al. 2023, Kim et al. 2014, Casanueva & Blake 2000, Brown et al. 2003).

A major challenge for enabling a feeling of co-location is the heterogeneity (i.e. difference) between the various physical spaces involved in the remote collaboration. Heath & Luff (2000) argue that these differences in spatial orientation and configuration create "incongruent environments for interaction" (2000, p.198). Without a common frame of reference it is difficult to engage in effective communication and facilitate intersubjective understanding of interaction (Gaver et al. 1993, O'Hara et al. 2011). Embodied actions, such as proxemics, f-formations and deictic cues, become more challenging to support (Grønbæk et al. 2023). For example, imagine Alice's whiteboard is to the right of her desk but Bob's whiteboard is to the left of his. When either Alice or Bob point at their local whiteboard they will appear to point at a blank wall from their remote partner's perspective. Wong, Grønbæk & Velloso (2024) term this situation the *Jamais Vu Effect*; when the spatial coherence of distributed spaces reach their limits and familiar collaborative situations suddenly feel strange and unfamiliar. Instead of fluidly shifting their attention to their whiteboards, Alice and Bob need to stop and clarify what the other person means. This difference between spaces causes a break in the illusion of co-location that immersive systems aim to uphold.

To support embodied actions, disrupted by the heterogeneity of distributed spaces, researchers have developed various techniques to reconstruct these interactions across the mapped environments. These techniques include redirected gaze, pointing or walking (Yoon et al. 2021, Wang et al. 2022, Sousa et al. 2019), animated transitions between functionally equivalent zones (Grønbæk et al. 2023, Fink et al. 2022, Wang, Kim, Panda, Ofek, Franco & Won 2024), and modifications to the avatar's scale or orientation (Grønbæk et al. 2024, Yoon et al. 2022).

While it is possible to reconstruct embodied actions across distributed spaces, these involve design decisions that require understanding the collaborative context. As the collaborative environment's spatial requirements increase, maintaining an objective world view requires more implementation and system effort to maintain (Benford et al. 1998). This brings into question *how much* spatiality a system should support if it requires considerable compute power to do so. For instance, in the previous example, if Alice and Bob are focused on writing at a desk, the system may only need to support embodied interactions around the desk. By contrast, if they are diagramming together, the system might adapt to support interactions around their whiteboards. More complex activities, such as working at the desk *and* whiteboard would need a more dynamic solution. When the spatial geometries differ significantly, this type of complex activity can introduce distracting perceptual distortions (Congdon et al. 2018). Design decisions must consider the benefits of supporting embodied actions and the costs of perceptual distortions due to additional computational interventions (Wong, Grønbæk & Velloso 2024). The decisions behind these trade-offs are informed by the context in which the collaboration takes place.

One aspect of context accounts for what exists in the external world; the identification and relationship between people, objects and space in a given situation. Dey (2001) defines context as:

> "Any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves" (2001, p.5).

This definition of context is particularly important for semantically mapping distributed spaces so that embodied actions, such as pointing, carry the same meaning from one space to another. This involves the process of identifying and matching spatial features based on their function or meaning, regardless of their exact physical form or location. For example, to enable redirected pointing, a blended realities system might identify the target object in the local user's environment (e.g. the professor's whiteboard, **Figure 2.1**), then calculate the spatial offset between this target and the remote user's layout. The local user's avatar is then rendered to point directly at the semantically mapped object in the remote space (e.g. the the student's wall, **Figure 2.1**). This mapping can be tightly or loosely coupled. In **Figure 2.1**, the student might be using their wall as a makeshift whiteboard, which is therefore loosely coupled with the professor's actual whiteboard. While the student's wall is not literally a 'whiteboard', it has the same semantic function as the

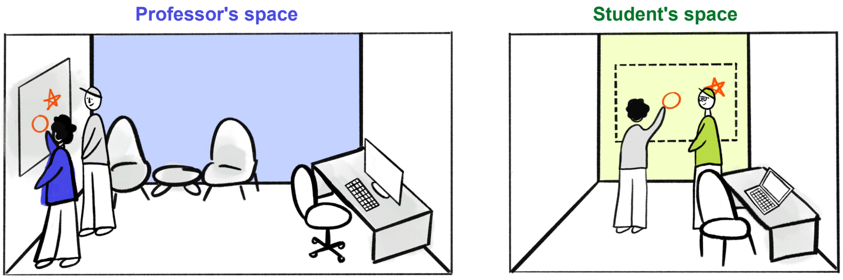


Fig. 2.1 The professor's *actual* whiteboard and the student's wall are loosely coupled.

professor's actual whiteboard, and therefore, the system has been developed to redirect the professor's gesture toward this wall. A contextual understanding of *what exists* in the space enables the system to preserve the intent behind collaborators' embodied interactions, ensuring they remain interpretable to remote collaborators even when spatial layouts differ significantly.

However, context is not solely made up of entities that exist in the world. As Dourish (2004) argues, "context cannot be a stable, external description of the setting in which activity arises. Instead, it arises from and is sustained by the activity itself" (2004, p.23). This stance aligns with with externalist accounts of cognition, where the physical world is part of a person's dynamic real-time cognitive feedback loop (Clark & Chalmers 1998). By this account, context is more than a *denotation of mapped objects* but instead *carries connotations of embodied action relative to the task at hand*. For example, Suchman's (1987) theory of *situated action* explains how the phrase "that's brilliant!"' might be sarcastic in one context, where a photocopier is malfunctioning, or a positive exclamation of success in another. In another example, Buxton (2009) explains how a person's role might change how they interact in a space. As he explains, "it would be very unusual for a stranger or someone with whom I was not working closely, or did not know, to stand [behind my desk]" (2009, p.220). In this space, Buxton's role as a professor culturally influences what appropriate interactions might look like. It therefore stands that context is made up, not just of the entities in space, but also through activities taking place that involve specific people, external conditions, internal expectations and cultural norms. These are the activities that turn a space into a culturally informed *place* with particular social expectations and norms (Harrison & Dourish 1996, Lawson 2007).

These different definitions of context play out in conversations that compare space and place; one focusing on the denotation of entities (Dey 2001) or objective spatial architecture and the other on how those objects are culturally appropriated to create places for collaboration (Dourish 2004). As Harrison and Dourish (1996) explain, "space is the opportunity; place is the understood reality" (1996, p.67), where space can be designed or invented by HCI researchers but place is culturally defined by how people appropriate these technologies and incorporate them into social activities.

This relationship between space and place is complex and should not be taken as unilateral. As Dourish (2006) later established, designers also use place-ness to inform the way they design spaces. This dynamic is exemplified in architectural disciplines, where the designer of a space might create it with a certain place-ness in mind (Lawson 2007). For instance, a state library might feature high ceilings and marble or wooden surfaces to evoke contemplation and quiet study, whereas a nightclub typically has lower ceilings, compact layouts, and dim lighting to encourage social intimacy (Lawson 2007). While the people who inhabit these spaces turn them into places for socially informed activities, the designer creates them with place-ness in mind.

Similarly, *blended realities* requires designing collaborative spaces, that consciously recognise the places established in different collaborative contexts. As such, in this thesis I examine both spatial *and* sociocultural aspects of heterogeneous environments. I unpack how these two dimensions can compliment and contradict each other.

2.2.2 Overcoming Spatial Heterogeneity Warrants a Shared Vocabulary

A growing body of research uses MR to help distributed collaborators seamlessly access digital and physical affordances for dynamic, collaborative work (Ens et al. 2019). This thesis explores approaches for *blending* physical spaces, despite differences in the spatial layout of the distributed physical environments, which I call *spatial heterogeneity*.

To cope with spatial heterogeneity, researchers have developed prototypes that demonstrate various approaches to mapping distributed spaces. These systems go beyond simply connecting remote collaborators and ***blend*** **physical elements of the collaborators' heterogeneous spaces, such that fixed, semi-fixed, or mobile objects (Marquardt et al. 2012) and the space around them are mapped to enable *proxemic cues* across distributed locations (Grønbæk et al. 2023).**

Solutions for blending heterogeneous spaces can be grouped by how dissimilar they assume the distributed spaces will be, i.e. the level of heterogeneity. These clusters inform the Spatial Heterogeneity Framework, and in Chapter 4 I will use these literary examples to explain how the framework can be used to compare and evaluate existing systems.

At the lowest level of heterogeneity, distributed spaces can be inextricably linked together by their very matter. An important example is Ivan Sutherland's (1965) fictitious Ultimate Display, where "a chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal" (1965, p.2). No current blended realities system fully enables this vision, but works that use robots to move physical objects in a remote space take steps towards it (Sakashita et al. 2023, Follmer et al. 2013, Barden et al. 2012, Brave et al. 1998, Kim, Kim, Choi & Woo 2024).

Some solutions assume distributed spaces are similar and can be aligned by simply rotating one onto the other. These gain most non-verbal cues "for free" by using pre-aligned physical spaces across two locations (Orts-Escolano et al. 2016, Pejsa et al. 2016). For example, Orts-escolano et al.'s (2016) *Holoportation* is a "3D teleconferencing system that allows both local and remote users to move freely within an entire space and interact with each other and with objects" (2016, p. 742) . In their study, a local and remote participant sat across from each other, using tables and chairs of similar size and position. Because systems at this level assume a high degree of spatial similarity, people cannot physically move the objects in their remote partner's space without breaking the alignment.

Other solutions apply simple transformations to gain consistent proxemics across different sized spaces or surfaces (Grønbæk et al. 2024, Herskovitz et al. 2022, Huang & Xiao 2024). This can be illustrated with Grønbæk et al.'s (2024) *Blended Whiteboard*, a "…collaborative whiteboard that allows two remote users to work in front of a local physical whiteboard [where] each physical whiteboard can have a different size" (2024, p.5). In this case, the system assumes that the whiteboard surfaces will differ in size and orientation. However, it also assumes the distance between the user and the whiteboard will remain constant across the distributed spaces and will not need to be warped to maintain continuity across the distributed spaces.

There are solutions that assume the different spaces will require dynamic remapping to support proxemics between and within the spaces between objects (Higuchi et al. 2015, Congdon et al. 2018, Hoppe et al. 2021, Jo et al. 2015). However, this kind of dynamic remapping can introduce certain anomalies, like avatar bodies becoming unusually warped. For example, Congdon et al. (2018) present a system that remaps the layout of one space to another, preserving the relative placement of objects and the space between these in each location. This dynamic remapping of the objects' layout transforms the spatial proxemics from one location to another. However, additional technical solutions are required to preserve the gestures and locomotion of avatars, to avoid their rendered 3D body being uncannily warped by the mapping technique.

At another level, some solutions assume the different locations may have immovable or complex spatial features that cannot be overcome by warping the space. Instead, these approaches splice the space into different areas, removing complicated spatial features and transitioning the user's avatar seamlessly from one defined area to another (Wang et al. 2022, Grønbæk et al. 2023, Fink et al. 2022, Wang, Kim, Panda, Ofek, Franco & Won 2024, Yoon et al. 2022, 2021, Kim, Kim, Shin & Woo 2024). For example, in Fink et al.'s (2022) system, the distributed spaces are spliced into "re-location" areas. The avatars transition between each "re-location" area by fading in and fading out. Limiting the spaces in such a way preserves the proxemic-based interactions inside the "re-location" spaces while overcoming the need to blend difficult parts of the different locations that are not integral to the collaborative task at hand. However, transition animations between the spliced-up areas can be confusing for the collaborators (Wong, Grønbæk & Velloso 2024, Grønbæk et al. 2023), and interactions between the designated areas will not be transmitted to the remote space.

Finally, there are some systems that only rely on semantic (and no spatial) mapping between task objects in two distributed locations (Johnson et al. 2023, Yang et al. 2024, Kang et al. 2023). To demonstrate, Johnson et al.'s (2023) system, *Unmapped*, provides the scaffolding for a future where mobile objects can be mapped on-to-one, regardless of the space between them. At present their system requires one user to be in VR, however it is not a stretch to imagine a future system where both collaborators can be in blended realities. To the best of my knowledge, there are no full MR

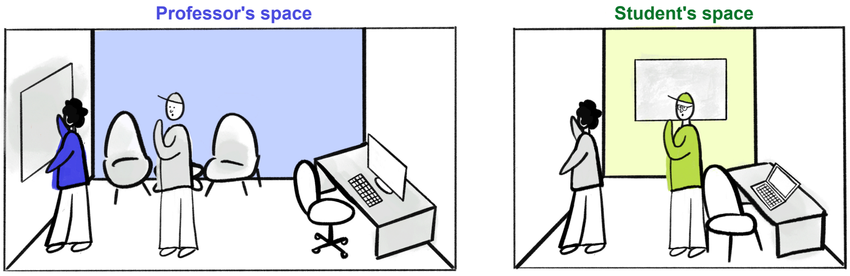


Fig. 2.2 The spaces have not been sufficiently blended. When the professor points at their whiteboard, the student sees them point at a blank space.

systems that enable this type of interaction. If we pretend all interactions are in blended realities, then mobile objects in the blended zone (e.g. cooking utensils) can be mapped one-for-one.

When a system reaches its limits for accommodating spatial differences, perceptual challenges can disrupt proxemic-based interactions. For instance, a recent study found that dynamic transformations can create perceptual distortions in the remote partner's rendered actions (Wong, Grønbæk & Velloso 2024). If spaces are insufficiently blended, interactions may appear misaligned, as seen in Figure 2.2 , where the professor appears to incorrectly point in space from the student's perspective due to improper avatar rendering. To address these distortions, various techniques such as deixis warping (Congdon et al. 2018), redirected walking (Wang et al. 2022), spatial segmentation (Fink et al. 2022), or teleportation (Yoon et al. 2022) have been employed. These "blending techniques" adapt distributed spaces to enable coherent proxemics.

With a growing body of available research solutions for blending heterogeneous spaces, we are faced with the challenge of understanding the differences between these solutions. It is difficult to understand what people stand to gain and lose in terms of non-verbal behaviour and proxemics, raising the need for a taxonomy and consistent vocabulary. Prior taxonomies (Benford et al. 1998, Ens et al. 2019, Schäfer et al. 2022, Sereno et al. 2022) and frameworks (Williamson et al. 2021, 2022) articulate key dimensions of collaborative immersive environments. However, none of these apply specifically to the problem of spatial heterogeneity. As Ens et al. (2019) explain in their recent literature review, "existing frameworks for describing groupware and MR systems are not sufficient to characterize how collaboration occurs through this new medium" (2019, p. 92). Enabling collaboration across heterogeneous spaces is a strong example of collaboration in blended realities that is yet to be supported by an underlying conceptual framework.

2.2.3 Blended Realities Must Consider Sociocultural Heterogeneity

While many blended realities systems rely on semantic mapping to enable embodied actions like proxemics, f-formations or deictic cues (Grønbæk et al. 2023, Yoon et al. 2021, Fink et al. 2022), these approaches do not fully account for sociocultural norms. For example, in **Figure 5.2**, if the system renders the student's avatar so they sit in the semantically equivalent chair in the professor's office, this makes sense from a purely functional, spatial alignment perspective. However, from the professor's perspective, the student's avatar appears to sit in their personal desk chair, rather than the guest seat, making the interaction socially awkward. The trouble is, the student believes they are sitting in their own socially appropriate office chair, unaware of invading the professor's personal space. Imagine unwittingly sitting in your boss's chair (how embarrassing!) or your student's intended seat (what an awkward power play!). This disconnect arises because sociocultural norms are contextually and culturally situated and cannot always be calculated using semantic mapping.

The non-verbal communication literature suggests that it is common for non-verbal behaviours to encode layers of meaning that can be interpreted differently depending on the sociocultural context (Hall & Knapp 2013, Burgoon, Floyd & Guerrero 2016). For example, proxemic norms vary based on both situational, cultural, and interpersonal factors (Burgoon 1993). Standing at an intimate distance from a newly introduced colleague may be perceived as uncomfortable or inappropriate, while standing too far from a romantic partner could be interpreted as aloof or emotionally distant (Hall 1966). Interpersonal communication literature shows that cultures perceive appropriate social distances differently (Hall 1966, Burgoon, Floyd & Guerrero 2016). Alternatively, the type of deictic gesture can also change depending on a complex mix of social roles, cultural expectations and emotional cues (Kendon 2004). These embodied actions go beyond solutions that use semantic mapping and follow much more nuanced rules with contextual interdependencies (Dourish 2004).

Systems for blended realities do not yet move past semantic mapping to understand how sociocultural norms can be translated between distributed spaces. This is an obvious gap in the literature, since various studies of human interaction in immersive environments show that people intuitively follow social norms like proxemics (Llobera et al. 2010, Williamson et al. 2022, Bailenson et al. 2001), object possessiveness (Poretski et al. 2018) and f-formations (Wang, Miller, Queiroz & Bailenson 2024). There are also some norms unique to immersive environments that are yet to be fully explored. For example, body crumpling, where the body tracking becomes distorted and the person's avatar is strangely rendered, or embodiment violations, when a person phases through someone else's avatar (Akselrad et al. 2023).

Prior work has shown that encoding social and spatial rules can come into conflict. Grønbæk et al. (Grønbæk et al. 2023) show that some participants felt their personal space was violated when a transition between semantic zones meant their partner's avatar phased through them. Yet, an animation that removes this feature can make it difficult for participants to manage visual attention. Does embodiment violation matter when there are other cognitive costs for the solution? Collaborative virtual environment and VR studies show people simply get used to or disregard social norm violations in immersive environments, writing these up as system glitches (Williamson et al. 2022). While other results indicate that these violations can be the source of social embarrassment or even online harassment (Poretski et al. 2018). In Chapter 5 I investigate how people navigate the trade-offs between sociocultural norms and coherent spatial embodiment, according to contexts with different social expectations.

2.3 Adapting Embodied Actions Require Ethical Considerations

Judee Burgoon's (1993) Expectancy Violations Theory (EVT) provides a useful framework for understanding how people evaluate unexpected behaviours. It shows how communicator traits, relationships, and context shape whether a violation is deemed acceptable. At the same time, the ambiguity between collaborator-driven and system-adapted actions raises ethical questions about identifying intent and the potential masking of harmful behaviours.

2.3.1 Expectancy Violations Theory: A Useful Lens for Understanding Sociocultural Norms Expectations

Expectancy Violations Theory (EVT) may offer a useful lens for understanding sociocultural norms in blended realities, where social expectations are fluid and not always semantically or spatially anchored (Dourish 2004). While prior research has primarily focused on spatial or semantic mapping techniques to support coordination in hybrid environments, sociocultural norms remain under-explored, likely due to their variability and context-dependence. EVT, developed by Judee Burgoon (1993), explains how people interpret and evaluate unexpected deviations from anticipated social behaviour. Crucially, these expectancy violations are not inherently negative; rather, their interpretation depends on both the nature of the behaviour and the perceived reward valence of the communicator (whether they are deemed friendly, familiar, competent, etc.). For instance, if a well-regarded friend stands at an intimate distance, the violation of personal space may be interpreted positively as an expression of warmth or familiarity. Conversely, if someone's boss did the same, the behaviour may be appraised as inappropriate or threatening.

EVT identifies three factors that inform these expectations: **communicator characteristics** (e.g. age, gender, personality or communication style), **relationship dynamics** (e.g. familiarity or status of the other person) and the **interaction context** (e.g. formality of the setting or nature of the task) (Burgoon 1993). When expectations are violated, EVT proposes a dual-appraisal process involving evaluation of both the behaviour and the communicator's reward valence. This mechanism helps to explain why the same action may be interpreted more favourably when performed in different sociocultural settings and begins to provide some structure to predict how people evaluate complex social interactions.

In HCI, EVT has informed how people respond to social robots and virtual agents that defy expected norms (Burgoon, Bonito, Lowry, Humpherys, Moody, Gaskin & Giboney 2016), offering insight into how people judge appropriateness based not only on actions and who performs them, but also under what conditions.

This perspective could be valuable for navigating design challenges in blended realities that require a trade-off between maintaining coherent spatial mapping and respecting sociocultural expectations.

2.3.2 Space Design has Ethical Considerations

In immersive environments, people find it difficult to discern between movements made by an avatar to communicate human motion and movements that are adapted by the system (Wong, Grønbæk & Velloso 2024). This ambiguity might arise because people instinctively apply Theory of Mind (ToM), the cognitive ability to attribute mental states to others, even when interacting with avatars (Wong, Grønbæk & Velloso 2024). ToM is closely related to the concept of mental models. These internal representations are used to predict how a system, or another person, will behave (Johnson-Laird 1983). In HCI, mental models have been used to assess usability and support group awareness in collaborative systems (Norman 2014). Wimmer and Perner's (1983) seminal study showed that by age four, children understand that others can hold beliefs different from their own. At this stage, they shift from predicting behaviour based on objective truth to predicting behaviour based on mental models. In blended realities, avatars are designed to act as proxies for remote collaborators, so it is natural that people extend ToM to them. However, as Wong, Grønbæk & Velloso (2024) show, this strategy is undermined when systems adapt avatar behaviours in human-like ways, making it difficult to discern collaborator-controlled actions from system-automated adaptations.

The ambiguity between collaborator actions and system adaptations could lead to misjudgements of intent. When people cannot see into their partner's physical environment, they must rely on the avatar as a proxy for behaviour. If the system modifies the avatar's actions, a person may infer intentions or mental states that do not correspond to their collaborator's reality. In highly heterogeneous contexts, this can create a mismatch, where people act on information that is coherent in their own environment but false in their collaborator's, leading to breakdowns in coordination (Wong, Grønbæk & Velloso 2024).

The unreliability of avatars as direct proxies for collaborators raises important ethical concerns. Adapting actions to preserve proxemic norms or avoid potential conflict may be considered a form of deception (Slater et al. 2020). Such adaptations could mask harmful behaviours, making them harder to detect or report (Blackwell et al. 2019). This demonstrates the ethical implications of adapting actions in blended realities to uphold assumed sociocultural norms.

Although people inhabiting a space will ultimately adust it to meet their own needs (Harrison & Dourish 1996), designers determine whether these adjustments are possible by making decisions about how the space is designed (Dourish 2006). From an externalist perspective, where the physical world is an integral part of cognition, responsible space design is non-trivial. By making decisions about spatial affordances and how actions are adapted by blended realities systems, designers influence cognitive processes and social interactions.

2.4 Summary

Based on an externalist perspective of cognition (Clark & Chalmers 1998) and theories of embodied interaction (Dourish 2001), space plays an essential and dynamic role in collaborative activities. MR technologies offer a compelling opportunity to reintroduce spatial affordances into remote collaboration, allowing participants to rely on embodied actions such as proxemics, f-formations, and deictic gestures (Ens et al. 2019). However, many existing MR solutions assume that distributed spaces are similar in size and orientation (Grønbæk et al. 2023). When spaces are heterogeneous, it becomes difficult to coherently transmit embodied actions across environments, disrupting cohesive communication and mutual understanding between participants (Wong, Grønbæk & Velloso 2024).

Blended realities have emerged as a promising direction to overcome this limitation by blending heterogeneous contexts so that remote collaborators' actions remain meaningful within each perceiver's environment (Grønbæk et al. 2024). Yet most work in this area has focused on spatial heterogeneity, akin to Dey's (2001) definition of context as external entities to be encoded. Drawing on Dourish's (1992) account of context as spatially and socially constructed, I suggest that blended realities need to investigate both spatial and sociocultural heterogeneity in unison, as heterogeneous contexts.

However, current systems lack a unifying language or conceptual theory to compare basic approaches to spatial heterogeneity, making it difficult to systematically analyse how different solutions address these challenges. More importantly, while spatial adaptation has received significant attention, the sociocultural aspects of collaboration, such as norms of personal space or role-based behaviours, are rarely considered. Prior work suggests that trade-offs often emerge between preserving spatial information and upholding sociocultural norms (Grønbæk et al. 2023), yet it is not known how systems should manage these conflicts or how collaborative contexts influence peoples' preferences.

Expectancy Violations Theory (EVT) may provide a valuable lens for understanding how people evaluate the conflict between upholding spatial information and sociocultural norms. Yet people struggle to apply Theory of Mind in immersive environments, making it difficult to discern whether norm violations stem from human intent or system adaptation. This raises ethical concerns about accountability, misrepresentation, and even harassment when sociocultural norms are computationally enforced or concealed. These challenges reinforce the need for a more nuanced theoretical foundation for designing blended realities.

This thesis addresses these problem spaces through two studies. The first, in Chapter 4, develops the Spatial Heterogeneity Framework, which provides a unifying language and conceptual tool for categorising and comparing existing systems that blend distributed spaces. The second, in Chapter 5, is a qualitative user study with 20 participants that investigates how people weigh the trade-offs between preserving spatial information and upholding sociocultural norms. Through semi-structured interviews and counterfactual scenarios, the study identifies contextual factors that appear to inform these trade-offs, and indicates when participants feel comfortable with system adaptations. Together, these contributions establish a foundation for designing blended realities that support collaboration across heterogeneous contexts.

Chapter 3

Transmission: Enabling Embodied Interactions Across Distributed Locations

This thesis investigates how embodied actions are transmitted across heterogeneous contexts in blended realities and the design trade-offs this involves. As demonstrated in Section 2.1, embodied actions (Dourish 2001) carried out in reference to external environments are active components in cognitive and social processes (Clark & Chalmers 1998, Suchman 1987). However, coherently transmitting these actions across distributed locations is challenging, particularly when spaces differ in size and orientation (Heath & Luff 1993, O'Hara et al. 2011, Benford et al. 1998). Blended realities offer a compelling solution by remapping spaces so that actions in the local environment make sense in the remote one. In Section 2.2, I presented an argument that current work in blended realities needs to extend this problem of spatial heterogeneity to consider heterogeneous contexts. By this I mean how both spatial *and* sociocultural actions are adapted across the distributed spaces. Unfortunately current research lacks a unifying language to talk about heterogeneous contexts from both a spatial and sociocultural perspective. In addition, the design trade-offs for adapting spatial and sociocultural actions are relatively unknown. I establish in Section 2.3 that is important to understand these trade-offs to design blended realities systems that consider important ethical implications.

The gaps identified in Chapter 2 motivate two core investigations in this thesis: the investigation in Chapter 4 defines a conceptual framework and unified language for understanding how blended realities currently transmit embodied actions across heterogeneous contexts, and in Chapter 5 I examine the trade-offs between adapting actions to preserve spatial information, or uphold sociocultural norms.

To structure these investigations, this section introduces three modes of transmitting actions in blended realities: (1) **transcribing**, which recreates an action as originally performed without adaptation; (2) **transforming**, which applies semantic mapping to align actions with the remote space while preserving spatial meaning; and (3) **translating**, which adapts actions to maintain sociocultural meaning across different contexts. I link transforming actions back to Dey's (2001) view of context as the denotation of spatial entities, while translation aligns with Dourish's (1992) notion of context as socially constructed. Finally, drawing on the Shannon–Weaver (1948) model of communication, I demonstrate how transformation and translation rules can conflict, creating a kind of "social noise". These tensions motivate the investigation in Chapter 5, which analyses how people weigh up these trade-offs in practice.

3.1 Challenges of Transmitting Embodied Actions in Blended realities

In the **transcribe** mode, the blended realities system transmits the local collaborator's action to the remote collaborator's environment without any adaptations. This mode requires the least computational interventions but relies on the different spaces being uniform in size and spatial layout. If the spaces are heterogeneous, then the original encoded message may not be interpreted correctly by the remote collaborator. For example, in **Figure 3.1** (top scenario), if the professor's actions are transcribed to the student's space, then they will not see the professor pointing at their whiteboard.

In the **transform** mode, the blended realities system maps the spaces so that features in the local space correspond with features in the remote space. Depending on the degree of heterogeneity between the spaces, different mapping strategies are required. The heterogeneity ladder introduced in Section 4.5 identifies six different levels of heterogeneity that blended realities systems will assume. At the *causal* level, every physical element in one space is perfectly mapped to its counterpart in the other space, and actions in one space have consequences in the other. At a *rigid* level of heterogeneity, only a rotational transformation is required to map the spaces coherently. At the *affine* level, the local space needs to be stretched or mirrored when mapped to the remote space. At the *non-linear continuous* level, the points in one space are warped to fit the remote space. If the spaces are *non-linear discontinuous*, then the local space is spliced into sections and these smaller homogeneous spaces are mapped to each other. Finally, at the *categorical* level, the spaces are so different that only the objects can be mapped to each other, but not the space between them.

The more heterogeneous the spaces are, the less embodied interactions are naturally afforded by the spatial mapping. For example, in systems like *Room2Room* (Pejsa et al. 2016), which assume low rigid level of heterogeneity, the spatial layout inherently supports interactions like walking to sit beside someone on their couch. However, in more heterogeneous scenarios non-linear discontinuous, systems like *Predict-and-Drive* (Wang et al. 2022) segment each space into semantic zones, enabling consistent representations within these zones but require transitions between zones to be interpolated with redirected walking. At the highest, categorical level of heterogeneity, systems like *Unmapped* (Johnson et al. 2023) can only map individual objects. Interactions like pointing or navigating the space between features must be explicitly reconstructed by the system. While such mappings make it technically possible to blend even highly dissimilar environments, they come at the cost of reduced access to interaction proxemics and deictic cues. Instead, more

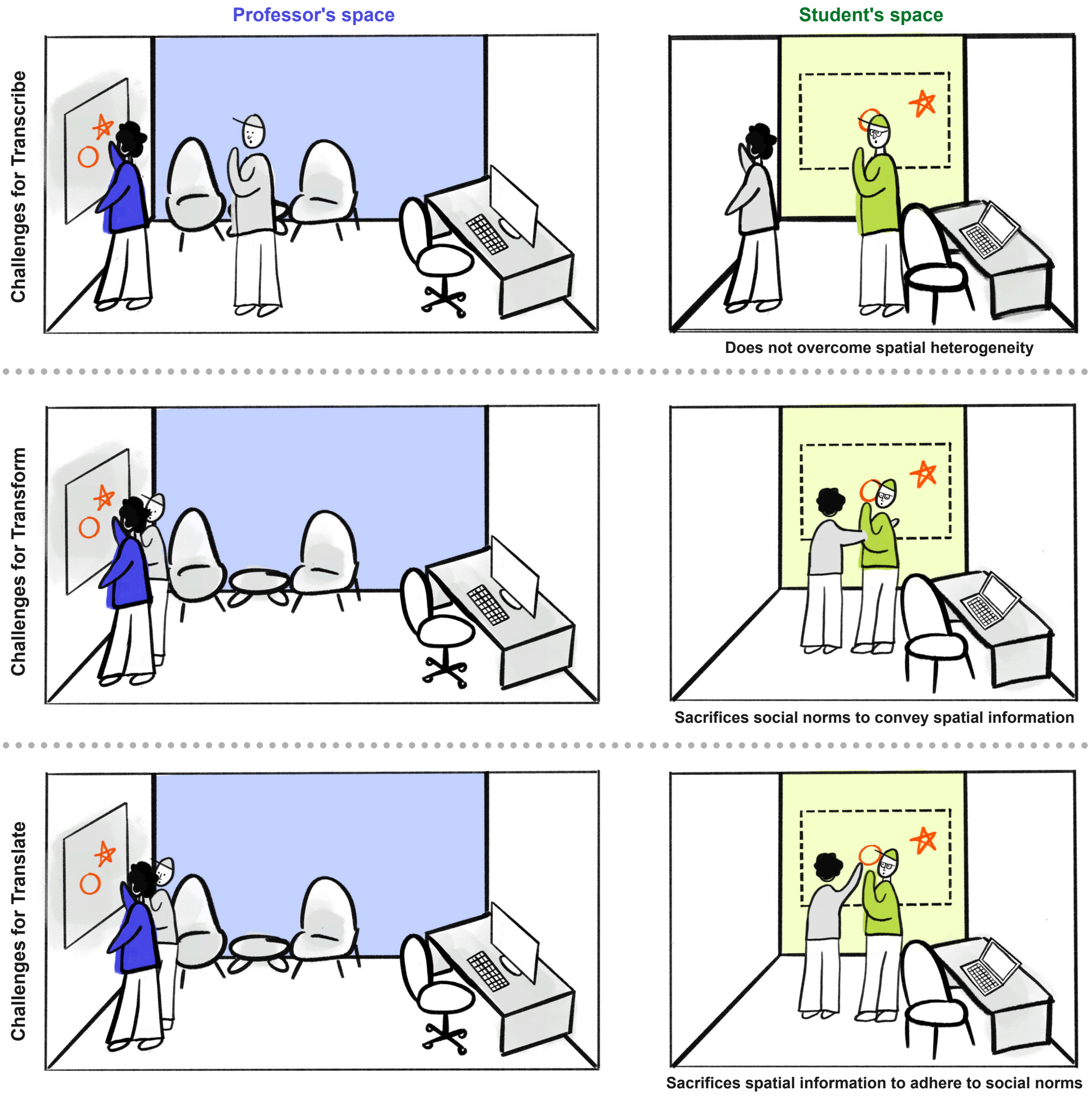


Fig. 3.1 In the **transcribe** mode, the professor's actions are transmitted directly to the student's space without adaptation. However, this creates confusion due to their different spatial layouts (top, right). In the **transform** mode, the professor's space is mapped onto the student's, spatially adapting their actions. In this case, an affine transformation is used to redirect the professor's pointing action, in coherence with the student's larger whiteboard. Yet this sacrifices social norms to convey spatial information and the professor to unwittingly points through their student (middle, right). In the **translate** mode, a collision forcefield prevents such violations and preserves sociocultural norms. However, by adhering to social norms, this rule sacrifices the spatial information conveyed by the professor's pointing trajectory. From the student's perspective, the professor gestures toward the circle instead of the star (bottom, right). This **social noise**, produced by conflicts between transformation and translation rules, highlights the design trade-offs inherent to blended realities.

fine-grained semantic mapping is required to calculate the difference between spatial features and redirect the embodied actions accordingly. For example, in **Figure 3.1** (middle scenario), the professor's pointing is redirected to match the student's larger whiteboard. However, prioritising spatial accuracy can sacrifice social coherence: the professor's avatar points correctly at the target symbol but in doing so, violates the student's personal space.

In the **translation** mode, the blended realities system considers additional requirements to support social norms across the distributed spaces. These social norms cannot be supported by spatial aspects alone and the context of the social interaction needs to be considered. A situation might call for the professor's avatar to be rendered in a different placement than what makes sense from an objective mapping point of view. For example, other rules might be introduced to avoid norm violations like pointing or teleporting through a collaborator's body (**Figure 3.1,** bottom scenario).

By analysing the sociocultural layer in blended realities through the lens of Burgoon's (1993) expectancy violations theory (EVT) (introduced in Subsection 2.3.1), this thesis connects the dynamics of social norms with the technical challenge of blending distributed spaces. Most blended realities research has focused on the **transform mode**, akin to Dey's (2001) view of context as "entities relevant to interaction", emphasising computational **denotation** of physical reality. In contrast, the **translate mode** requires grappling with the more elusive, interpretive layer of meaning-making that arises from **connotations** associated with the activity itself (Dourish 2004) (e.g. anticipating how a professor might react to someone sitting in their personal chair).

Reconciling these tensions is challenging because they rest on different epistemologies: one treats context as an objective state, the other as socially enacted and interpreted. Informed by Dourish's (2004) research on contextual awareness, I argue that social information in blended realities is created through users' situated actions and that these situated actions require bridging technical and social dimensions. This means it is important to understand both the objective affordances of the physical world and the subjective, culturally embedded interpretations of those affordances. This thesis therefore investigates what happens when people navigate transform and translate modes of interaction.

The three modes of transmission are inspired by the Shannon-Weaver model of communication (Shannon 1948). Actions are encoded by the local collaborator and transmitted by the blended realities system before being decoded by the receiver in their remote space. However, in the Shannon-Weaver model, noise, like static on a telephone line, is unintentionally introduced by the channel carrying the message. In this case, additional noise is created when the rules supporting different modes of transmission conflict with each other (**Figure 3.2**). A kind of social static occurs that can make it confusing for the remote collaborator to piece together whether the intent behind the interaction was the person or the machine. **Figure 3.1** (bottom scenario) demonstrates how design trade-offs occur in each of these modes.

Design tensions often emerge when the two modes and epistemological stances conflict, creating social noise. This sort of conflict is illustrated by **Figure 3.1** (middle and bottom scenarios),

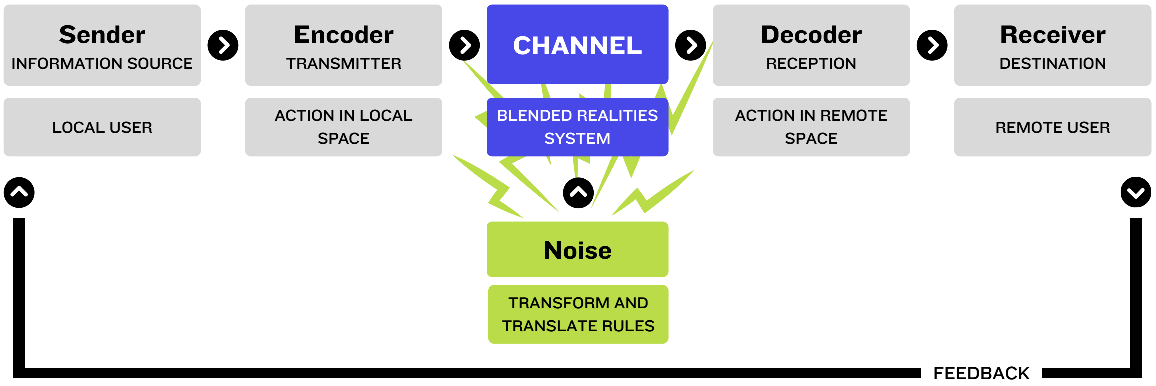


Fig. 3.2 Inspired by the Shannon-Weaver model of communication (Shannon 1948), blended realities can be understood as encoding a local collaborator's actions, carrying them through the system channel, and decoding them as actions in relation to the remote space. However, noise arises when transformation and translation rules conflict. This creates a kind of 'social static', making it difficult for the remote collaborator to interpret the intent behind an action. Since the source of this distortion is invisible, it can appear to them as if it originated from their collaborator rather than the system (Wong, Grønbæk & Velloso 2024).

where a transformation has been applied so that the professor can point directly at the intended item on the student's larger whiteboard. However, this redirected pointing means the professor unwittingly points through the student. At the translation layer, perhaps a collision forcefield is introduced, so that the professor will never point through their student. Yet this means that the professor's pointing trajectory will be slightly off. This kind of social noise is likely redundant for pointing at larger objects, but in high-stakes situations that require accuracy, such misconceptions can slow down the task and create additional cognitive demand (Johnson et al. 2023).

In this thesis, I focus specifically on conflicts between the transformation and translation modes, aiming to understand how people weigh the trade-offs involved in navigating an interface that adapts to both spatial and social contexts. In doing so, I contribute to a broader debate about the degree to which immersive technologies should control the rules of reality. I recognise the duality of space and place (Dourish 2006), offering new insight on how the design of spaces in blended realities can influence culturally imbued place-making.

Chapter 4

Transformation: Blending Heterogeneous Spaces

As established in Chapter 3, there are three modes of transmitting embodied actions across distributed spaces. The investigation in this chapter focuses on how current blended realities systems transmits embodied actions in the transform mode. This is in response to Section 2.2 of the related work, where I argue that the emerging field of blended realities needs a unified language and theoretical construct to help identify, discuss, and compare solutions. By identifying how current systems seek to overcome spatial heterogeneity (RQ1) and characterising these approaches in a theoretical framework (RQ2), I aim to help researchers examine and improve these approaches in a structured way.

In this chapter, I make the over arching argument that blended realities systems assume that distributed spaces will have a certain degree of spatial heterogeneity. This assumption dictates the embodied interactions that are afforded by the spatial layout without any adaptations, and those that need to be enabled by the system. As the degree of heterogeneity increases, the less embodied interactions are enabled by the spatial layout “for free”.

Explaining the data behind these arguments, in Section 4.2, I present the 14 canonical blended realities systems I examined in response to RQ1. I then show how I used these to derive the Spatial Heterogeneity Framework, which was validated against a further 32 blended realities systems in response to RQ2. In Section 4.3 I present the final iterated version of the Spatial Heterogeneity Framework, explaining the four components that researchers must navigate to develop blended realities systems for heterogeneous contexts. These components consist of: (1) activity zones (Section 4.4), which explains how designers of blended realities systems identify the functional areas of the spaces that need to be blended; (2) the heterogeneity ladder (Section 2.2), which identifies six different levels of spatial heterogeneity that blended realities systems generally assume; (3) blended proxemics (Section 4.6) identifies the embodied interactions that are either afforded by the spatial layout or system; and (4) the solutions matrix (Section 4.7) compares how different blended realities systems enable embodied actions across the distributed spaces by assuming a certain degree of heterogeneity. Finally, in Section 4.8 I make a case for future work to understand how blended realities might enable a large group fo collaborators to join from many different sized spaces, and how AI might augment these systems, making it easier to seamlessly transition between different levels of spatial heterogeneity.

4.1 Aim and Focus

Mixed reality (MR) technologies offer compelling solutions for distributed collaboration by reintroducing spatial communication as a shared resource. By embodying remote users as avatars, these solutions can make virtual encounters feel similar to face-to-face ones. MR also makes it possible to combine the flexibility of virtual content with the tangible benefits of physical interactions, creating an optimal environment for collaboration (Ens et al. 2019). For instance in MR, a local collaborator might approach a shared physical surface to draw a digital diagram, prompting the remote participants to also move closer and observe or annotate the content. This ability to re-anchor collaboration in the physical space accommodates embodied interactions that videoconferencing tools often omit, facilitating embodied turn-taking and contextual understanding (Kiyokawa et al. 2002, Grønbæk et al. 2020).

Blended realities is a growing subset of MR technologies that enable users to bring collaborators into their *own physical space*. This experience can be mirrored from every collaborator's perspective so that *each person perceives their team members as present in their own local environment*, advancing the integration of local physical affordances with distributed collaboration. However, this creates a fundamental challenge: if every user perceives their collaborators as avatars mapped to their unique local environment, how can researchers create a **unified experience** across **different** physical spaces? Resolving this tension is critical for blended realities systems to fully harness the affordances of the physical environment as a collaborative space.

In this chapter, I present a theoretical framework to guide practitioners on how to solve this problem, which I formalise as the *challenge of spatial heterogeneity*—**the ability to coherently blend a collaborator's actions and objects from their local space with other dissimilar remote spaces**. When I refer to 'objects', I mean the position of fixed (e.g. walls), semi-fixed (e.g. chairs), or mobile (e.g. laptops) artefacts (Marquardt et al. 2012) within a bounded area. Different spaces will have varying degrees of spatial heterogeneity, either having a very similar size and/or layout of objects (low heterogeneity) or a very different size and/or layout of objects (high heterogeneity).

For example, in **Figure 4.1**'s first scenario, Alice and Bob are two collaborators distributed across home offices that have a low degree of spatial heterogeneity; their spaces are the same size and have a similar layout of objects. Using MR, they are both able to see their remote partner as a 3D avatar naturally walking around. They can sit together at the "same" desk and draw on their respective whiteboards as if they were one. However, in the second example, Alice has replaced her whiteboard with a larger one. She has also moved her desk to face the window on the other side of the room. Each of these actions change the level of spatial heterogeneity: Alice creates

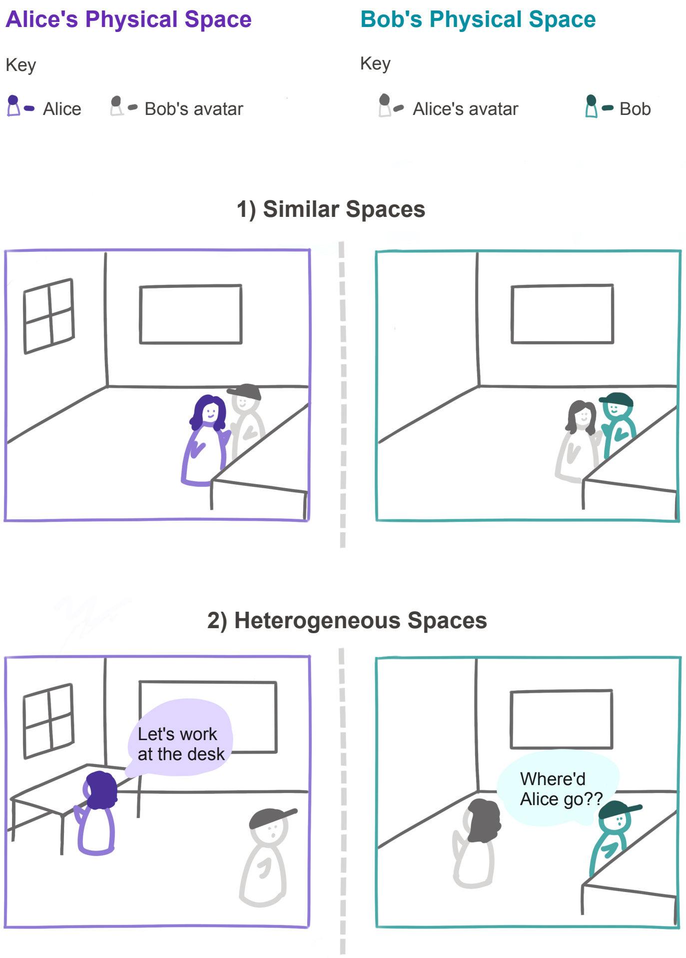


Fig. 4.1 Alice and Bob start with similar spaces. However, when Alice moves her desk and changes the size of her whiteboard, she increases the level of heterogeneity between their spaces. The increased spatial heterogeneity complicates how Alice and Bob's avatars need to be rendered in the space.

less compatibility between their layout and objects. It becomes more challenging to map Alice's movements one-for-one to Bob's local space; in this instance, she appears to point at an empty space. Consequently, if Alice and Bob go from working together on a shared whiteboard and move to their respective desks, they no longer appear to move in the same direction, leading to the situation in **Figure 4.1** (bottom left).

Recent work has started to engage with the problem of blending heterogeneous spaces for collaboration. This emerging body of research has produced a rich set of approaches to create blended physical-digital spaces, such as redirecting deictic gestures (Yoon et al. 2021, Fink et al. 2022), providing different task-based models of shared space (Herskovitz et al. 2022), mapping floor spaces Lehment et al. (2014), or mapping functional interest points across spaces in a tightly coupled (Jo et al. 2015, Congdon et al. 2018) or loosely coupled manner (Grønbæk et al. 2023, Johnson et al. 2023, Grønbæk et al. 2024). However, there is a lack of shared terminology and foundational theory to describe the different benefits of these options or when they should be implemented.

Blended realities practitioners, including researchers, designers, developers, and other application decision-makers, need a better way to navigate solution types and optimise them for the level of heterogeneity at hand. **The framework I present aims to help practitioners navigate spatial heterogeneity to enable blended realities systems.** This framework was informed by researchers' experiences working with blended realiteis systems and further validated through a literature review. Its scope focuses on MR technologies across the reality-virtuality continuum (Milgram & Kishino 1994, Speicher et al. 2019, Skarbez et al. 2021) that deliberately blend physical and digital environments to support collaborative interactions. This excludes experiences in augmented reality (AR), where the physical environment is not crucial for collaboration, but could include experiences in virtual reality (VR) that are grounded in physicality (e.g. substitutional reality (Simeone et al. 2015)).

In the remainder of this chapter, I explain how the framework can be used to describe, compare and generate new solutions for blending heterogeneous spaces in MR collaboration. Overall, this chapter makes the following theoretical claims, supported by empirical and artefact contributions:

Theoretical claims

- To maintain spatially coherent interactions across distributed locations, blended realities systems must make assumptions about the degree of spatial heterogeneity. These assumptions can range across causal, rigid, affine, non-linear continuous, non-non-linear discontinuous, or categorical mappings.
- As spatial heterogeneity increases, the potential for perceptual distortions also grows, since fewer proxemic cues are afforded "for free" by the physical layout of objects.

Empirical and artefact contributions

- The Spatial Heterogeneity Framework, a conceptual model grounded in proxemics theory, for identifying heterogeneity problems, mapping potential solutions, and analysing how current systems address them.
- The concept of blended proxemics, which can be used to categorise the different proxemic-based interactions that support non-verbal communication and are enabled by blending heterogeneous spaces.
- Validation of the Spatial Heterogeneity Framework through an analysis of 32 papers (1994–2024) describing systems for blended realities.
- Identification of four components that determine how systems overcome spatial heterogeneity: (1) activity zones that different functions of space used in the collaborative activity,

(2) assumptions made about spatial difference on the heterogeneity ladder, (3) blended proxemics either afforded by the physical space layout or enabled through blending technique, and (4) how technical solutions enable these affordances using different blending techniques or by relying on the physical space layout.

- A practical worksheet that guides researchers step by step through applying the Spatial Heterogeneity Framework.

I envision that MR technologies that blend physical spaces will permeate people's work habits, fostering richer interactions than current remote work tools. By formally presenting the challenges of spatial heterogeneity, this research provides pathways for blended realities systems to bring remote users into a shared, physical-digital, blended environment.

4.2 Method

The need for a unifying framework arose from our research lab's experience developing systems that blend heterogeneous spaces in mixed reality (MR). Using a set of 14 canonical examples (Pejsa et al. 2016, Grønbæk et al. 2017, Grønbæk et al. 2024, Johnson et al. 2023, Orts-Escolano et al. 2016, Yoon et al. 2022, 2021, Wang et al. 2022, Fink et al. 2022, Jo et al. 2015, Huang & Xiao 2024, Herskovitz et al. 2022, Sra et al. 2018, Hoppe et al. 2021), I iteratively designed an initial framework proposal through a series of workshops within the authorship team. These canonical examples were chosen based on their unique approaches to blending heterogeneous spaces, and in collaboration with the final paper authorship team (Adélaïde Genay, Jens Emil Grønbæk, and Eduardo Velloso). This was presented at a workshop with 6 additional researchers specialising in MR. The feedback from this workshop was used to iterate the framework.

To validate the framework, I performed a comprehensive literature review from 1995 to 2024. Since blended realities is an emerging field and subset of MR, I started with Ens et al.'s (2019) systematic literature review on MR collaboration as an initial dataset covering 1995 to 2019. From this dataset, I identified 80 papers that described remote collaboration systems. I then supplemented this sample with papers from the last five years, performing a keyword search *("MR" OR "mixed reality" OR "AR" OR "augmented reality" OR "telepresence" OR "VR" OR "virtual reality" OR "XR" OR "extended reality") AND ("remote" OR "distributed" OR "hybrid")* from 2019 to 2024 in the same databases as Ens et al. (2019) (CHI, CSCW, and ISMAR). I added any additional papers the authors were aware of and included a search of IEEE VR proceedings. In total, I identified 345 papers from Ens et al.'s (2019) systematic literature review (80), a search from the last five years of CHI (98), CSCW (6), ISMAR (60) and IEEE VR (84), and papers supplemented by the authors (17).

The inclusion criteria were as follows: (1) the system focused on remote collaboration and (2) it blended two or more dissimilar physical spaces. I first skimmed the abstract of each record and excluded any that clearly did not meet the above criteria (302). Then, I read the full text of the remaining articles (43) and met with members from my research lab to discuss and remove any articles that did not meet the criteria (11).

This left a total of 32 articles, of which 18 were published in the last five years (2020 to 2024), 8 in the five years before that (2015 to 2019) and only 6 before 2015. This shows how the topic of spatial heterogeneity and blended realities in MR is relatively small but has been growing extensively since the 2020 pandemic.

I analysed each system described in the 32 papers according to the Spatial Heterogeneity Framework. A database of all 32 papers and how they relate to the framework can be found in an additional Notion database[1] and a hard copy spreadsheet is also provided in the supplementary materials. Subsection 2.2.2 in the related work gives an overview of key examples used to demonstrate the framework.

I found the framework covered the majority of systems uniformly. The only minor outlier was Yang et al.'s (2024) system, which prompts the user to move into an optimal area for collaborating rather than relying on the system to enable proxemic-based interactions. However, rather than invalidating the framework, it surfaced a different approach to enabling proxemics, which I discuss further in Section 4.8. The remainder of the paper presents the final framework and discusses additional considerations presented by edge cases.

4.3 The Spatial Heterogeneity Framework

Choosing a blended realities solution for diverse collaborative tasks is challenging, as activity requirements and spatial heterogeneity often change during collaboration. Solutions assuming spatial similarity inherit affordances from the physical environment but struggle to adapt to highly heterogeneous spaces, limiting their flexibility. Users may also address spatial challenges by modifying the layout of their physical space or adapting tasks rather than changing the blended realities solution.

To address the lack of a theoretical basis for designing and comparing systems that overcome spatial heterogeneity, identified in Section 2.2, I propose the **Spatial Heterogeneity Framework** (Figure 4.2, right). This framework externalises abstract concepts, enabling researchers to better discuss, compare, and develop systems for collaboration across dissimilar spaces.

The framework comprises four components: activity zones, the heterogeneity ladder, blended proxemics, and a solutions matrix. By considering each component, practitioners can identify and debate the challenges of blending heterogeneous spaces. A practitioner can start with any of the outer components (**Figure 4.2**), depending on their constraints and priorities, and then move inward to consider the heterogeneity ladder before moving out again to consider the remaining outer rectangles. The worksheet in Appendix A provides a more detailed step-by-step process for this approach. At a high level, the components are defined as follows:

---

[1] https://spatial%2Dheterogeneity%2Dliterature.notion.site/18afa49cb74480938616d17edbf5112c?v=18afa49cb744813d9853000caed6bfbd

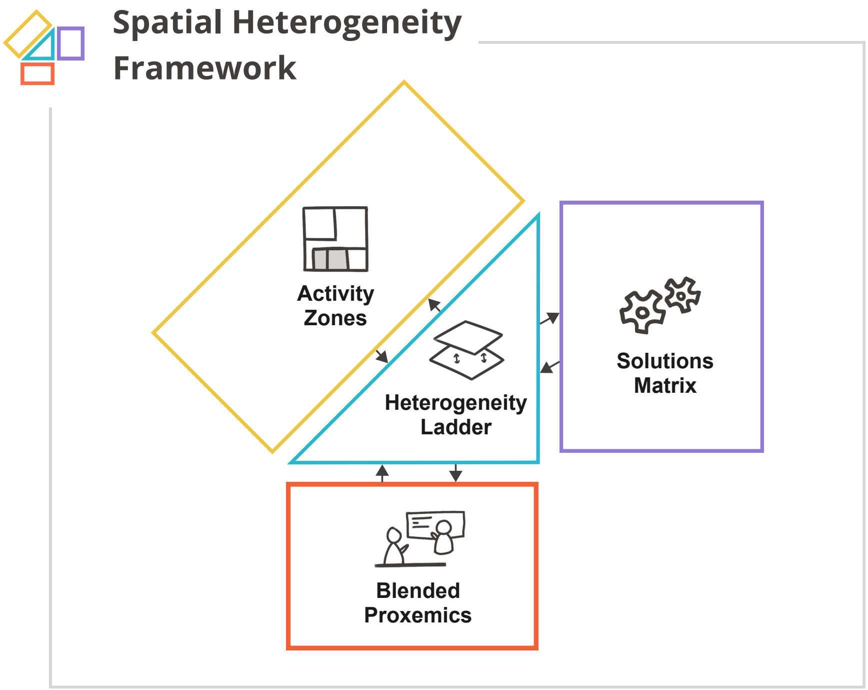


Fig. 4.2 The Spatial Heterogeneity Framework consists of four components: the activity zones, heterogeneity ladder, blended proxemics, and solutions matrix.

(1) **Activity zones** (Section 4.4), which are used to define *the roles of each area of the physical space in the collaborative activity*. This component identifies the *"blended zones"* in each space to be mapped to each other.

(2) The **heterogeneity ladder** (Section 4.5) describes the level of *physical similarity* between the remote spaces' respective blended zones. The level of similarity between blended zones of distributed spaces depends on the layout of the rooms' fixed (e.g. walls), semi-fixed (e.g. a desk) and mobile (e.g. whiteboard pens) features (Marquardt et al. 2012). The ladder serves as a hierarchical metaphor, where each step up the ladder represents an increase in spatial heterogeneity—the degree of dissimilarity between distributed spaces. As the level of heterogeneity increases, it becomes more challenging to create a sense of shared space across distributed locations.

(3) **Blended proxemics** (Section 4.6), which refers to the *social interactions between people, objects and space* supported across the distributed locations. For example, in **Figure 4.1**, a blending technique might adjust Alice's avatar in Bob's space so that her pointing gesture aligns correctly with the table in Bob's environment, ensuring the proxemic cues are meaningful and contextually appropriate.

(4) **Solutions matrix** (Section 4.7) compares solutions that *enable blended proxemics* not already afforded by the direct mapping between the distributed spaces. For example, a system

might use re-directed walking (Yoon et al. 2021, Wang et al. 2022) to make a person's avatar move correctly toward a desk in a remote location.

In subsequent sections, I detail these components, **using the following case study to illustrate the interplay of the framework's components:**

> Alice and Bob are in different locations and plan to meet using a blended realities system to workshop a paper they are writing. In their meeting, Alice and Bob sit at their desks. Ideally, they want to configure their headsets so it feels like they are sitting opposite each other without needing to move the desk away from the wall. After their initial conversation, Alice and Bob agreed to extend their meeting so they could work on some visual diagrams. However, to sketch their ideas, they now must incorporate their whiteboards while occasionally accessing the material on their desks. When they try to use their whiteboards, they notice something strange; because Bob's space has different dimensions, his avatar appears to walk beyond her office wall (**Figure 4.3**).

I use the Spatial Heterogeneity Framework to show how to help Alice and Bob negotiate the spatial heterogeneity between their locations. The aim here is to improve their spatial communication and sense of being in the same place. For the purpose of this demonstration, I start with the activity zones component. However, it is possible to start at any of the outer rectangles in the framework. I deliberately use a relatively simple scenario to describe the framework, noting that the validation also considered more complex situations.

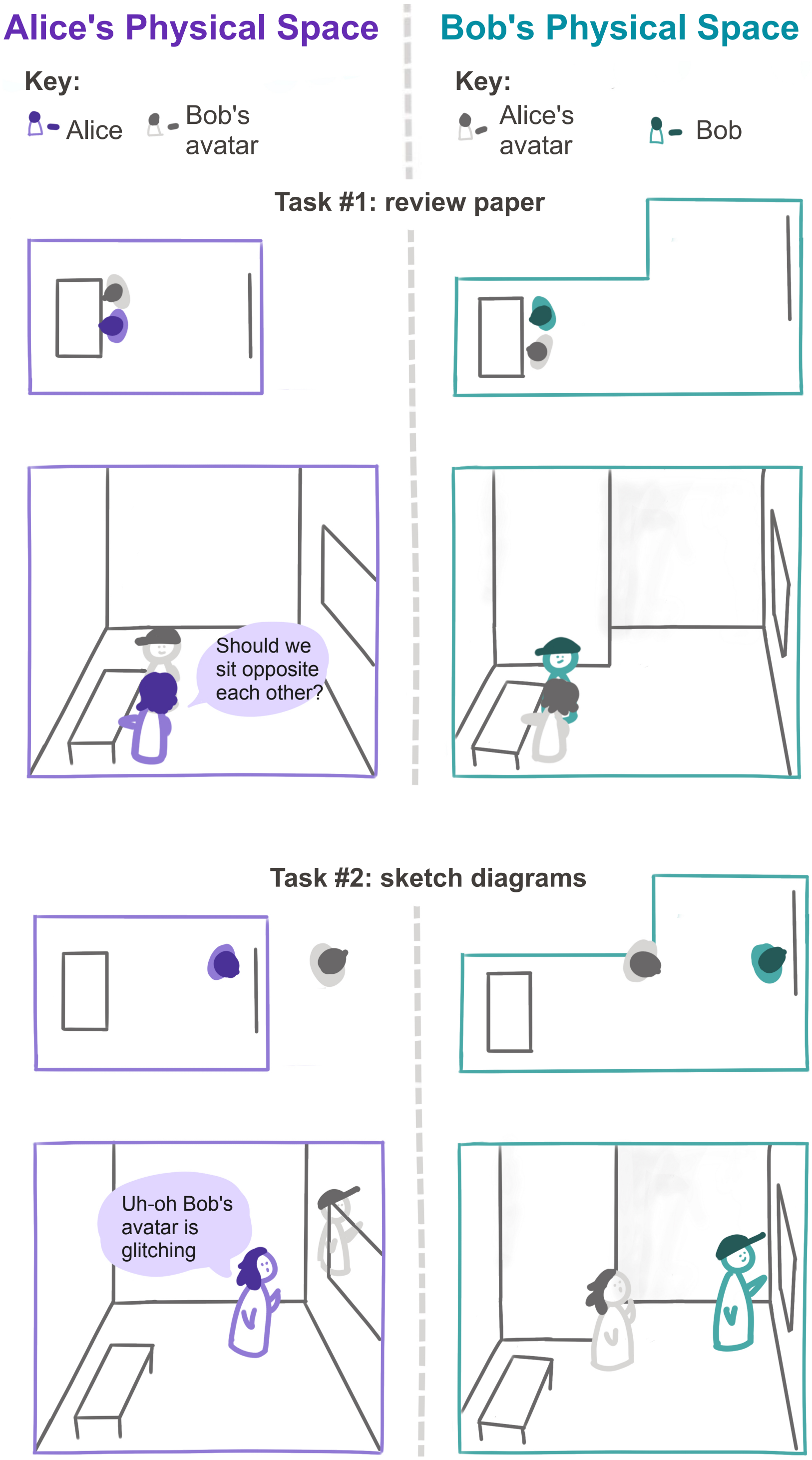


Fig. 4.3 In task one of the case study, Alice and Bob review their paper at their desks. They would like to sit opposite each other, but there is not much room behind their desks. In task two, Alice and Bob want to sketch some diagrams on their whiteboards. Yet when they walk to the whiteboards in their respective spaces, they realise their shared blended zone needs to be adjusted.

4.4 Activity Zones: Defining How the Space Will Be Used

The activity zones component helps practitioners visualise how spaces are used and which areas should be blended. **By zone, I mean the dynamic areas defined in the blended realities system that show where different types of interactions will take place across the distributed spaces**. Elements such as fixed (e.g. wall-mounted whiteboard), semi-fixed (e.g. desk and chairs) and mobile objects (e.g. laptop and pens), as well as the space between them, are all included in the zones.

For example, in their second task, if Alice and Bob were in the same space, it would be simple for Alice to indicate her intent to draw something by moving closer to the whiteboard and picking up a pen. This proxemic-based interaction is more complex in a blended realities system, since Bob's whiteboard is in a different location to Alice's. Unless the system is explicitly told otherwise, Bob will appear to walk through Alice's wall whiteboard (**Figure 4.3**)*. As such, there needs to be an explicit definition of how Alice and Bob's spaces are configured so that the system can blend these in a way that will support their second task's embodied interactions.*

I propose a taxonomy of these zones informed by prior work that shows how people divide up physical spaces into areas for a specific group or individual activities (Buxton 2009, 2006, Wong, Sánchez Esquivel, Leiva, Grønbæk & Velloso 2024). Figure 4.4 illustrates the different areas in the zoning taxonomy. This division of space shows where the distributed users need their spaces to support their embodied interactions. This gives practitioners a shared vocabulary to agree on how different areas of the space should or should not be blended by the system. The taxonomy is hierarchical and divides the space according to whether it is used for the collaborative activity (active vs. inactive), visible to another user (public vs. private), and mapped to another user's space (blended vs. independent).

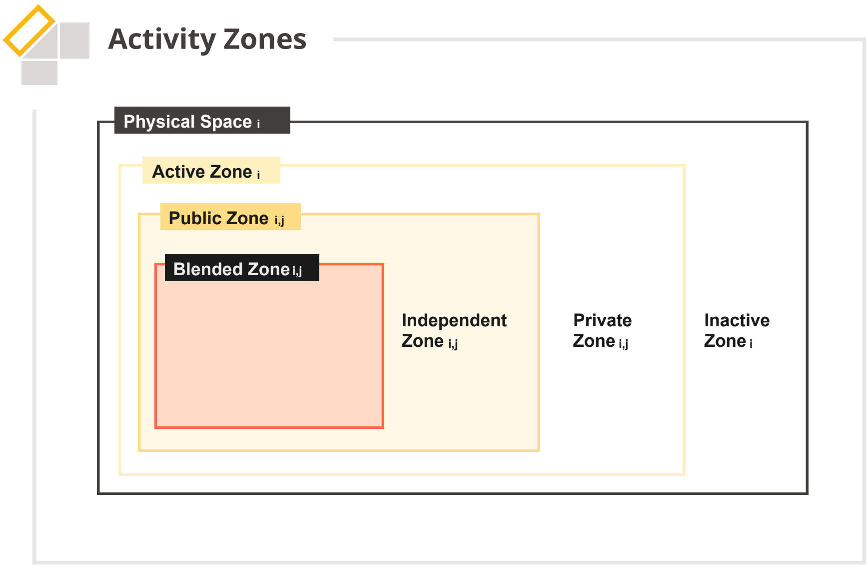


Fig. 4.4 In this figure i represents the local user and j the remote user. The activity zones component shows how physical spaces are divided into zones for different collaborative activities (Buxton 2009, Wong, Sánchez Esquivel, Leiva, Grønbæk & Velloso 2024). This zoning is hierarchical, splitting the space according to whether it is involved in the activity, shared with another user, or mapped to another user's space. Importantly, the blended zone in the centre is where the distributed spaces should be aligned so that interactions are translated across the different locations.

Physical space$_i$: fixed and semi-fixed features of the user i's environment. *Alice and Bob's homes, including their furniture, delimit their physical space. It is important to note that the entire physical space is not necessarily defined in the blended realities system but has the potential to be included for larger activities and is, therefore, included in the taxonomy.*

Active zone$_i$: the selected area inside the physical space where the entire collaborative activity happens for user i. *In the first task (Figure 4.5), Alice makes her entire office part of the active space, her physical space is quite small and she wants Bob to still see if she moves away from the desk. Bob only wants to make a small section of his larger office part of the active zone. In the second task, Bob decides to use his entire office space; the active zone is expanded to include his whiteboard on the other side of the room.*

Inactive zone$_i$: the zone inside the physical space that the user is neither using nor planning to use during the collaborative activity. It is the complement of the active zone within the physical space. *For Alice and Bob, all areas outside the active zone (e.g. their bedrooms, their kitchens, etc.) form this zone (Figure 4.5* ***and*** *4.6).*

Public zone$_{i,j}$: the selected area inside user i's active zone that is visible to remote user j. As long as user i is in the public zone$_{i,j}$, user j can see them. *In the case study, Bob initially sets his public zone smaller than his entire office space so that Alice is not distracted by the notes on his whiteboard behind him (Figure 4.5).*

Private zone$_{i,j}$: the area within user i's active zone that is *not visible* to remote user j. The local user i can access, view and interact with items in this zone, but not remote user j. They may decide to give some users access to this zone but not others. For example, in Irlitti et al. (2023), only the instructor and his workbench were streamed to remote users, who could not see the students co-located with the instructor. These students were in the private zone between the instructor and the remote students. *In the scenario (**Figure 4.5**), the private zone includes the rest of Bob's active office space, where he keeps notes on his whiteboard private until he decides to sketch. Alice does not have a private space large enough to be used since her office is quite small and she wants Bob to see all of her embodied interactions.*

Blended zone $_{i,j}$: the area of user i's public zone that should overlap with remote user j's public zone. It represents the features of the physical space with corresponding mapped features in the other user's space. *In task one (**Figure 4.5**), Alice and Bob would like to blend their desk area. In task two (**Figure 4.6**), they extend this blended zone to include their whiteboards and the space between these objects.*

Independent zone$_{i,j}$: the area inside the public zone that is visible to other users but is inaccessible or does not guarantee a mapping with other users' public zones. This is different to the private zone since remote collaborators *can still see interactions* in this area but are not able to act in this area themselves. For example, in Grønbæk et al.'s (2024) *BlendedWhiteboard*, two users can face each other through their whiteboard. In this case, the blended zone is the whiteboard wall, and the independent zone is the space on either side of the blended whiteboard that the participants occupy and can walk around. They are able to both act on the whiteboard wall (blended zone) but cannot access each other's spaces. Certain solutions overcome spatial heterogeneity by only blending a small compatible area and leaving the rest independent (Grønbæk et al. 2024, 2023, Herskovitz et al. 2022). *In task one (**Figure 4.5**), this includes Alice's whiteboard areas since it has not been included in the blended zone for collaboration. If she moves to the whiteboard, Bob will see her stand up and move away from the desk, but he will not see her move to the whiteboard in his physical space. In task two (**Figure 4.6**), they both expand their blended zone to include the whiteboard. The independent zone becomes the vacant space between the blended zone and the public zone.*

Each distributed space involved in the collaborative task can be classified according to this taxonomy. These zones are used to understand which parts of the distributed locations must be blended. If these zones are not mapped directly, they are culturally decided (Dourish 2006) and emerge from how the participants subconsciously use the space. Thankfully, it is not necessary to perfectly align each of the different zones to support the collaborative tasks (Sra et al. 2018, Grønbæk et al. 2023). Instead, the different spaces must only align attributes in the blended zone to support the activity requirements. This makes it possible to have an area of the local user's space (public and independent zone) that does not need to be blended with the remote user's space. Whereas, to enable coherent proxemic cues, it is essential for the blended zones to be mapped one-to-one, either naturally through the physical environment or with a blending technique like warping deixis (Sousa et al. 2019). These zones can be asymmetric or dynamically defined as the task changes, which I consider later in the discussion, they can also be discontiguous, which corresponds to their level of heterogeneity introduced in Section 4.5.

The following section shows how to use these zones to determine the level of spatial heterogeneity and, therefore, the scope of blended realities solutions available for translating interactions across distributed spaces. The case study below demonstrates the activity zones applied to Alice and Bob's first and second task:

> *Considering Alice and Bob's spaces in task one* ***Figure 4.5****), it can be seen that Alice and Bob primarily use their desk area, which can be explicitly defined or implicitly emerge as the blended zone. Under the zoning taxonomy, initially, features in the blended zone share a similar layout. In task two (****Figure 4.6****), however, Alice and Bob need to collaborate around their whiteboards. When they incorporate the whiteboard into their blended zones, it is clear that their spatial layout no longer matches.*

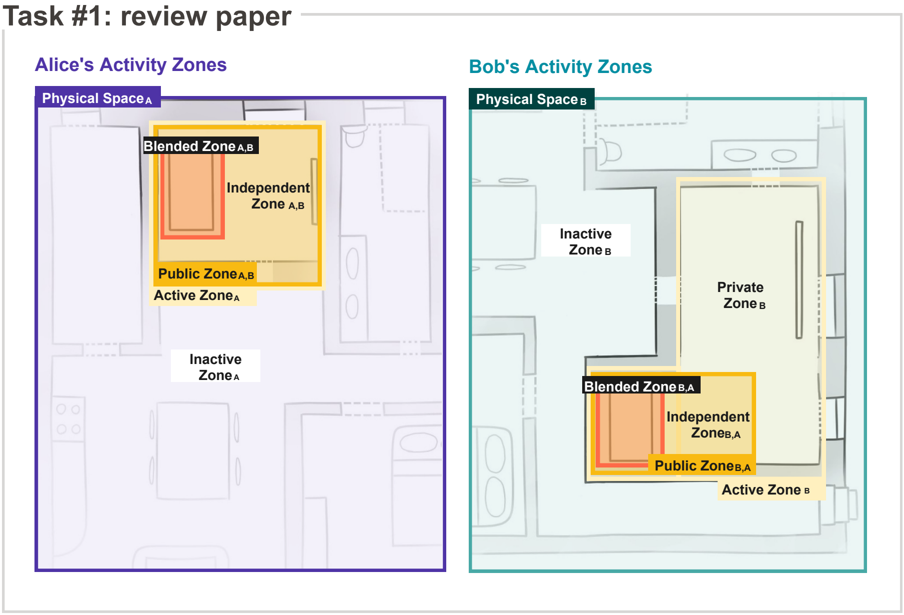


Fig. 4.5 In task 1, Alice sets her active zone as her entire office space, whereas Bob only sets his active zone as part of his office. Bob's whiteboard is currently in his private zone where he is keeping some notes hidden until they start sketching. Alice has maximised her independent zone because she wants Bob to be able to see her moving around her smaller office space. This means that the private zone is negligible and while it still exists, lies dormant and does not play an active role in the zone interactions. However, note that the whiteboard is in Alice's independent zone, if Alice moves out of the blended zone to the whiteboard, Bob will see her move away from the desk but not toward *his* whiteboard. For both, the blended space is the desk area.

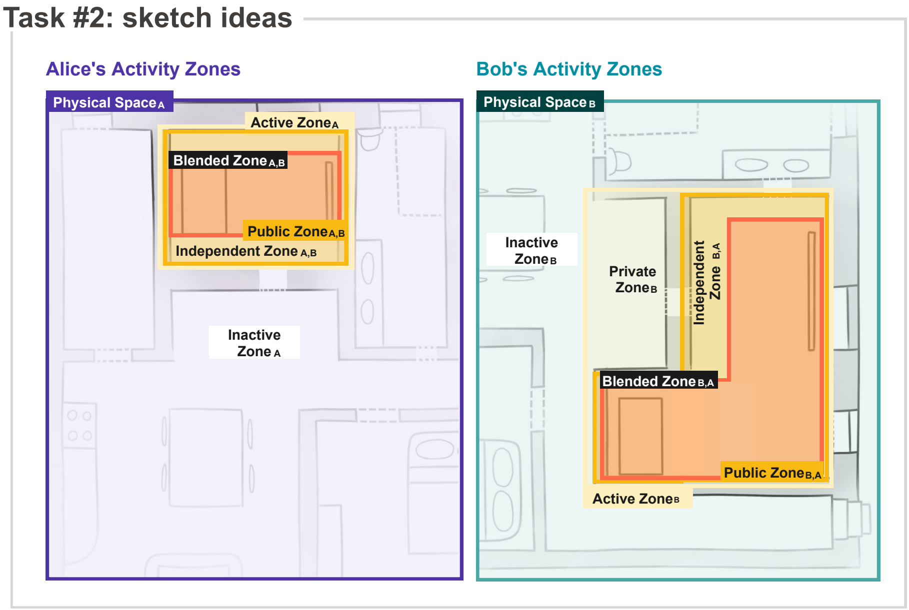


Fig. 4.6 In task 2 Alice sets her active zone as her entire office space, while Bob includes an area just outside of his office that he can use as a private zone, where he can step out if he needs to talk to his co-worker Carla. Since Alice's office space is quite small, she doesn't have a large independent zone. Similar to task 1, her private zone does not play an active role in the zone interactions. Bob has limited his blended zone to optimise it, leaving on a small area that is independent. In both spaces Alice and Bob have included their desks and the surrounding space in the blended zone. Note that the shapes here are quite different.

4.5 Heterogeneity Ladder: Defining The Spatial Difference Between Distributed Spaces

The **level of spatial heterogeneity between distributed spaces can be determined by how similar their blended zone layouts are**. The greater the difference between the arrangements of objects in the blended zones, the higher the degree of spatial heterogeneity.

> *In Alice and Bob's first task, according to their zone layout, the level of spatial heterogeneity is low, since the shape and orientation of their blended zones are very similar. However, in their second task (**Figure 4.6**), the blended zone boundaries change, increasing the degree of spatial heterogeneity for their collaborative task. In this instance, spatial heterogeneity requires further adjustments to enable collaborative activities across different spaces.*

To navigate spatial heterogeneity, I propose using a hierarchical ladder (going from low to high) and can be used as a tool for (1) characterising the level of physical disparity between the blended zones, (2) validating the range of blended proxemics supported by the physical space layout at each level, and (3) deducing the scope of blended realities solutions available. The different rungs on the ladder were informed by the canonical examples introduced in Section 4.2, and the supplementary material shows where each solution surfaced in the literature review sits on the ladder.

This heterogeneity ladder is defined in terms of the spatial transformations required to map points from Blended$_{A,B}$ onto Blended$_{B,A}$ (**Figure 4.7**). I note that the ladder is solution-agnostic—it simply describes the transformation required to create a one-to-one mapping of objects and the space around them in the blended zone. However, all technological solutions for enabling blended realities make assumptions about the level of heterogeneity they support; defining these assumptions helps to discuss and compare their characteristics (explored further in Section 4.7). In general, the more similar the blended zones, the wider the range of supported blended realities solutions and the fewer blending techniques are required to enable a congruent representation of embodied interactions across the distributed spaces. In simple terms, the rung of the ladder corresponds to how much practitioners must stretch and warp the two blended zones to create a one-to-one mapping of objects and the space around them across the blended zones. **For simplicity, these operations will be referred as "transformations".**

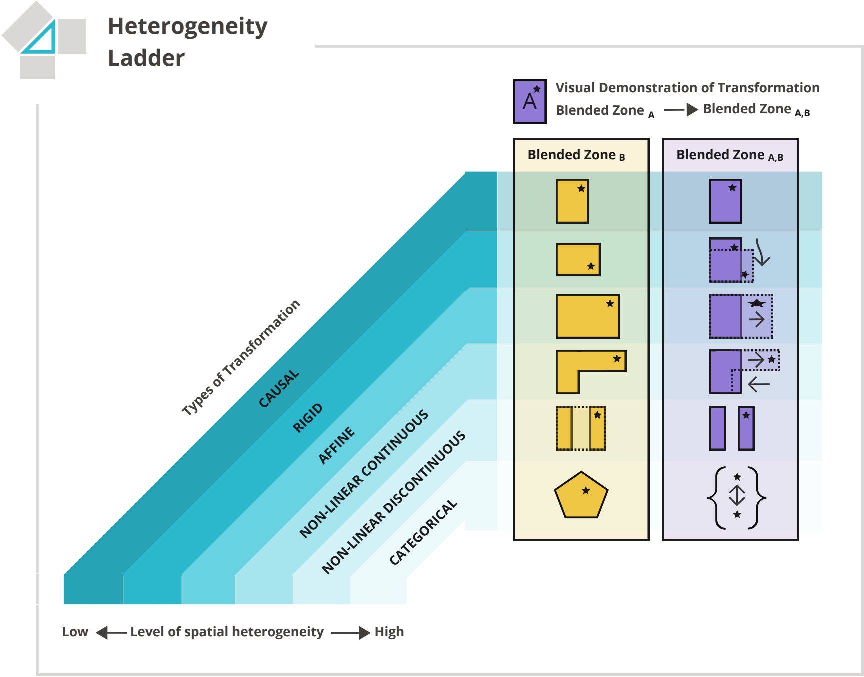


Fig. 4.7 The heterogeneity ladder represents the kind of transformation required to map the points in a blended zone (Blended$_{A,B}$) onto another (Blended$_{B,A}$), from most similar (top) to most heterogeneous (bottom).

I define space configurations on the heterogeneity ladder from the lowest to the highest level of spatial heterogeneity as follows:

Causal transformation: At the lowest rung of the ladder is the causal transformation, which is the idealised gold standard, equivalent to meeting face-to-face. Every physical element in one space is perfectly mapped to its counterpart in the other space, and actions in one space have consequences in the other. If the spaces are blended at this level, spatial heterogeneity is not a challenge. However, they require substantial infrastructure to guarantee such bidirectional causation. ***Examples of systems:*** *Ultimate Display* (Sutherland et al. 1965), *PSyBench* (Brave et al. 1998), *inFORM* (Follmer et al. 2013), *Telematic Dinner* (Barden et al. 2012), *ReMotion* (Sakashita et al. 2023).

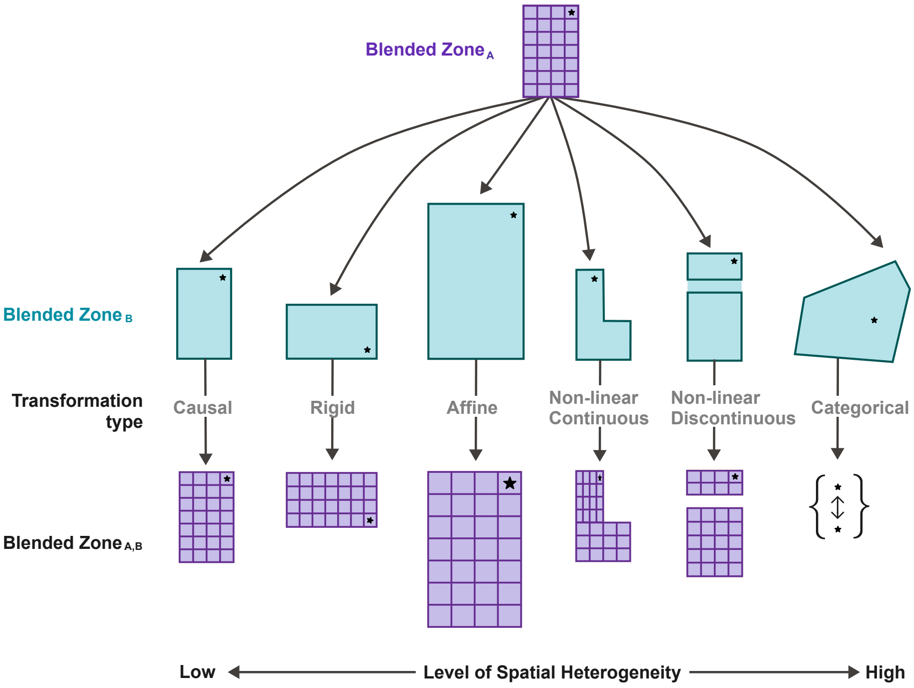


Fig. 4.8 The different transformations required to map Blended Zone A onto Blended Zone B to create Blended Zone A,B. Note that the transformations will result in different trade-offs. Causal is the only transformation where the spaces are perfectly blended at all times. In rigid transformation, some mobile objects or features in the blended zone may be rotated the wrong way, as indicated by the star. In affine transformation, warping may occur where some features are displayed as too big or too small in a remote location. In non-linear continuous transformation, some features may be warped in an attempt to maintain the space between objects. In non-linear discontinuous transformation, there is a gap in the blended space that may cause perceptual glitches in the rendered environment. In categorical transformation, the space around objects is ignored, leading to a blended space with no spatial mapping between objects.

Rigid transformation: The blended zones have identical fixed and semi-fixed structures, requiring only a suitable orientation for the mapping. Distances are preserved, and actions in either space are equivalent. However, physical alterations in one space are not reflected in the other. This level requires carefully arranging the space so that the physical objects and the space around them perfectly align. The spatial consistency across environments preserves the ability to effectively point, gesture, and use bodily orientation and mutual gaze for natural group collaboration ***Examples of systems**: Holoportation* (Orts-Escolano et al. 2016), *Room2Room* (Pejsa et al. 2016), *SurfShare* (Huang & Xiao 2024), *Subspace Allocation* (Kim, Kim, Choi & Woo 2024), *IllumiShare (Junuzovic et al. 2012).*

Affine transformation: The blended zones are similar but have different dimensions, which requires stretching or reflection of one or more dimensions. If such a transformation was used to map the two zones directly, users' actions might appear distorted (e.g., a step in one space might be equivalent to two steps in the other or writing on a surface might appear backward if not properly reflected). However, the direction of pointing is still preserved. ***Examples of systems:** Blended Whiteboard (Grønbæk et al. 2024), Loki (Thoravi Kumaravel et al. 2019), X-space (Herskovitz et al. 2022), Mutual Space Generation (Kim et al. 2022), ImmerseBoard (Higuchi et al. 2015).*

Non-linear continuous transformation: At this level, there is no affine transformation that could map objects in one blended zone to the other, requiring a non-linear transformation. Spaces share features, but they are not easily aligned. For example, if trying to map a straight corridor onto a curved one, the user's movement in one space can be plausibly rendered onto the other, but deictic gestures lose their meaning without additional inference. ***Examples of systems:** Merging Environments (Congdon et al. 2018), SpaceTime (Jo et al. 2015).*

Non-linear discontinuous transformation: At this level, the blended zones might have discontinuities, amplifying the problems of the previous rung. For example, a person walking from one blended area of their space to another separated by private space may appear as suddenly "pausing" their walk in the remote partner's space and then resuming it as they finish traversing the private space. Conversely, the remote user might appear as teleporting between the two blended zones in the local user's space. ***Examples of systems:** Re-locations (Fink et al. 2022), RealityBlender (Grønbæk et al. 2023), Predict-and-drive (Wang et al. 2022), MRTransformer (Wang, Kim, Panda, Ofek, Franco & Won 2024), Your Place and Mine (Sra et al. 2018), Placement Retargeting (Yoon et al. 2022), Avatar Telepresence (Yoon et al. 2021), Object Cluster Registration (Kim, Kim, Shin & Woo 2024).*

Categorical transformation: Only individual objects and features are mapped, but not the space between them. For example, users might have the same set of objects in each space, but there is no assumption about their relative positioning. Spatial references are meaningless, though users might still be able to make verbal references to the objects themselves. ***Examples of systems:** Real-time Retargeting (Kang et al. 2023), Unmapped (Johnson et al. 2023), Visual Guidance (Yang et al. 2024).*

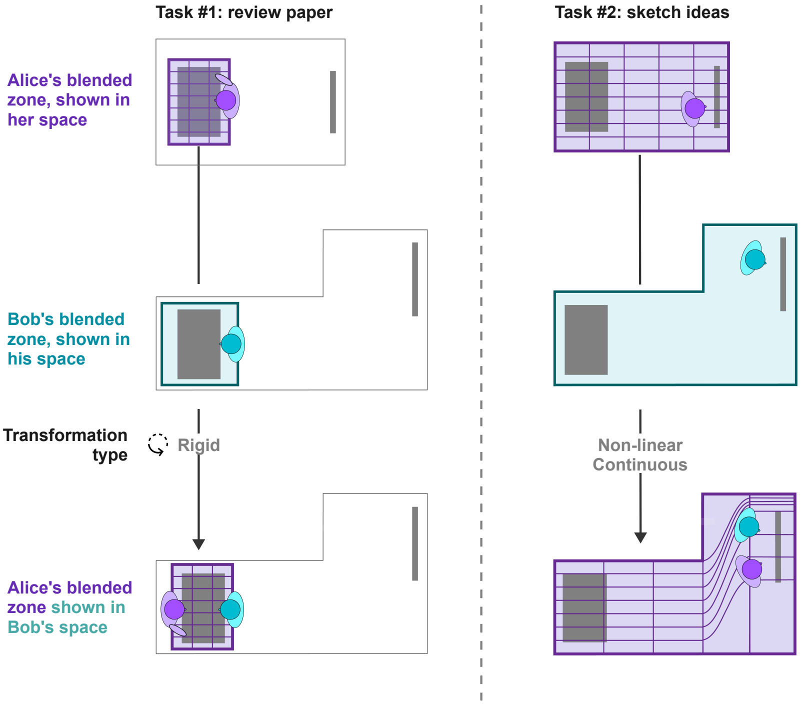
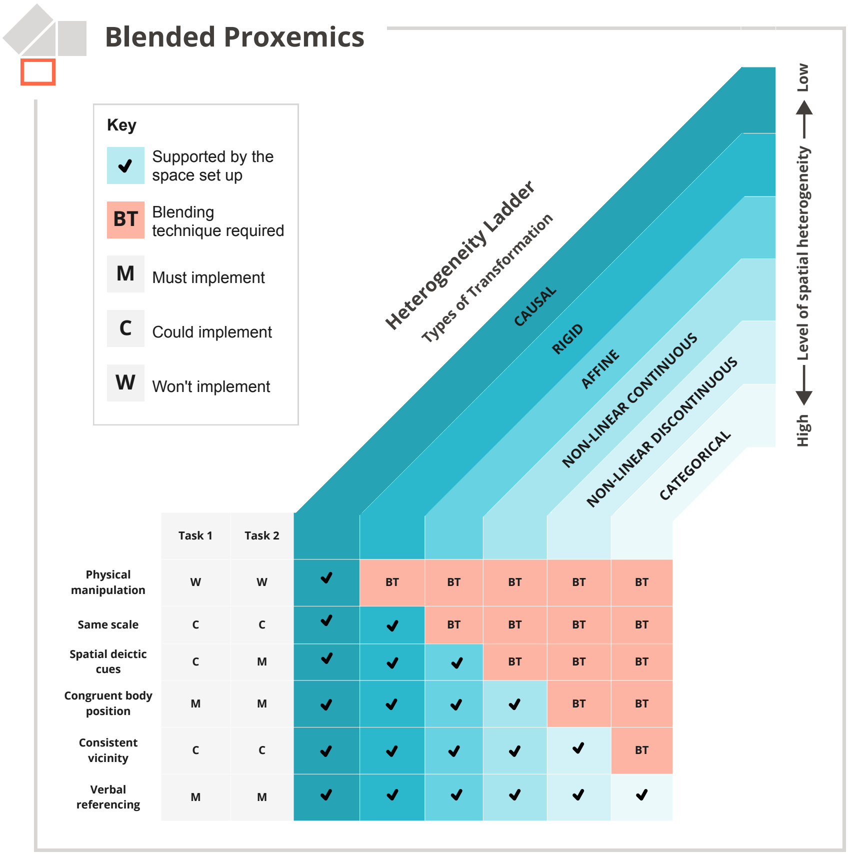


Fig. 4.9 In task 1, a rigid transformation is required to map Blended Zone Alice to Blended Zone Bob so that it appears in Blended Zone Alice, Bob that Alice's avatar is facing him. I note that with the rigid transformation, similar to facing opposite someone in the real world, if Alice points to her right, then from Bob's perspective, it would appear Alice is pointing to her left. In task 2, a non-linear continuous transformation is required to map Blended Zone Alice to Blended Zone Bob so that it appears in Blended Zone Alice, Bob that Alice's avatar is in front of Bob's whiteboard.

Returning to the example of Alice and Bob, in task one, their spaces are zoned such that a rigid transformation is possible. When they configure their headsets to sit opposite each other, this can be achieved by rotating the desk area (Figure 4.9). This transformation type is relatively high on the ladder, and their blended zones are already physically aligned, so all of the technology examples below could be readily applied to support this task. That being said, the extra computing power used to support the solutions below this rung is not necessary to adequately support the collaborative interactions. In task two, the blended zones expand. The level of spatial heterogeneity increases, moving the viable scope of MR solutions for blending the distributed locations down three rungs. In this case, Alice's avatar must be translated so that the interactions between the desk and the whiteboard make sense in Bob's space. With this configuration, it is no longer possible to apply a rigid transformation approach to the spaces or an affine transformation; Alice and Bob limit their scope of MR solutions to the ones that sit in the non-linear continuous transformation category and below. This is because of the extra transformations required to warp the space so that their avatars appear to move correctly through their partner's different physical layout (Figure 4.9).

The next section demonstrates how the ladder can be used to understand **when proxemics are naturally supported by the spatial layout and when a technical solution needs to be applied.**

4.6 Blended Proxemics: Identifying the Embodied and Proxemic Requirements for Collaborative Tasks



	Task 1	Task 2						
Physical manipulation	W	W	✓	BT	BT	BT	BT	BT
Same scale	C	C	✓	✓	BT	BT	BT	BT
Spatial deictic cues	C	M	✓	✓	✓	BT	BT	BT
Congruent body position	M	M	✓	✓	✓	✓	BT	BT
Consistent vicinity	C	C	✓	✓	✓	✓	✓	BT
Verbal referencing	M	M	✓	✓	✓	✓	✓	✓

Fig. 4.10 The blended proxemics component can be used to identify the type of interactions that should be afforded by the shared collaborative environment. Notably, as the level of spatial heterogeneity increases, so too does the number of technology fixes required to enable proxemic interactions between the different spaces. The checklist in the second column shows the interactions Alice and Bob must, could and won't have to support their collaborative activities.

Collaboration is most effective when participants can keep track of their individual and group interactions Gutwin & Greenberg (2002). During face-to-face collaboration, much of this tracking is done implicitly, through the unspoken rules of proxemics Hall (1966), Lawson (2007) and the interplay of people, objects and space Mentis et al. (2012). For example, a group member might move closer to a whiteboard when they want to draw a visual diagram. This non-verbal proxemic cue indicates their intent to draw, maintaining group awareness. **In distributed collaboration,**

these proxemic cues must be translated across different locations; I call these interactions blended proxemics. Defining a task's blended proxemics helps to determine how the distributed spaces will be used to support group awareness.

> *As demonstrated in the case study, different collaborative tasks have different proxemic requirements. In task one, Alice and Bob want to have a* ***congruent body position*** *so that they can easily see when the other person is reading or editing the paper and when they are engaged in conversation. However, in task two, Alice and Bob need to monitor each others'* ***consistent vicinity*** *with the whiteboard and desk so that they know when their partner wants to sit and edit the paper or sketch a diagram on the whiteboard. In each task, the participants use different proxemic cues to stay aware of the group dynamics.*

The spatial heterogeneity of distributed spaces and their zones determines which blended proxemics can be already supported by the blended zones and which may need to be enabled by some sort of technical solution. The blended proxemics component of the Spatial Heterogeneity Framework provides a starting point for discussing interactions that will or will not be supported across the distributed spaces and the strategies that can be used to make them available.

When exploring MR solutions, I suggest using a checklist to identify the blended proxemics that **must (M)**, **could (C)** or **won't (W)** be supported by the activity. The current checklist specifies six types of blended proxemics that can be supported to varying degrees across the different spatial heterogeneity levels. These were derived from current literature cited in the related work, yet the list is not exhaustive and can grow to include different kinds of interactions. The list currently includes:

1. Physical manipulation—moving physical objects: when moving an object in the local space, it also moves the equivalent object in the remote user's space. For example, if Alice moves her desk, then Bob's physical desk moves in his space. This blended proxemic interaction is the most difficult to support and requires substantial technical infrastructure. ***Examples of blended proxemics enabled by MR systems****: blended props or robotics (Auda et al. 2021, Onishi et al. 2022, Brave et al. 1998, Barden et al. 2012, Sakashita et al. 2023).*

2. One-to-one scale—interactions look like they happen at the same scale relative to the physical space: The avatar's movements and interactions in their local space are at the same scale as their actions in the remote space. For example, since Alice's desk is the same size as Bob's, her avatar translation will make sense in Bob's remote space. However, if Alice gets a bigger desk, then her movements will no longer be at the same scale as Bob's smaller desk. ***Examples of blended proxemics enabled by MR systems****: surface re-scaling (Grønbæk et al. 2024, Huang & Xiao 2024).*

3. Spatial deictic cues—pointing at objects in the environment or using directional verbal references: If the user points or verbally references the position of an object in their local space, their communication method points to the corresponding object in the remote space. For example, if Alice and Bob are sitting side-by-side at their respective desks, when Alice points at the right corner of her desk, then Bob will see her avatar point to the right corner of his physical desk. ***Examples of blended proxemics enabled by MR systems***: *deixis warping (Sousa et al. 2019, Yoon et al. 2021, Hoppe et al. 2021, Higuchi et al. 2015).*

4. Congruent body position and trajectory—there is a shared understanding of trajectory between landmarks: if the local user stands, sits, walks or turns to face a particular object, every step of this trajectory has a corresponding point in the remote user's space. The shapes of the trajectories might be different, but every point has a corresponding mapping. For example, as Alice walks to the right side of her space, Bob will see her avatar walking on a continuous trajectory in his space. ***Examples of blended proxemics enabled by MR systems***: *spatial anchors (Yoon et al. 2022, Wang et al. 2022).*

5. Consistent vicinity—meaningful interactions around select landmarks (fixed and semi-fixed objects) in the space: proxemic interactions typically only occur within a smaller subset of the active collaboration area to save blending the entire space. For example, in the first task, Alice and Bob's interactions make sense around their desks. However, this is only because they have used a smaller subset of the active collaboration area. When they include their whiteboards, this limited consistency no longer works. At this level, the space between landmarks might not be elegantly mapped, but at least the interactions on and around the landmarks are consistent. ***Examples of blended proxemics enabled by MR systems***: *partitioned areas for interaction (Grønbæk et al. 2023, Sra et al. 2018, Fink et al. 2022).*

6. Verbal referencing—knowing about the presence of objects common to both spaces: the system handles verbal communication so that the local user can verbally reference objects in the space. For example, if Alice says let's sit at the desk, then Bob will hear her avatar say this in his remote space. This interaction is the easiest to support using microphones and speakers. One challenge, however, is if Bob in fact does not have a desk in his space. It should be noted that this blended proxemic does not include deictic cues, such as "go right" or "put that there", these sit in spatial deictics since they have a directionality associated with them. ***Examples of blended proxemics enabled by MR systems***: *physical objects are matched across spaces (Johnson et al. 2023).*

An important relationship to note is that as the level of spatial heterogeneity increases, so does the number of technical fixes required to support interactions in the physical space. The checklist in Figure 4.10 shows the different blended proxemics ordered from most to least difficult to implement across heterogeneous spaces.

To use the checklist, the practitioner should first specify the task/s that need to be completed. Once the task is specified, each blended proxemic on the checklist is assessed to determine whether these interactions must, could or won't need to be supported to complete the task.

In task one of the case study, Alice wants to review her paper with Bob. In this task, ***congruent body position must be supported*** *to show the user's attention. Spatial deictics and consistent vicinity could be supported, but it is not crucial to their collaborative task. Physical manipulation won't be supported since the information object is digital, and no physical objects must be moved to satisfactorily complete the task (**Figure 4.10**).*

In the second task, Alice and Bob's blended proxemics requirements change. In the new checklist, Alice and Bob still want to have the body position at both their desk and whiteboard translated congruently. However, they also want to keep track of their partner's ***consistent vicinity*** *to their whiteboard or desk so they can monitor when their partner wants to sit and edit the paper or sketch an idea on the whiteboard. Additionally, spatial deictics must be supported to be able to point to particular parts of their diagrams (**Figure 4.10**).*

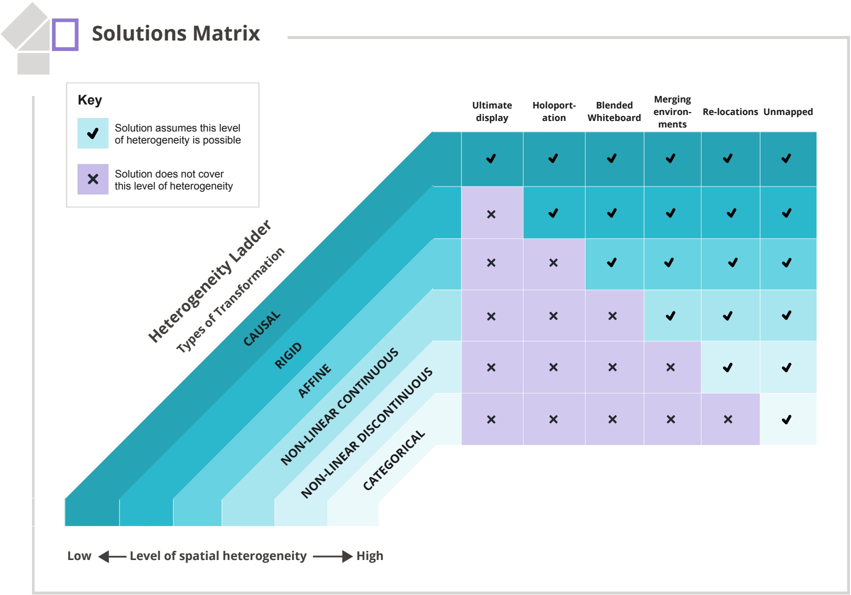
Completing this checklist creates a prioritised set of proxemic cues to be blended across the distributed spaces. This prioritisation will inform the trade-offs made when a strategy for blending the heterogeneous spaces is chosen.

4.7 Solutions Matrix: Assumptions Made About the Level of Spatial Heterogeneity and Blending Techniques

The framework is designed to be technology-agnostic, and it should be used it to compare and assess solutions. To explain how this might be done, I use six of the canonical examples shared in Section 4.2. I note that there is an array of other solutions beyond these six, which are referenced in the related work (Subsection 2.2.2), however I look to these to exemplify the framework.

Blended realities for distributed collaboration rely on assumptions about spatial heterogeneity to support blended proxemics, either through blending techniques or by leveraging the similarity between physical layouts. For instance, *Holoportation* (Orts-Escolano et al. 2016) assumes identical local and remote spaces (aside from orientation), preserving embodied interactions seamlessly but limiting versatility in mismatched spaces. In contrast, *Blended Whiteboard* (Grønbæk et al. 2024) accounts for differences like size and orientation, scaling surfaces to maintain meaningful proxemic cues.

Understanding how different blended realities systems enable blended proxemics makes it easier to weigh the different trade-offs. If a solution only supports blended proxemic workarounds for spaces with low heterogeneity levels, then it is likely this solution won't adequately support a collaborative task across spaces with highly heterogeneous blended zones. As such, the solution chosen mediates the number of blended interactions enabled across different levels of heterogeneity, and each solution will have its own set of trade-offs.



	Ultimate display	Holoport-ation	Blended Whiteboard	Merging environ-ments	Re-locations	Unmapped
CAUSAL	✓	✓	✓	✓	✓	✓
RIGID	×	✓	✓	✓	✓	✓
AFFINE	×	×	✓	✓	✓	✓
NON-LINEAR CONTINUOUS	×	×	×	✓	✓	✓
NON-LINEAR DISCONTINUOUS	×	×	×	×	✓	✓
CATEGORICAL	×	×	×	×	×	✓

Fig. 4.11 The solutions matrix shows the assumptions different systems make about the distributed spaces' level of spatial heterogeneity. Systems that support the highest level of spatial heterogeneity will support all the levels below. However, it is important to weigh up the blended proxemics already afforded by the space when selecting a solution, to use compute power wisely and avoid interface distortions.

(1) Ultimate Display: Ivan Sutherland's (1965) *Ultimate Display* describes a fictional interface where a computer can control matter itself. In this case, the assumption made about the blended zone is that all blended proxemics will be afforded by the physical space using a **causal transformation**. If you change something in one physical space, it will translate into another. The downside of Ultimate Display is, of course, that such a system is yet to exist.

(2) Holoportation: Orts-Escolano et al.'s (2016) *Holoportation* assumes that the blended zone (the table, chairs and blocks on the table) involves a **rigid** level of spatial heterogeneity. An inability to directly manipulate the remote user's set of blocks on their table rules out causal. However, the semi-fixed objects are the same in size and orientation across the distributed locations; therefore, the blended proxemics supported 'for free' by the space include the same scale, spatial deictics, congruent body position, consistent vicinity and verbal referencing. However, if the local user increases the scale of their table, they change their level of spatial heterogeneity to affine. In this case, a technical solution would be needed to translate their interactions relative to the size of the table. For example, placing the virtual representation of their blocks so they land on the smaller table, not in mid-air.

(3) Blended Whiteboard: In Grønbæk et al.'s (2024) *Blended Whiteboard*, the system assumes that the blended zones (the whiteboard surface) differ in size and orientation and, therefore, supports an **affine** level of heterogeneity. To enable same-scale blended proxemics, the system uses a blending technique to re-scale interactions from one blended zone to the other. In their examples, if the local user has a drastically larger whiteboard, then their avatar's position is translated relative to the remote user's smaller whiteboard. However, if these users increase their blended zone to include their desks then, unless their desks are in the same position, spatial deictic cues will not be available. In this case, they have increased their level of heterogeneity to non-linear continuous, which is a level of heterogeneity for which Blended Whiteboard does not provide spatial deictic technical solutions.

(4) Merging Environments: Congdon et al.'s (2018) system for merging environments assumes that the blended zones (the entire room, its objects and the space between them) must be dynamically remapped and warped, supporting a **non-linear continuous** level of heterogeneity. At this level, congruent body position, consistent vicinity and verbal referencing blended proxemics are given for free by the mapping between the distributed spaces. However, by warping the spaces to fit each other, the authors found that [p.4] "...mapping objects individually did not produce good results as the mapping interfered with physical movements (e.g., gestures and locomotion)" (2018, p.4). To avoid over warping the avatar movements the system enables same scale and deixis by "preserv[ing] the relative position and orientation of associated objects after mapping" (2018, p.4).

(5) Re-locations: In Fink et al.'s (2022) *Re-locations*, it is assumed that the desired blended zone (the whiteboard, desk and space between) will need to be spliced apart in order to preserve proxemic cues around the desk and whiteboard, supporting a **non-linear discontinuous** level of heterogeneity. In this case, by setting specific "re-location" areas they overcome immovable objects that may block these areas. Alternately, by splicing the space up into smaller more manageable areas, they effectively reduce the number of blending techniques required to enable blended proxemics. In each "re-location" area, the level of heterogeneity is effectively rigid, which gives them same scale, deixis, congruent body position, consistent vicinity and verbal referencing for free. However, the downside is that any interactions the local user has between the different "re-location" areas will not be transmitted to the remote user's space. Alternately, some studies have found that transition animations as the user moves between these areas can be disconcerting for the user (Grønbæk et al. 2023, Wong, Grønbæk & Velloso 2024).

(6) Unmapped: In Johnson et al.'s (2023) system *Unmapped*, mobile objects in the blended zone (cooking utensils required to bake a cake) can be mapped one-for-one at a **categorical** level of transformation. In this case, the same scale, spatial deictics cues, congruent body position, and limited consistency are enabled only with reference to the mapped mobile objects. However, this approach to mapping mobile objects one-for-one in the blended zone has its flaws. If an object is not mapped, then blended proxemics are lost. This is particularly challenging for environments that may have many objects. Alternately, if an object is incorrectly mapped, verbal referencing, along with all other blended proxemics, will be lost.

The solution matrix uses the literary examples above to visualise the assumptions made about the degree of spatial heterogeneity and, therefore, the blended proxemics they enable. **To map the assumptions about the spatial heterogeneity, practitioners would ideally ask themselves (1)**

which blended proxemics the system assumes will be supported by the physical space set-up, and (2) which blended proxemics are being supported with a technical workaround?

As mentioned before, Alice and Bob start out with a ***rigid*** *level of spatial heterogeneity. Since the blended proxemics they need for task one are already supported by their physical layouts, they opt to select a Holoportation-type templated solution. However, in task two, their activity changes, which prompts them to* ***rezone*** *their spaces so that their blended zones include their whiteboards. This rezoning increases their degree of spatial heterogeneity from rigid to* ***non-linear continuous****. Unfortunately, at this level of spatial heterogeneity,* ***congruent body position*** *proxemics are not supported by their physical space layout. Therefore, they need to make sure their technical solution has a blending technique that supports it.*

The case study shows how navigating the solution matrix component can help find a solution that will support the end user's desired blended proxemics that are not already afforded by the space. It also shows how a change in activity alters the zoning set-up, prompting us to walk through the framework again to assess the blended realities solutions that will support their new needs. In this way, the framework is a dynamic tool, and practitioners can cycle through the different stages depending on changes to their spatial situation or collaborative goals.

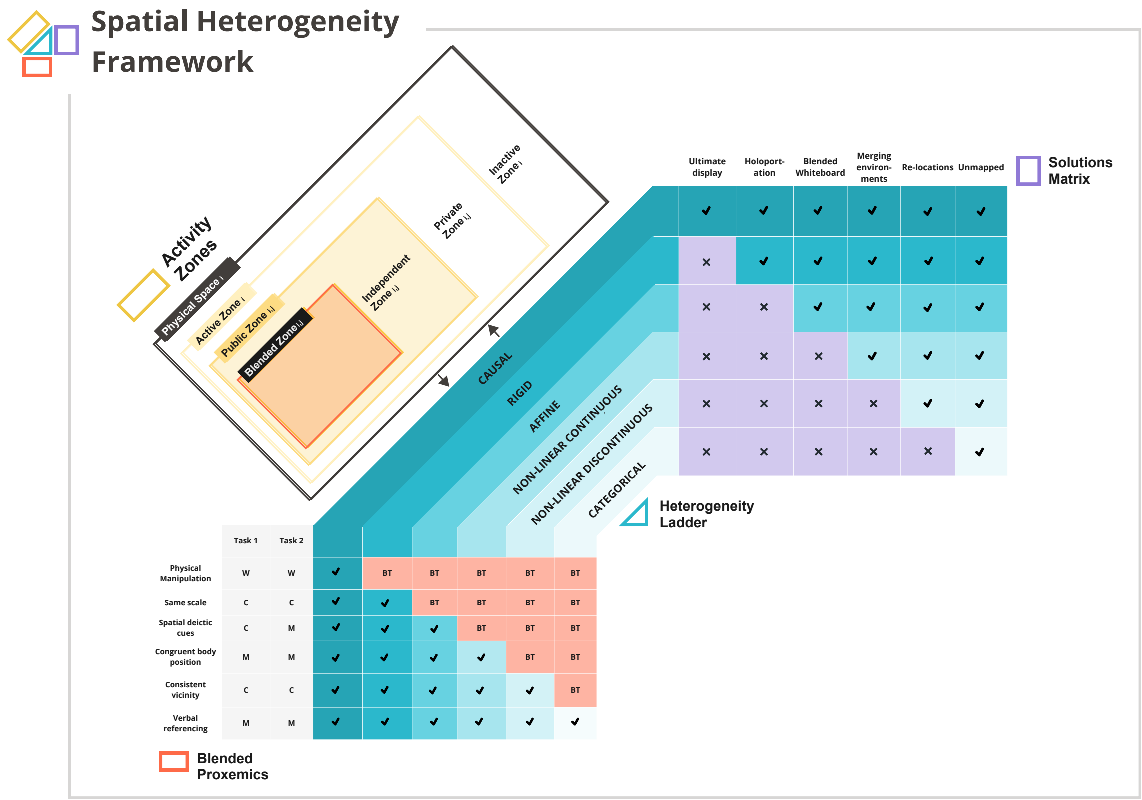
4.8 Future Challenges for Overcoming Spatial Heterogeneity

While the Spatial Heterogeneity Framework has a practical use for discussing and comparing solutions, it also reveals opportunities for future work. In this section, I outline key assumptions and challenges identified during the framework's validation process, offering potential directions for further exploration.

4.8.1 Operating at Different Levels of Heterogeneity

The demonstration of the framework in this thesis assumes that each pair of collaborators will participate at the same level of spatial heterogeneity. However, some situational and privacy constraints could make it impossible for all pairs to share the same position on the heterogeneity ladder. For example, this thesis's demonstration did not account for situations where someone is forced to be stationary (e.g. joining from an aeroplane seat) while others move around an office space. Nor did it cover how the different levels of privacy might introduce varying degrees of spatial heterogeneity.

In some circumstances, it may be impossible to find a blended state that all users find acceptable. For example, there may be situations where two people want to blend their larger, more heterogeneous spaces, while five people only want to blend a small section of their location (maybe they are on public transport or in a public space). In this situation, does it matter that two par-



Ultimate display	Holoport-ation	Blended Whiteboard	Merging environ-ments	Re-locations	Unmapped
✓	✓	✓	✓	✓	✓
×	✓	✓	✓	✓	✓
×	×	✓	✓	✓	✓
×	×	×	✓	✓	✓
×	×	×	×	✓	✓
×	×	×	×	×	✓

	Task 1	Task 2						
Physical Manipulation	W	W	✓	BT	BT	BT	BT	BT
Same scale	C	C	✓	✓	BT	BT	BT	BT
Spatial deictic cues	C	M	✓	✓	✓	BT	BT	BT
Congruent body position	M	M	✓	✓	✓	✓	BT	BT
Consistent vicinity	C	C	✓	✓	✓	✓	✓	BT
Verbal referencing	M	M	✓	✓	✓	✓	✓	✓

Fig. 4.12 The components in the Spatial Heterogeneity Framework influence each other. The way one component is navigated will depend on the outputs of another. Practitioners should start with one of the outer rectangles first before moving into the heterogeneity ladder and then using the outputs to inform the other outer components.

ticipants experience different embodied interactions and proxemics to the other five (Johnson et al. 2021)? Does place-making remain similar to videoconferencing, where users are unable to reach into their collaborators' spaces to create a shared place for collaboration? Using the Spatial Heterogeneity Framework as a theoretical lens, researchers might explore whether there can be multiple instances of a blended zone that sit at different levels of the heterogeneity ladder.

Similarly, privacy settings introduce new challenges that might be investigated by future work. There may be situations where a workshop facilitator and assistant give each other access to their private zone, making it part of the blended zone between these two participants. Consequently, their level of spatial heterogeneity might change in comparison to the whole-of-group blended zone. This presents challenges for how spaces transition between these different zoning taxonomies and levels of spatial heterogeneity. Should the workshop facilitator and assistant manually trigger their blended zone configuration to change when they want to talk privately, or should this happen automatically? What will the other participants see when this happens? The current framework provides the scaffolding to discuss these questions. I suggest that activity zones, outlined by the framework, can be used to discuss and interrogate how people blend and navigate heterogeneous spaces to accommodate different roles and privacy settings.

4.8.2 Who or What is Doing the Blending?

In theory, the blended zones could be adjusted indefinitely to find the perfect level of heterogeneity and blended proxemics for the task at hand. However, more realistically, the user will stop their modifications when they reach their optimal, desired environment or settle for an option that is 'just good enough' or less costly to put in place. Future work could explore why users choose certain modification strategies and patterns in adjusting blended zones. Systems could reduce the burden on users to change their set-up by using these patterns to auto-select templates or suggest changes to physical space and activities. Ideally, systems should consider the user's context, integrating meeting agendas or using AI to adapt to work styles and collaborative practices.

During my validation of the framework, Yang et al.'s (Yang et al. 2024) recent paper on visual guidance for user placement in dissimilar spaces stood out as choosing a different modification strategy from other papers. Instead of using a blending technique to change the way the avatars were rendered, they showed users where the optimal places were to stand, relying on them to make decisions about how they move about the space in a way that will make sense for their remote partner's space. This highlights the need to balance system automation with user agency in modifying blended zones. While automated systems can simplify adjustments by providing templates or suggestions, approaches like Yang et al.'s (2024) demonstrate the value of empowering users to actively shape their environment. By offering visual guidance rather than automating spatial adjustments, their approach enables users to make decisions that align with their understanding of the collaborative context and their partner's spatial needs.

Although blended zone modification can and should eventually be automated, allowing users to personally adapt the space remains paramount for effective collaboration. As Harrison and Dourish (1996) suggest, "space is the opportunity, place is the understood reality" (1996, p. 69). While blended realities produce the spatial opportunity for collaboration, allowing people to modify its features creates a shared place that is optimised for collaborative tasks. A well-designed blended zone should not only support intuitive spatial interactions but also let users customise and "re-place" (Harrison & Dourish 1996) its features to make it suit their context.

4.9 Summary

The challenge of spatial heterogeneity—the difference in physical space configurations—poses a significant barrier to enabling distributed collaboration in blended realities systems. Overcoming this requires a theoretical understanding of how spaces can be "blended" to support shared interactions and place-making.

Currently, the field lacks the vocabulary to address these challenges. The Spatial Heterogeneity Framework provides a foundation for researchers to discuss and compare how blended environments can be modified to enable collaboration across disparate spaces. Additionally, it offers guidance for designers and developers navigating trade-offs and constraints associated with different space types.

The framework is designed for adaptability, allowing practitioners to start with different components depending on their particular priorities and constraints. It also accounts for various strategies to address spatial heterogeneity, including changes to technology, physical layouts, or social expectations.

As with any model, the framework prioritise conceptual clarity over procedural details, sacrificing some nuance in the process. A worksheet is provided in Appendix A to help practitioners step through the process of using the framework in more detail. However, this guide should not be mistaken for directly representing all possible distributed collaboration situations or uses of the framework. Like any good model, the framework and worksheet should be used as a common starting point, and the rules can broken by practitioners to serve a well-defined purpose.

The intent here is for practitioners to use the framework, in concert with theoretical principles of spatial interactions and place-making, to interrogate and advance best-practice solutions for distributed collaboration in blended realities. This is all in pursuit of a future where physical spaces are blended to enable shared spatial interactions and place-making, no matter where people are located.

Chapter 5

Translation: Blending Places and Sociocultural Norms

In Chapter 4, I showed how spatial heterogeneity challenges blended realities systems in their effort to coherently transmit embodied actions across distributed locations. As the blended zones of the spaces become more heterogeneous, fewer blended proxemics are afforded by the physical layout “for free” and must instead be enabled through blending techniques. Most current systems in this emerging area focus on blending techniques in the transform mode that prioritise spatial information, such as aligning facing direction (Pejsa et al. 2016, Orts-Escolano et al. 2016, Grønbæk et al. 2024, Sra et al. 2018) or maintaining consistent locomotion paths (i.e. enabling congruent body position and trajectory blended proxemics) (Wang et al. 2022, Congdon et al. 2018), coordinating movement around shared objects (i.e. enabling consistent vicinity) (Wang, Kim, Panda, Ofek, Franco & Won 2024, Grønbæk et al. 2023, Yoon et al. 2022, Fink et al. 2022), or redirecting deictic gestures to the correct target (i.e. enabling spatial deictic cues) (Yoon et al. 2021, Johnson et al. 2021).

As demonstrated in Section 2.2, supporting collaboration across distributed environments requires attention to both spatial information *and* sociocultural norms. In Chapter 3 I showed how sociocultural expectations, such as maintaining personal space, are captured in the translate mode and can come into conflict with transformation blending techniques. Figure 3.1 shows how actions adapted in the transform mode redirects the professor’s pointing gesture so that their avatar points at the intended object in the student’s space. However, in doing so, the professor’s avatar unwittingly points through the student’s body in their remote location. This creates a conflict between transform rules that prioritise spatial accuracy and translate rules that uphold sociocultural norms.

In this chapter, I argue that to design coherent systems for heterogeneous contexts, researchers and practitioners must understand when users prioritise spatial information or sociocultural norms in different collaborative situations. I therefore investigate how people weigh up the trade-offs between upholding spatial information and sociocultural norms during remote collaboration

(RQ3), and what designers and developers should consider when incorporating transformation and translation modes into blended realities systems (RQ4).

To answer these questions, Section 5.2 introduces a user study with 20 participants that draws on Comparative Structured Observation (Mackay & McGrenere 2025), counterfactual thinking (Oulasvirta & Hornbæk 2022), and card sorting methods to probe for the contextual factors inform peoples' preferences for interactions in the transform or translate mode. Section 5.3 presents the findings, showing how participants' preferences both align with and extend Burgoon's (1993) expectancy violations theory (EVT). In Section 5.4, I argue that extending EVT provides a structured way to explain how people evaluate sociocultural norm violations, and why they sometimes prioritise upholding these norms in the translate mode over preserving spatial accuracy in the transform mode. Finally, I show how participants' awareness of system mediation informed their judgments, often leading them to attribute violations to "system glitches" and relax their mental model of appropriate proxemics.

5.1 Aim and Focus

Mixed reality (MR) promises to create a sense of co-location for collaborators working across physically distributed spaces. In practice, the difference in spatial layouts can make it difficult to coherently enable embodied interactions. Consider a professor meeting with a student in MR. To faithfully transmit the professor's actions to the student's space, most systems would require the professor's and student's spaces to be identical and aligned (Grønbæk et al. 2023). The challenge here is that people's workspaces often have different physical layouts. The professor's office might have a desk, office chair, whiteboard, and guest chairs, while the student's space is smaller, with only a desk and an office chair. If the professor walks toward their whiteboard, and the system renders this movement without adapting to the student's spatial layout, the student might see them walking through a wall (**Figure 5.1**). Such mismatches in the spatial layout disrupts the intelligibility of remote participants' interactions within the local space.

Blended realities, an emerging subset of mixed reality (MR) interfaces, seeks to address the challenge of spatial heterogeneity by remapping elements of a local participant's environment to align with those in the remote collaborator's space (Grønbæk et al. 2023). This approach allows the system to display the remote participant's actions coherently within the local workspace, despite the physical disparities between remote and local spaces (Grønbæk et al. 2023). In the previous example (**Figure 5.1**), a blended realities system can map the professor and student's spaces to each other and transform the professor's actions, so that their avatar appears to stand in front of the student's wall, which they might use as a makeshift whiteboard. This approach allows the system to display the remote participant's actions coherently within the local workspace, despite the physical disparities between the remote and local spaces.

Recent research has introduced a diverse range of solutions to enable blended realities; from design (Wong, Sánchez Esquivel, Leiva, Grønbæk & Velloso 2024), engineering (Sra et al. 2018,

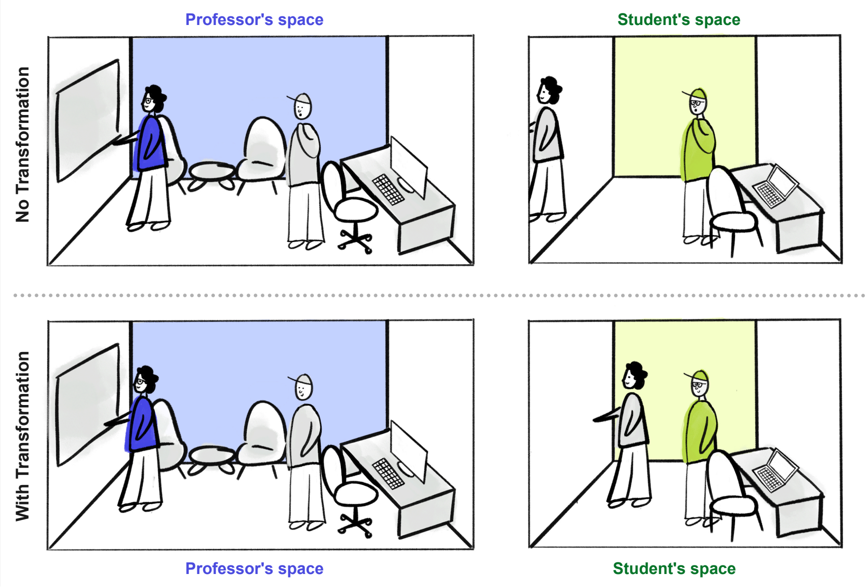


Fig. 5.1 Spatial heterogeneity makes it difficult to transmit embodied interactions across locations. Without a spatial transformation, the professor's avatar will appear to walk through the student's wall (top, right). In blended realities, the spaces are mapped so the professor instead stops in front of the wall, which the student might use as a makeshift whiteboard (bottom, right).

Grønbæk et al. 2023), and user experience perspectives (Wong, Grønbæk & Velloso 2024). Yet, much of this work has concentrated on the spatial transformation of actions, with less attention on the translation of *sociocultural norms* (e.g. expectations about personal space) that shape how people interact with each other in particular contexts.

This chapter builds upon the realisation that *when systems overcome spatial differences, new challenges arise related to the heterogeneity of sociocultural expectations.* These are moments where the sociocultural norms that people expect themselves and others to uphold differ, due to their dissimilar spatial layouts and therefore dissimilar actions being rendered in each space. For instance, let's imagine the student in the previous example sits at their desk in their own physical environment (**Figure 5.2**). Without appropriate translation, this interaction might appear strange from the professor's point of view. They might think, *"Why is the student sitting in my personal space?"* To account for the social etiquette attached to personal objects and spaces (Poretski et al. 2018), the blended realities system might add rules about the mapped chairs, so that when the student sits in their office chair, they appear to sit on the guest chair in the professor's office (**Figure 5.2**). Likewise, a similar rule would need to be introduced so that if the professor sits in their chair, their avatar might stand next to the student rather than appearing to sit on top of them. This is especially complicated since, without these rules, from the student's and professor's point of view, they are simply sitting in their usual chairs and not breaking any sociocultural rules.

The study presented in this chapter takes a step toward explaining how sociocultural norms, like sitting in an appropriate chair, are supported in blended realities. In particular, the complexity

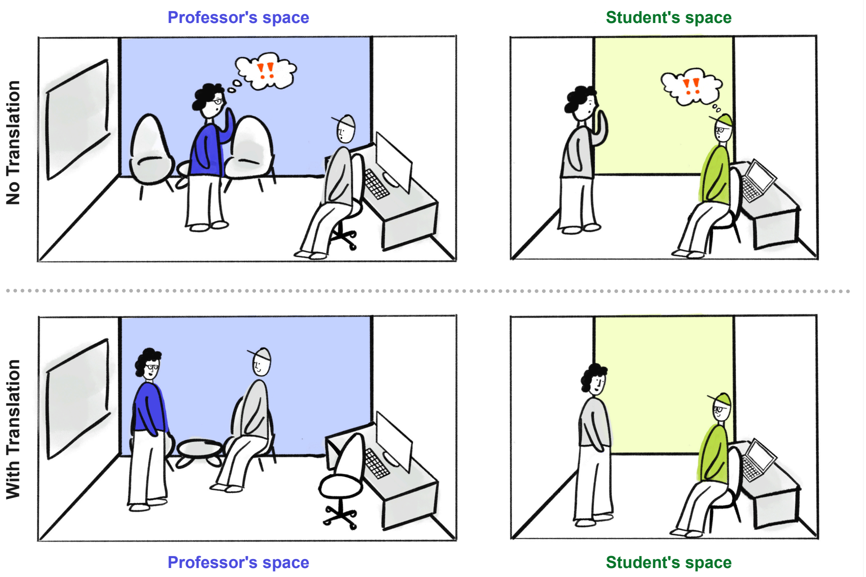


Fig. 5.2 In the condition without translation, the student's avatar appears to sit at the professor's desk. From the professor's perspective, this violates sociocultural expectations and feels like an invasion of personal space (top, left). However, for the student, the action seems natural since they are simply sitting at their own desk in their local environment (top, right). In the translation condition, the blended realities system adjusts the student's position so that their actions are in coherence with the sociocultural norms of the professor's office, making the action appear more socially appropriate (bottom, left).

of translating these norms across different environments, and the design trade-offs this creates. As outlined in Chapter 3 the complexity of the programmed rules that dictate how actions are transmitted increases, so too does the likelihood of these rules contradicting each other. Some actions may be faithfully preserved and transmitted to the remote environment, at the cost of other actions becoming distorted or omitted. For example, one rule might tell the system to uphold the spatial transformation, while the other overrides it, instead opting to uphold coded sociocultural norms. These contradictions force the system designer to compare the spatial and sociocultural affordances in the context of not just the spatial layout but also the collaborative scenario: what benefits do coherent, embodied interactions like sitting in the appropriate chair provide? How might the sociocultural rules change if the collaborators are not professor and student, but rather close friends? These design dilemmas highlight the need to understand what people value about embodied interactions and how these trade-offs should be negotiated. They also raise questions about the extent to which such experiences should be preserved and the ethics of potentially deceptive situations, where the local participant cannot see how they are represented in their remote collaborator's space and vice versa.

This chapter focuses on the trade-offs between *transforming* and *translating* actions across the distributed spaces, presented in Chapter 3. Coming back to the previous example, this can be demonstrated as the difference between *transforming* the spaces to preserve the facing-direction and trajectory of the professor (**Figure 5.1**), as they walk toward the whiteboard and *translating* the sociocultural norms to avoid the student accidentally sitting in the professor's chair (**Figure 5.2**).

To understand how people navigate the trade-offs between transformation and translation rules, I carried out a low-fidelity, counterfactual user study with 20 participants. This approach was inspired by Oulasvirta and Hornbæk's *Counterfactual Thinking* (Oulasvirta & Hornbæk 2022), Mackay and McGrenere's *Comparative Structured Observation* Mackay & McGrenere (2025), as well as card sorting design research methods Fincher & Tenenberg (2005). I opted for a low-fidelity qualitative study that gave us flexibility to test specific interactions and a wide scope to probe unexpected elements that contributed to the study participants' preferences. The goal of this research is to advance the design of intuitive systems for distributed collaboration, not by replicating face-to-face embodied interactions but by deepening the field's understanding of how people interpret and weigh up different types of non-verbal cues. In doing so, I aim to identify when transformation and translation rules might add value to a blended realities interface; answering how people weigh the trade-offs between upholding spatial information and sociocultural norms during remote collaboration (RQ3) and what designers and developers should consider when incorporating transformation and translation modes into blended realities systems (RQ4).

The findings show that participants' decision-making processes broadly aligned with Judee Burgoon's (1993) Expectancy Violations Theory (EVT) , where individuals assess situational factors, such as cultural context, power dynamics, personality traits and communication styles, to inform expectations about sociocultural norms. I extend this theory to blended realities, recognising that participants' awareness of system mediation introduced new layers to this evaluative process. Specifically, the inability to see into their partner's physical environment limited their ability to apply theory of mind (Wimmer & Perner 1983), leading them to withhold judgment or reinterpret ambiguous behaviours as unintentional. For example, when avatars pointed through a participant's body, many chose to overlook the apparent sociocultural norms violation, attributing it to a technical limitation. While some treated blended realities as a space with relaxed norms, others expressed unease when they could not reconcile mediated behaviours with their internal model of appropriate norms. These findings extend EVT by illustrating how an awareness of system mediation influences social evaluation in blended realities. Overall, this chapter makes the following theoretical claims, supported by empirical and artefact contributions:

Theoretical claims

- I show that an extension of Burgoon's (1993) Expectancy Violations Theory (EVT) can explain how people assess trade-offs between transformation (spatial information) and translation (sociocultural norms) in blended realities.
- An awareness of potentially divergent interaction perspectives (their own and their partner's) can obscure Theory of Mind, leading participants to relax their judgments of whether certain behaviours are intentional or unintentional.

Empirical and artefact contributions

- Identification of five themes from semi-structured interview data showing how participants' evaluations of interactions broadly aligned with Expectancy Violations Theory (EVT). These themes describe how participants considered: (1) public stakes and time pressure; (2) personal stakes, power, and hierarchy; (3) communication preferences; (4) predicted collaborator reactions; and (5) their model of appropriate proxemics versus actual proxemics.
- Identification of two themes showing how EVT should be extended to account for blended realities: (1) participants' familiarity with technology and their tolerance of glitches, and (2) how awareness of system mediation complicated their Theory of Mind.
- An extension of the EVT flow diagram, which provides researchers with a tool to design blended realities systems that predict when people might desire to uphold spatial information or sociocultural norms across heterogeneous spaces.

The contributions in this chapter demonstrate the trade-off between different modes of transmitting interactions across heterogeneous spaces and contexts, as well as concerns for maintaining transparency about the source of these interactions. By articulating how users interpret and respond to these trade-offs, I provide a foundation for designing more nuanced, socially attuned blended realities. In doing so, this work pushes forward conversations about when digital systems should faithfully replicate the dynamics of physical reality and when they should intentionally diverge.

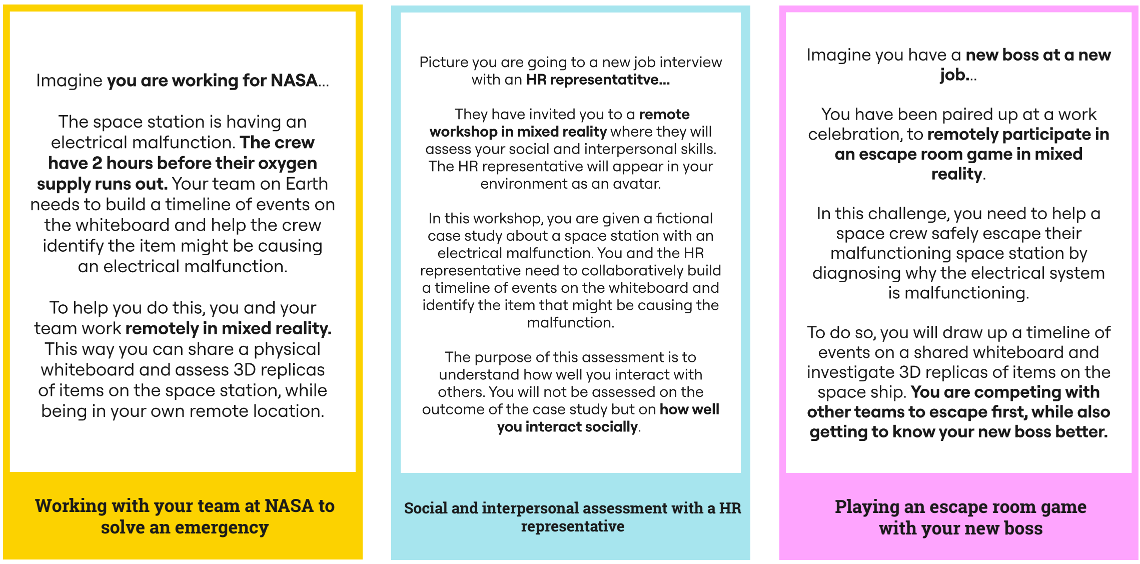
5.2 Study Design

The study design draws on recent work on theory-building in HCI (Velloso & Hornbæk 2025, Oulasvirta & Hornbæk 2016, Oulasvirta & Hornbæk 2022), which highlights the difficulty of applying existing theories to novel interfaces. Building on Oulasvirta and Hornbæk's (2022) concept of counterfactual thinking, I developed *counterfactual cards* that acted as a "speculation pump" to explore factors shaping preferences for transformation or translation modes during conflicts between them.

Inspired by card-sorting techniques, these cards communicated scenario-based prompts that instantiated contextual factors informed by literature on non-verbal interpersonal communication (Burgoon, Floyd & Guerrero 2016, Hall & Knapp 2013). These are grouped into: public stakes (e.g. time pressure or lives at stake) (Gower et al. 2022, Cone & Rand 2014, Wu et al. 2022), personal stakes (e.g. a promotion) (Pooladvand et al. 2025) or power structures (e.g. organisational hierarchy) (Burgoon & Dunbar 2006, Webb 2001, Pooladvand et al. 2025). This contextual information does not aim to be exhaustive; rather, it acts as a starting point for this explorative research. I leaned on the flexibility of the qualitative study to probe for other factors that people consider when assessing the different interaction options, since the prior work that informs the scenarios' contextual information is not an exact match with the particular problem space.

Each scenario card (Figure 5.3) placed participants in a blended realities system with the same task, where they needed to collaboratively diagnose a malfunction on a space station with their remote partner; examining virtual objects, and constructing a timeline on a whiteboard. However, in each scenario, this task is completed in different contexts:

1. **Critical:** Working with a NASA colleague to solve an oxygen-critical malfunction (high public stakes, low personal stakes, low power imbalance).
2. **Assessment:** Solving a case study with an HR assessor during a job evaluation (low public stakes, high personal stakes, high power imbalance).
3. **Game:** Teaming with a new boss in a competitive escape room (medium public stakes, medium personal stakes, medium power imbalance).



Imagine **you are working for NASA**...

The space station is having an electrical malfunction. **The crew have 2 hours before their oxygen supply runs out.** Your team on Earth needs to build a timeline of events on the whiteboard and help the crew identify the item might be causing an electrical malfunction.

To help you do this, you and your team work **remotely in mixed reality.** This way you can share a physical whiteboard and assess 3D replicas of items on the space station, while being in your own remote location.

Working with your team at NASA to solve an emergency

Picture you are going to a new job interview with an **HR representatitve...**

They have invited you to a **remote workshop in mixed reality** where they will assess your social and interpersonal skills. The HR representative will appear in your environment as an avatar.

In this workshop, you are given a fictional case study about a space station with an electrical malfunction. You and the HR representative need to collaboratively build a timeline of events on the whiteboard and identify the item that might be causing the malfunction.

The purpose of this assessment is to understand how well you interact with others. You will not be assessed on the outcome of the case study but on **how well you interact socially**.

Social and interpersonal assessment with a HR representative

Imagine you have a **new boss at a new job...**

You have been paired up at a work celebration, to **remotely participate in an escape room game in mixed reality**.

In this challenge, you need to help a space crew safely escape their malfunctioning space station by diagnosing why the electrical system is malfunctioning.

To do so, you will draw up a timeline of events on a shared whiteboard and investigate 3D replicas of items on the space ship. **You are competing with other teams to escape first, while also getting to know your new boss better.**

Playing an escape room game with your new boss

Fig. 5.3 Participants were shown three different scenarios that instantiated different contextual elements. In **scenario 1 (critical)** participants are working with a NASA colleague to solve an oxygen-critical malfunction (high public stakes, low personal stakes, low power imbalance). **Scenario 2 (assessment)**, asks participants to imagine they are solving a case study with an HR assessor during a job evaluation (low public stakes, high personal stakes, high power imbalance). Finally, in **scenario 3 (game)** participants are team up with a new boss in a competitive escape room (medium public stakes, medium personal stakes, medium power imbalance).

The degree to which each scenario instantiates the different contextual aspects (i.e. high, medium, low) is informed by the research team's interpretation. Each participant was asked during the interview what they thought the difference was between the scenarios, to understand how they personally assessed the different cards. This assessment informed the interpretation of the results and identification of any contextual aspects that the researchers had not considered.

For each scenario, participants rated their preference for transformation (A) or translation (B) interactions on a 7-point Likert scale (Figure 5.5). Ten pairs of interaction cards illustrated actions in the blended realities scenario, using the Rain rig in Blender[1], with each pair describing a **benefit** (e.g. direct pointing) and a **compromise** (e.g. pointing through their partner's avatar) (Figure 5.4).

The cards were swapped in and out during semi-structured interviews to prompt discussion, with participants adjusting ratings and explaining their reasons for doing so. This dynamic card swapping allowed us to observe how participants weighed up the trade-offs for their actions to be transformed or translated, in different contexts.

Condition	Option	Card 1 (benefit)	Card 2 (compromise)	Trade-off
1	Option A (transform)	While assessing the space station's items, you **can point directly** at the 3D replicas on the desk...	...but sometimes you end up **pointing through your partner** to do so.	**Pointing through their partner's avatar.**
	Option B (translate)	While assessing the space station's items, your **partner never perceives you as pointing through** them...	...but your partner **keeps thinking you're referring to a different item**.	**Misdirected pointing.**
2	Option A (transform)	While assessing the space station's items, you **can point directly** at the 3D replicas on the desk...	...but your **partner looks shocked** for some reason.	**Negative non-verbal feedback.**
	Option B (translate)	While assessing the space station's items, your **partner never perceives you as pointing through** them...	...but your partner **keeps thinking you're referring to a different item**.	**Misdirected pointing.**
3	Option A (transform)	While assessing the space station's items, you **can point directly** at the 3D replicas on the desk...	...but your **partner says "Why are you reaching through me?"**, yet from your point of view you are not touching them.	**Negative verbal feedback.**
	Option B (translate)	While assessing the space station's items, your **partner never perceives you as pointing through** them...	...but your partner **keeps thinking you're referring to a different item**.	**Misdirected pointing.**
4	Option A (transform)	While assessing the space station's items, you can **point directly** at the 3D replicas on the desk...	...but your **collaborator keeps pointing through you**.	**Their partner keeps pointing through them.**
	Option B (translate)	While assessing the space station's items, your **partner never perceives you as pointing through** them...	...but your partner **keeps thinking you're referring to a different item**.	**Misdirected pointing.**
5	Option A (transform)	You can **teleport to follow your partner** between the desk and the whiteboard...	...but when you teleport to the desk you **suddenly appear halfway through your partner's avatar**.	**They teleport into their partner's avatar.**
	Option B (translate)	You know you **won't teleport into each other**...	...but when you suggest walking to the desk, your **partner walks in a different direction**.	**Misdirected walking.**
6	Option A (transform)	While assessing the space station's items, you can **point directly** at the 3D replicas on the desk...	...but your **partner's arm is glitchy and warped through their body**.	**Their partner's arm is warped.**
	Option B (translate)	You know you **won't teleport into each other**...	...but when you suggest walking to the desk, your **partner walks in a different direction**.	**Misdirected walking.**

Fig. 5.4 Participants were shown a set of **counterfactual cards** that displayed an action illustrating either the transform (option A) or translate (option B) mode. Each option had a **benefit** that either *supported* spatial coherence (option A) or sociocultural norms (option B) and a **compromise** that either *compromised* sociocultural norms (option A) or spatial coherence (option B). I swapped out these cards in the order captured by column 1, asking the participant their preferences for the different options in each new set. The last column of the table summarises the tensions between each option.

---

[1] www.studio.blender.org/characters/rain/v3/

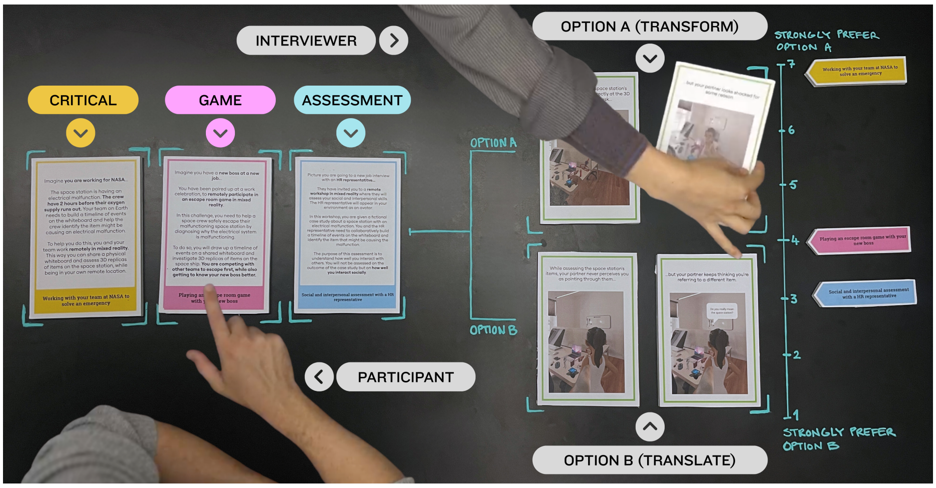


Fig. 5.5 Participants completed a table top activity activity, inspired by counterfactual thinking (Oulasvirta & Hornbæk 2022), card-sorting design research methods (Fincher & Tenenberg 2005) and Comparative Structured Observation (Mackay & McGrenere 2025). The interviewer placed the three different scenario cards (critical, game and assessment) on the left and then iteratively swapped out the interaction cards on the right. Option A cards, at the top, illustrated interactions in the transform mode and option B cards, at the bottom, illustrated interactions in the translate mode. These cards were structured so that choosing option A (transform) required a trade-off for sociocultural norms, and option B (translate) required a trade-off for spatial coherence. On the far right, participants moved the markers to indicate if they strongly preferred option A (7, at the top of the scale) or strongly preferred option B (1, at the bottom of the scale). Each of the three markers corresponded to one of the scenarios.

5.2.1 Method

I recruited 20 participants (10 men and 10 women) for a three-part user study (**Figure 5.6**), which consisted of: (1) a blended realities demo, (2) a tabletop counterfactual cards activity paired with a semi-structured interview and (3) a questionnaire that collected demographics and their experience using extended reality technologies. Since the study relied heavily on conversational elements, I primarily recruited participants who were native or fluent English speaking via a recruitment survey. Figure B.1 in Appendix B gives a full overview of the participants' demographics and experience using extended reality technologies. In total, each interview lasted 90 minutes, and the participants were reimbursed $40 each for their time.

Pre-work Participants read a plain-language statement, signed consent forms, and agreed to video/audio recording. The researcher explained the vision of blended realities, its role in remote collaboration, and challenges in mapping spaces, showing a video illustrating issues like avatar phasing. Transformation and translation modes were not explicitly introduced and only described as "perceptual distortions" corrected through changes to the way avatars are rendered.

Part 1: Blended realities demo Using a Meta Quest Pro and Shapes XR[2], the researcher demonstrated what a blended realities system might look and feel like. In separate rooms, the participant and researcher interacted as avatars in each other's physical space, exploring object pointing and controller-tracked gestures. The demo concluded when the participant reported that they understood the experience and had no more questions.

Part 2: Counterfactual cards Sitting at a table, participants were asked to imagine collaborating in blended realities (**Figure 5.5**). The researcher then showed them one scenario card first (scenario order randomised) and asked them to evaluate two interaction cards: Option A (transformation) and Option B (translation). After discussing each set of interaction cards, participants rated their preference on a 7-point Likert scale while thinking aloud (7 = strongly preferred A (transformation), 1 = strongly preferred B (translation). The researcher then showed the participant another scenario, repeated the discussion and rated the interactions in the first condition, according to the new scenario context. They repeated this until all three scenarios were presented and sitting on the table.

Once all the scenarios were on the table, the participant was asked to compare them and reflect on their similarities or differences. The researcher then swapped out the interaction cards, progressively showing them the different interaction conditions shown in **Figure 5.4**. The order of these interaction cards was kept constant across scenarios. The researcher used a semi-structured script **(see Appendix B)** to probe participants' reasons for choosing option A or option B.

Part 3: Questionnaire The participants completed an electronic questionnaire using Qualtrics. The survey collected information on age, gender, and experience using MR.

5.2.2 Data Collection and Analysis

During the study, both video and written data were collected. Video recordings were transcribed and analysed using a user research tool called Dovetail[3]. Following Braun and Clarke's (2006) Thematic Analysis, I re-read the transcripts, performed low-level coding, refined these codes, and discussed them with others in my research team before forming high-level themes.

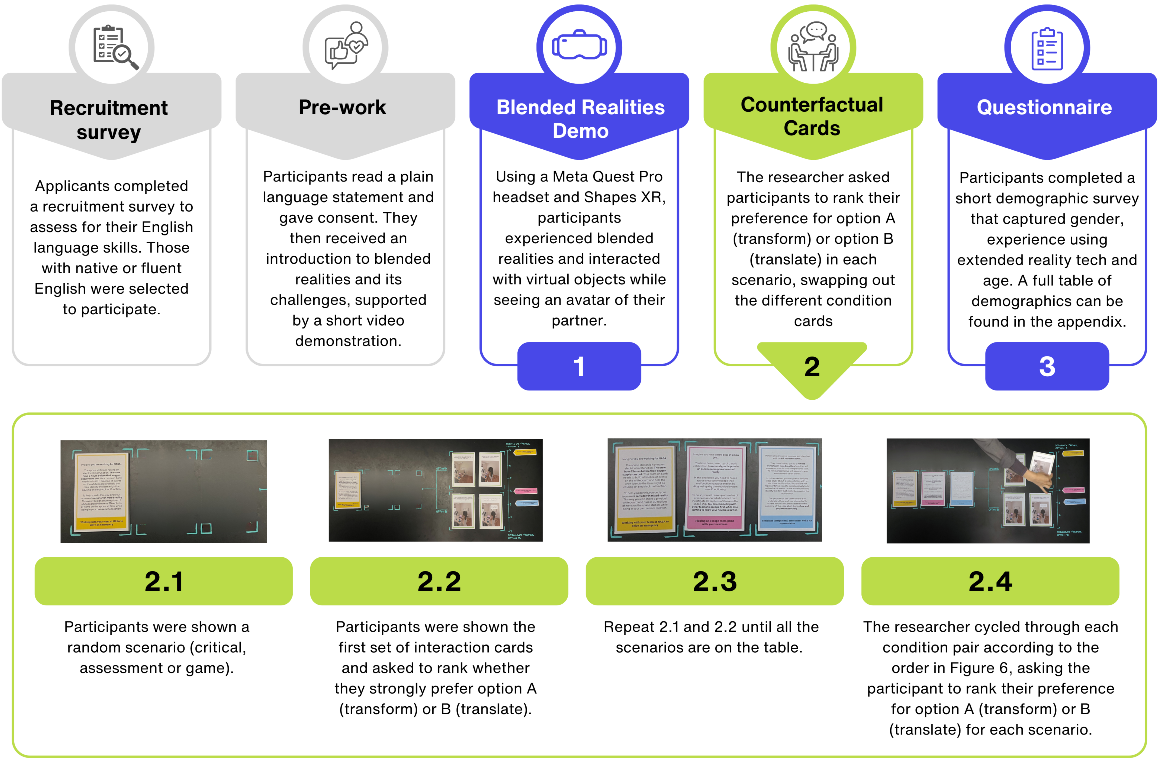
5.3 Results

I analysed the results to answer (RQ3) and (RQ4). In doing so, I identified seven thematic areas that explain how participants weighed up the trade-offs between transformation and translation modes. These themes include: (1) public stakes and time pressure, (2) personal stakes, power and hierarchy, (3) different communication preferences, (4) internal model of appropriate proxemics

---

[2] www.shapesxr.com

[3]



Recruitment survey

Applicants completed a recruitment survey to assess for their English language skills. Those with native or fluent English were selected to participate.

Pre-work

Participants read a plain language statement and gave consent. They then received an introduction to blended realities and its challenges, supported by a short video demonstration.

Blended Realities Demo

Using a Meta Quest Pro headset and Shapes XR, participants experienced blended realities and interacted with virtual objects while seeing an avatar of their partner.

1

Counterfactual Cards

The researcher asked participants to rank their preference for option A (transform) or option B (translate) in each scenario, swapping out the different condition cards

2

Questionnaire

Participants completed a short demographic survey that captured gender, experience using extended reality tech and age. A full table of demographics can be found in the appendix.

3

2.1

Participants were shown a random scenario (critical, assessment or game).

2.2

Participants were shown the first set of interaction cards and asked to rank whether they strongly prefer option A (transform) or B (translate).

2.3

Repeat 2.1 and 2.2 until all the scenarios are on the table.

2.4

The researcher cycled through each condition pair according to the order in Figure 6, asking the participant to rank their preference for option A (transform) or B (translate) for each scenario.

Fig. 5.6 A recruitment survey was completed to identify native or fluent English-speaking participants. During the pre-work participants read a plain-language statement and signed consent forms. The study itself consisted of three stages: (1) a blended realities demo, (2) a tabletop counterfactual cards activity paired with a semi-structured interview and (3) a questionnaire that collected demographics and their experience using extended reality technologies. The figure breaks down each step of the counterfactual cards activity in more detail.

and actual proxemics, (5) prediction of their collaborator's actions, (6) Familiarity with the tech and acceptance of glitches, and (7) Theory of Mind (ToM) approach.

As explained in Section 5.2, I focus primarily on the qualitative data, since the purpose of the quantitative results was to support the semi-structured interviews and serve as a tool for prompting discussion. Accordingly, I do not present an extensive statistical analysis to formally test or falsify hypotheses. Instead, numerical data are reported to illustrate the patterns observed by the interviewer during the tabletop activity with participants. Consistent with McKay and McGrenere's (2025) work on Comparative Structured Observation, I encourage readers to place greater emphasis on the qualitative findings. However, unlike Comparative Structured Observation, the goal was not to declare one mode superior, but to use dynamic counterfactuals (Oulasvirta & Hornbæk 2022) to develop an initial theory of how users negotiate the trade-offs between transformation and translation in blended realities. The quantitative results are not designed to support statistical inference but rather to facilitate an information-rich process for surfacing unknown elements and contributing to the development of theoretical constructs in this emerging area of research.

5.3.1 Theme 1: Public Stakes and Time-pressure

In the *critical* scenario, all participants explained that time pressure and life-or-death risk to others increased their likelihood to prefer transformation over translation. All participants expressed that in high public stakes and time-pressure situations, accuracy should be favoured over *'social niceties'*.

> *P9W: "[the transformation mode is] very clear and especially in a limited time situation, you need that quickly. You don't have as much time for being nice about things [...] even if people don't enjoy it, [pointing through someone is] the correct option; it's better because it's faster."*

Interestingly, not all participants had a strong preference for transformation in the *critical* scenario. Some placed their markers slightly toward translation, especially if their partner displayed negative feedback. When probed further, P1W, P9W, P10W, P19M explained that social conflict could hinder progress on the mission and so they preferred to avoid it if possible.

> *P1W: "I think [the transformation mode] could turn into a distraction really quickly if someone is kind of bothered by it. [...] emotions are gonna be high [in the critical scenario] because of how high pressure and how stressed and upset everyone is."*

5.3.2 Theme 2: Personal Stakes, Power and Hierarchy

Participants weigh up the personal stakes involved in the activity and how they want to be perceived by a person with more power. Participants adjusted their score depending on the perceived personal risk of a social faux pas and the other person's power over them. For example, P5M weighs up the risk of losing their job and the influence their boss has on this outcome.

> *P5M: "There's power dynamics involved when you're dealing with your boss, [...] fulfilling what [they are] wanting out of the scenario because [...] your job's in their hands."*

Participants agreed that the assessment scenario had high personal stakes, impacting their desire to come across as polite and cognisant of social norms. Perhaps, as P14W suggested, to help them avoid an awkward social situation that would reflect poorly on their character.

> *P14W: "[...] because it is an HR representative, you worry they might judge you for [pointing through them]."*

Conversely, according to participants, the NASA hierarchy in the *critical* scenario felt equal and focussed on public rather than personal stakes, making it more socially acceptable to point through someone's avatar.

> *P13W: "I think [pointing through them] is OK because we're the same hierarchy. There's no one who's like above us [...] [my team is] gonna mostly think about getting the job done and saving people."*

5.3.3 Theme 3: Different Communication Preferences

Some participants expressed a slight preference for a particular communication style. While preferences for accurate (transformation) or interpersonal (translation) tactics varied by scenario, many participants often talked about their preference for a particular communication style. For example, those most strongly preferring transformation (P11W, P12W, mean score = 6.3) prioritised accuracy, even in high-stakes or emotionally charged situations. P11W argued that translation could create confusion and harm interactions:

> *P11W: "if there's the confusion about [...] what I'm pointing to, that would [...] hinder our social interaction much more than if it simply appears as pointing through them."*

For P12W *"clear communication trumps everything else"* and they want to avoid *"assumptions"*. Yet, interestingly, their definition of *"clear communication"* does not necessarily mean accurate pointing; it could mean an accurate display of how their partner is feeling so that they can address these emotions directly.

> *P12W: "[the transformation mode] is ideal for social interpersonal skills. If someone looks shocked, you can, [...] actually lead down to a line of questioning and problem solving and demonstrate [interpersonal skills]."*

Conversely, those most strongly preferring translation (P3M, mean score = 3.2 and P10W, mean score = 3.1) prioritised preserving relationships and proxemic norms. P10W considered discomfort a barrier to collaboration, even in time-critical tasks:

> *P10W:"I think in [the NASA] scenario, if this kind of interaction is causing my partner to be uncomfortable, I think that would disturb the collaboration and [...] even though it might be more efficient, I feel like it would have its [...] drawbacks would kind of slow our work down."*

P3M avoided proxemic violations entirely, even in competitive scenarios:

> *P3M: "Well, I mean, if that was in real life, it would be inappropriate [...] it feels real even though they are avatars, so it would be uncomfortable for me."*

Most women (e.g., P1W, P6W, P7W, P9W, P11W, P12W, P13W, P14F, P16W) viewed a faux pas from transformation as an opportunity to demonstrate interpersonal skill:

> *P14W: "I don't really see many downsides any more just because you're pointing at the right thing now and getting kind of like the, emotional feedback as well, yeah."*

In contrast, men preferred to avoid situations where partners might react negatively:

> *P19M: "I wouldn't want to be interviewed with someone that is, uh, expressive. Or like responses negatively towards my performance, even if I'm doing a bad job."*

Men appeared to more strongly prefer translation when there was a negative response to their actions in condition 2 and 3 (**Figure 5.4**). The difference in response between women and men could be related to the avatar appearing like a woman in the examples. P7W explained their response would change if their collaborator were a man or a woman, however this was not explicitly echoed by any of the men.

Finally, participants linked their communication style to personality traits such as being confrontational, competitive, or avoidant, suggesting deeper personal dispositions influence communication preferences.

5.3.4 Theme 4: Predicting The Collaborator's Reaction

Participants use predictions of their collaborator's response to guide their preferences, drawing on an understanding of the partner's personality, perceived goals and prior behaviour. Several participants (P1W, P2M, P3M, P4M, P7W, P8M, P12W, P13W, P18M) emphasised the importance of considering their partner's personality when determining their preference for transformation or translation modes.

> *P2M: "[...] what I'm finding difficult about answering these is I feel like it's so dependent on the personalities of the people involved.*

This was particularly evident in the *game* scenario, where participants changed their preferences depending on whether they thought their boss was competitive, easy-going, or sensitive to social norm violations. Some thought winning the competition and having fun would leave a good impression on their boss (P2M, P6W, P7W, P8M, P10W). Others aimed to avoid offending their boss, paying attention to sociocultural norms (P1W, P4M, P9W, P14W, P15M, P16W, P18M).

Participants used their partner's behaviour to build an opinion of their personality, as well as their goals and level of technological familiarity. For example, P7W initially demonstrated a strong preference for transformation and prioritised *'accurate'* communication. However, when their partner responded negatively to their behaviour, P7W's perception of their partner's personality and competency shifted, influencing how they plan their actions.

> *P7F: "I think for [the boss and HR] ones, they're more uptight and I have to consider their emotions, I have to consider how they feel and not only how they feel about me or the conversation, but how they feel about the technology."*

5.3.5 Theme 5: Model of Appropriate Proxemics and Actual Proxemics

Participants create a model of *appropriate* proxemics, checking whether this model aligns with the *actual* proxemics they see and what they theorise their partner sees. A range of factors influenced participants' perceptions of appropriate proxemics within each scenario. These factors include the previously identified themes: public stakes, time pressure, personal stakes, hierarchy, perceived power, communication style and their partner's personality. The interviews showed that participants applied these factors, both consciously and unconsciously, to construct a mental model of what constituted appropriate proxemics for the task at hand.

In the *critical* scenario, all participants were willing to relax proxemic norms in a life-or-death, time-pressured context. In contrast, when interacting with figures perceived as having more power, such as an HR representative or boss, participants adopted a stricter interpretation of appropriate behaviour, tending to prefer the translation mode. As P18M explained, there are *'rules'* or culturally coded norms that define what is considered appropriate in different contexts.

> *P18M: "Work environments have rules about being [...] suitable. It's just a different vibe, different rules."*

For instance, reaching through someone's avatar might be okay with friends but not in a professional situation, where their personal reputation may be at stake. When probed further, participants (P3M, P4M, P5M P7W, P8M P9W, P10W, P13W, P14W, P18M, P19M) explained if they were more familiar with their partner, then obeying strict social distances would not matter. According to P5M and P7W, this could be because they know where the boundaries are, and how their interpersonal actions will be interpreted by the other person.

> *P5M: "So my closest friend and, and friends, like there's nothing I could really do that would like in this kind of like scenario that would sort of like offend them."*

> *P7W: "I've had chill bosses where they don't really mind or they might find [reaching through someone] funny. I'm fine with it, but if I knew that they would be uncomfortable or if I could sense that they were more professional or official that I wouldn't want to [reach through them]."*

Participants (P1W, P2M, P4M, P6W, P7W, P8M P10W, P18M) described a heightened sense of risk when interacting with strangers, due to uncertainty about the other person's personality and potential reactions. This perceived risk was further amplified when the stranger held a position of power. As a precaution, participants tended to apply their internal model of appropriate proxemics more rigidly in unfamiliar professional settings, deliberately avoiding behaviours that might violate an acceptable *'social distance'*.

However, participants described feeling anxious if their partner reacted in a negative or unexpected way. Some theorise this is because they cannot see whether the *actual* proxemics from their perspective and their partner's perspective match.

> *P1W "I think it really depends on where they are in the space as well. I wonder if, like, they would have a bespoke problem because in their perspective, I'm placed somewhere a bit weird and unusual."*

P1W goes further to illustrate how it could be tricky to navigate these different perspectives when they can't see their partner's *'spatial rules'*.

> *P1W: “Here, it’s trickier because it might be like, "oh, [you] might need to move away", but [they might say] "oh no, you’re not in the way” [...] because if you don’t have that like-for-like [...] if you ask the person to move, it may not work because the spaces are different and there’s different [...] spatial rules.”*

5.3.6 Theme 6: Familiarity with Tech and Acceptance of Glitches

Participants’ familiarity using technology and readiness to accept glitches modifies their perception of acceptable proxemics and behaviours in MR. Many participants (P3M, P4M, P5M, P7W, P9W, 10W, P12W, P13W, P14W, P15M, P17M, P18M) expressed that as MR becomes more common, people may become more accustomed to virtual anomalies, such as phasing through someone’s avatar. Participants (P4M, P5M, P7W, P9W, P10W, P12W, P18M, P19M) often compare this acceptance to familiarity with video games and understanding the limitations of the technology.

> *P12W “I just could get used to it. And there’s plenty of games that you can play where you’re just like, you know, they, they don’t have any sort of like collision or anything like that, so phasing through people is just, you know, the limitations of the software at the time.”*

As a result, these participants express a relaxed attitude toward real-world proxemic norms, viewing violations as an expected artefact of virtual spaces. They assume their actions are likely to be excused as the computer’s fault or even use the glitch as a source of amusement to “break the ice” with their partner.

> *P15M “Because it’s in a virtual reality space like even if you [...] cross your boss’ [...] avatar or anything [...] because it’s not real [...] it won’t be like a serious thing like people um they wouldn’t perceive that as too inappropriate.”*

> *P1W “You could have a little joke with them, and that actually helps break the ice with them a lot more.”*

In the *critical* and *assessment* scenarios, most participants (P1W, P3M, P4M, P6W, P7W, P8M, P11W, P12W, P13W, P14W, P15W, P18M) assumed their partner was familiar with the technology and would understand when glitches occurred. However, in conditions 2 and 3, when their partner responded negatively to their interactions, several participants (P4M, P6W, P7W, P11W, P12W, P15W, P18M) revised their assumptions about their partner’s technological familiarity. This sometimes caused frustration and participants questioned whether they shared a mutual understanding of appropriate proxemics with their partner.

> *P6W “So what might not be shocking to me might be shocking to them. So maybe they’re working in a mixed reality setting for the first time. I know that if I can point through*

> *them and I can maybe walk through them, but they might not know that. It just makes things a little bit difficult because now we're not on the same page and I have to [...] explain it to them."*

While some participants were more accepting of technical glitches, others argued that because MR simulates real-world contexts, it should also maintain real-world social norms (P3M, P6W, P7W, P8M, P14W, P15M, P16W, P17M, P18M, P19M, P20M). However, most participants reported preferring transformation when their partner pointed through *them* and were largely unbothered by this social norm violation. This may be attributed to an awareness that the interaction was 'not real' or a reluctance to speak up and correct their partner.

> *P8M "I'm completely fine with it. I don't have any issues with that because I know that it's not being done on purpose. It's just a system which does it."*

> *P13W "I don't think I'd care. I think I'd be OK, because [...] I wouldn't feel offended or anything like that [...] because I wouldn't confront anyone."*

There were some participants (P4M, P7W) who mentioned they may feel uncomfortable if the other person is pointing through them due to power or gender differences.

> *P4M "In a way I am judging the boss too, cause like the boss could be [...] weird or [...] creepy [...] I think that would be alarming to me if all of a sudden the boss was like putting their arm on me all the time in this reality."*

> *P7W "If it's a girl, I don't care. If it's a guy, it depends if they know they're pointing [through] me [...] if it's a creepy guy pointing through me [...] I wouldn't be totally comfortable with it."*

5.3.7 Theme 7: Theory of Mind Capability

Participants describe using Theory of Mind to interpret their partner's perspective, intentions, or temperament. However, this process was often complicated by uncertainty about whether an observed behaviour reflected their partner's intentional action or a technical glitch. Participants consistently reported relying on their partner's avatar, particularly facial expressions, gestures and gaze direction, to infer how their own actions were being received, and to understand the reasons behind their partner's reactions.

Several participants (P3M, P4M, P5M, P7W, P10W, P12W, P13W, P14W, P19M) expressed that they could not fully trust that the avatar's movements or facial expressions are accurate reflections of their partner's real behaviour or emotional state. In particular, P19M questioned the avatar's facial expression meaningfully captures their partner's emotional response.

> *P19M: "I'm not 100% sure whether the technology can capture that emotion as effectively as if it were face-to-face."*

While P5M addressed more broadly how avatars risk misrepresenting the subtle cultural and behavioural cues that typically inform in-person social interactions.

> *P5M: "So much is carried by, you know, how we look and appear and act and move [...] in a virtual sense, like what we do with that will sort of alter how we perceive other people [...] it allows for people to misrepresent who they might be."*

This issue was amplified by the fact that participants could not see into their partner's physical space, preventing them from verifying real-world proxemics or non-verbal cues. As highlighted by P14W and P4M, this lack of shared situational awareness could be frustrating and even anxiety-inducing, particularly if technical issues are misattributed to them in ways that could negatively impact perceptions of their professionalism or character.

> *P14W: "I think it would [...] make me panic a bit, and I'd wonder why they were almost like blaming a technical issue on me, as if I was the one running the mixed reality [...] I'd expect them to have run into that issue before, and I'd be a bit offended that they were instantly blaming me."*

> *P4M: "Oh my goodness, I hope they don't think that I'm, yeah, like stroking them or something weird like that [...] I would already feel uncomfortable going into this situation with a new boss and an HR representative, being hyper-aware of how I am presenting myself."*

This led some participants to reflect on the challenge of distinguishing between glitches and intentional behaviours. For example, P3M noted that the lack of feedback made it difficult to know whether they were being perceived as intended, questioning the validity of blended realities without this feature.

> *P3M: "I guess there's no feedback. If you're not coming across like you think you are [...] it's a weakness in the technology [...] you're better off just being on a video call."*

5.4 Extending Expectancy Violations Theory to Blended Realities

I interpret the findings through the lens of Expectancy Violations Theory (EVT), which helps to understand how participants weighed communicator, relational, and contextual factors when evaluating translation and transformation modes. In blended realities, mediation adds ambiguity, prompting participants to uphold proxemic norms but apply them more flexibly when intent was unclear. While some believed users might adapt to proxemic violations over time, the findings and prior work (Bailenson et al. 2001) suggest such violations still trigger a physiological response. Even when suspecting the system as the source of a misstep, some participants remained wary of their partner's intent and concerned their own actions might be misrepresented. This raises ethical questions about when to use translation modes and how to signal whether a violation stems from machine processes or human intent.

5.4.1 Participants Weigh Up Factors According to Expectancy Violation Theory

Most of the qualitative themes from the study align closely with Expectancy Violations Theory (EVT) (**Figure 5.7**). As proposed by the theory, participants drew on communicator characteristics (communication style), relational characteristics (personal stakes, perceived power, and hierarchy) and context characteristics (public stakes and time-pressure) to form their expectancies and to assess communicator reward valence. Burgoon (1993) describes expectancies as "an enduring pattern of anticipated behaviour" (1993, p.31), which I observed in participants' mental models of appropriate proxemics, shaped by these three categories of characteristics.

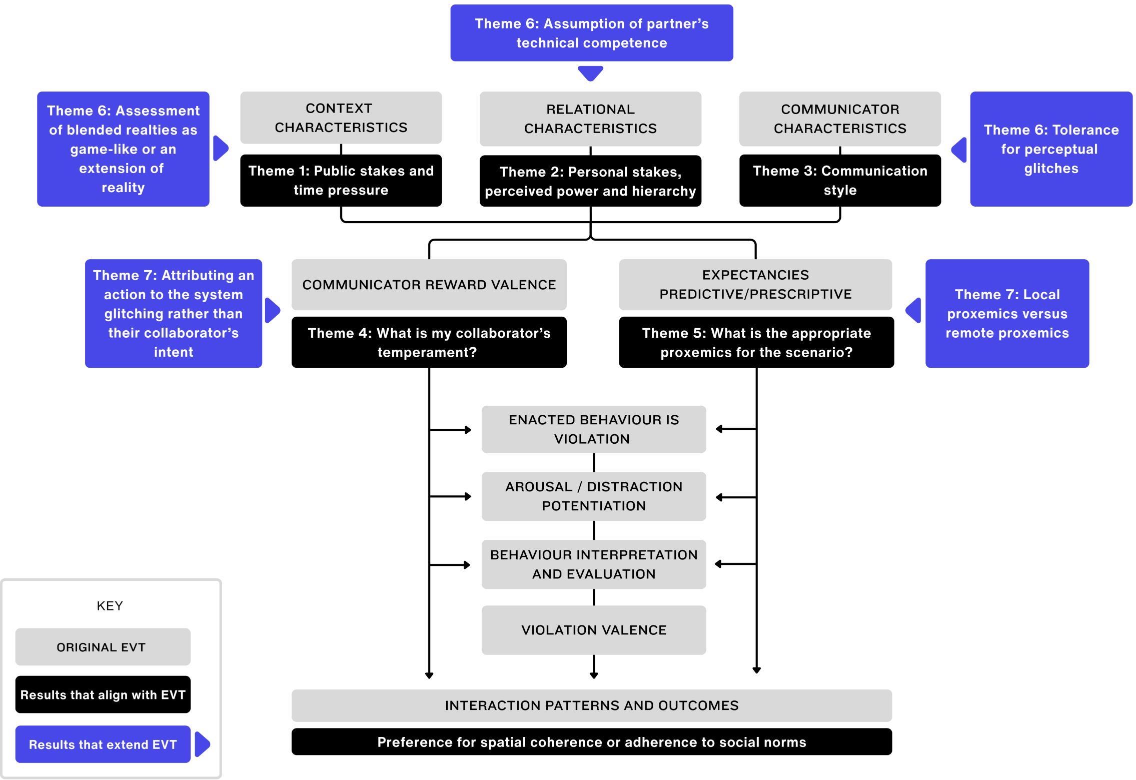


Fig. 5.7 Themes 1-5 from the results align closely with Burgoon's (1993) EVT. However, I extend the theoretical framework according to Theme 6 (familiarity with tech and acceptance of glitches), which influences how participants evaluate the context characteristics. I also extend the framework according to results in Theme 7 (Theory of Mind capability) that show how blended reality systems influence the communicator reward valence and expectancies. Specifically, the participants' willingness to attribute an action to the system glitching may result in a slightly more positive communicator reward valence and the potential difference between local and remote proxemics alters their model of expected appropriate proxemics.

Similarly, the communicator reward valence, participants' judgment of their collaborator's temperament (Burgoon 1993), was also constructed through the same lens. Participants actively predicted how their collaborator might react based on communicator, relational and contextual cues. For instance, in the *critical* scenario, participants often drew on relational factors such as a reduced sense of hierarchy, context factors like urgency due to time pressure or their direct communication style to infer that their collaborator would not react negatively to proxemic violations. These factors led them to anticipate a positive or neutral communicator reward valence for the interaction.

Consistent with EVT, participants interpreted and evaluated actual behaviours against their expectancies and communicator reward valence to assess the extent of the violation valence, i.e. how acceptable an interaction would be. This evaluation influenced their decision to prefer the translation mode and uphold proxemic expectations or to go with another option that may satisfy other situational characteristics, for example, timeliness or accuracy in the *critical* scenario. These findings illustrate how classic expectancy-based reasoning aligns with the way participants consider different priorities in blended realities.

5.4.2 Compensating for the 'Machine' Relaxes Expectancies

The findings generally support EVT, however system mediation influenced how participants judged their own and their partner's actions at all stages of the EVT framework (**Figure 5.7**). *Communicator characteristics* were influenced by participants' tolerance of perceptual glitches. For example, during teleportation in condition 5, some found accidental avatar collisions particularly unsettling, while others with gaming experience dismissed these glitches and instead valued the efficiency that this communication style provided. *Relational characteristics* were impacted by assumptions about their partner's technological competence. *Contextual characteristics* varied according to whether participants saw blended realities as playful, game-like spaces with relaxed norms or as extensions of the physical world requiring adherence to formal expectations. Each of these characteristics informed participants' proxemic *expectancies* and whether perceptual glitches that violate proxemic boundaries, such as phasing through avatars, should be tolerated. Others, especially in formal contexts, expected the system's visual realism to uphold conventional proxemic rules. Similarly, the *communicator reward valence* was often softened by attributing social missteps to system errors, leading participants to excuse both their own and their partner's violations. As such, the participants' awareness of a mediating system appears to impact each stage of EVT.

The way blended realities mediate expectancy violations raises important considerations for the design of these systems. Uncertainty about the remote physical environment limited participants' ability to accurately apply Theory of Mind (ToM). In the study, this ambiguity in distinguishing system-driven actions from human intent often led to a more forgiving view of norm violations: "it's just the computer glitching, right?" However, recognising that multiple proxemic "realities" could coexist across physical spaces, some participants felt disempowered by their inability to control the representation of their actions according to their own desired model of appropriate proxemics. This raises issues of accountability and trust; what if the action was not the computer glitching but their partner intentionally performing an expectancy violation? Perhaps there should be clearer distinctions between human and system behaviour. Further research is required to study the boundary between human intent and technical glitches. I suggest that to do so, EVT should be extended with an extra stage to spotlight and deliberately consider how people evaluate mediating systems as almost a ‘third participant’ in the scenario. **Figure 5.7** shows how I propose the framework could be extended specifically for Human-Computer Interaction studies, based on the results.

5.5 Summary

Blended realities offer a compelling opportunity to support embodied interaction across distributed, heterogeneous contexts. However, in Section 2.2 I show how current approaches tend to focus on spatial heterogeneity and are yet to examine the sociocultural dynamics that influence collaborative interactions. Additionally, in Chapter 3 I demonstrated how solutions that uphold spatial information or sociocultural norms can come into conflict, generating “social noise”, which can be confusing for collaborators to decipher.

To address these problem spaces, this chapter investigated how people weigh up the trade-offs between transform and translate modes of transmitting embodied actions in blended realities. The results of this investigation extend Expectancy Violations Theory (EVT) (Burgoon 1993), showing how an awareness of the mediating system can obscure Theory of Mind, leading participants to relax their judgments of whether certain behaviours are intentional or unintentional. This extended theoretical framework can be used to explain how people assess the trade-offs between upholding spatial information (transform mode) and sociocultural norms (translate mode) in different collaborative contexts.

Chapter 6

Discussion

This thesis presents the results from two investigations that show how embodied actions are transmitted across heterogeneous contexts in blended realities and examines the design trade-offs this involves. In Chapter 4, I showed how blended realities systems assume that the distributed spaces will have a certain degree of spatial heterogeneity. This means some blended proxemics will be afforded by the spatial layout "for free" and others need to be enabled with a blending technique that adapts the local collaborator's actions to the remote collaborator's spatial layout. In Chapter 5 I revisited how spatial information and sociocultural norms are navigated in blended realities, aiming to understand participants' preferences when these goals conflict with each other. The findings from the presented user study show how participants loosely follow Burgoon's (1993) expectancy violations theory (EVT) when assessing these trade-offs. However, an awareness of the mediating system influenced the way participants interpret actions, leading them to relax their social expectations.

In this chapter, I discuss what designers and developers should consider when adapting embodied actions across heterogeneous contexts, further addressing (RQ4). In Section 6.1 I argue that deciphering system-automated actions from human intent is an important ethical consideration for blended realities space design, since ambiguity can excuse missteps and even conceal harmful behaviour. In Section 6.2 I reiterate how the Spatial Heterogeneity Framework and extended EVT provide a conceptual foundation for navigating when actions should be adapted. I discuss other options to coherenetly transmit embodied actions, such as physically rearranging spatial layouts or changing the collaborative activity itself. However, these options place the burden of adaptation on the user, stressing a need for future work to consider when these options are best implemented. Finally, in Section 6.3 I provide commentary on the extent to which the investigation results can be generalised, making suggestions as to how future research can use the presented frameworks and qualitative results to inform larger-scale quantitative studies.

6.1 Should Adaptations in the Translate Mode Look Like 'the Machine'?

The ability to determine whether an action in blended realities is system-automated or human-controlled connects to a broader debate about how closely immersive environments should mimic reality (Slater et al. 2020). Prior research indicates that including the physical environment in systems for remote collaboration can promote engagement (Kiyokawa et al. 2002, Grønbæk et al. 2024). However, blended realities may introduce interactions that negatively impact interpersonal relationships (Slater et al. 2020). For example, the system might prioritise spatial coherence over social norms, rendering a person's avatar so that they point directly at an intended object but in doing so appear to reach through their collaborator's body. Although it is tempting to dismiss this interaction on the grounds of it not being "real", prior research shows that these expectancy violations may feel more intense due to the immersive nature of extended reality (Blackwell et al. 2019). Indeed, while several participants in Section 5.3 believed that the effects of these expectancy violations would diminish over time, others worried that persistent breaches or actions perceived as deliberate could undermine trust, comfort, and professionalism, negatively impacting their collaborator's impression of them. As such, expectancy violations in blended realities may have consequences that extend beyond the virtual environment, informing real-world interpersonal relationships. While it would be excessive to stop the development of blended realities based on these findings, researchers should be aware of how systems that mimic embodied actions may have interpersonal impacts beyond the designed environment.

Alternatively, while upholding sociocultural norms in blended realities may be well-intentioned, these engineered solutions can also introduce new forms of deception. Several participants expressed concern that if a collaborator knowingly crossed personal boundaries and the system automatically corrected this behaviour, they would be unable to accurately evaluate or respond to the unwanted conduct. This introduces new challenges for recognising harassment, particularly given the ephemeral nature of virtual spaces, which already makes reporting difficult (Blackwell et al. 2019). Existing work on ethics and harassment in immersive environments acknowledges that norm violations often stem from inexperience, with people unaware of how their actions are perceived by others (Slater et al. 2020, Blackwell et al. 2019, Akselrad et al. 2023). However, the findings in this thesis suggest that the reverse dynamic is also possible. A person may intentionally engage in inappropriate behaviour that the system then conceals through auto-correction. Even when the system *did* render the collaborator's avatar pointing through the study participant's body, participants dismissed this as a technical glitch, being uncertain whether the action reflected an error in the system or their collaborator's intent. Yet, when prompted to reflect, many voiced concern that such behaviour could be deliberate without their awareness. These insights highlight the ethical implications designers face when embedding sociocultural norms into blended realities systems. As the philosopher Paul Virilio (1989) reminds us, "to invent the ship is to invent the shipwreck". While blended realities promises to improve remote collaboration, designers must equally anticipate the novel risks they create and work proactively to mitigate them.

These ethical concerns about reproducing sociocultural norms in blended realities bring into question what these adaptations should look like and how their origin is communicated. Should adaptations look like regular actions or should the system in some way indicate how they have been changed? Many of the participants in Chapter 5 discussed the benefits of being able to identify whether an expectancy violation was caused by the system or their partner. Their concerns are further supported by Reidel et al.'s (2014) research findings, which shows that people are better at predicting the trustworthiness of humans than avatars and are at a disadvantage when discriminating trustworthiness based on avatar representations. Participants believed that "unmasking" the intent of interactions would help them interpret behaviour more accurately and maintain social harmony. This need is especially salient when the cost of misinterpretation is high and could have social impacts beyond the collaborative environment. Yet overly explicit cues could, in turn, disrupt the natural flow of non-verbal interaction. It is conceivable that this would add to the cognitive load of collaboration. Too many signals could lead to effects akin to "Zoom fatigue" (Bailenson 2021), where people are overwhelmed by the amount of time spent examining close-up facial expressions. Future research might investigate how to establish transparent intent when collaborators cannot see into their participant's actual environment, and whether these measures impact collaboration outcomes.

Together, these ethical implications of adapting embodied actions across distributed environments highlight the responsibility researchers and designers share in building collaboration spaces. While Harrison & Dourish (1996) originally emphasised that people transform spaces into meaningful places through interaction, Dourish (2006) later emphasised the designer's role in creating the affordances that make such place-making possible. Take for instance the 'raise hand' function common in video conferencing technologies. In co-located meetings, the idea of raising one's hand to speak feels child-like and inappropriate, perhaps because this behaviour is reserved for classroom culture (Lawson 2007). However, people readily use this interaction in video conferencing, perhaps because it has been made so readily available by the interface design. This shift demonstrates how designed affordances can give people the means to establish new etiquette in digital spaces, changing what is considered socially acceptable. Similarly, in blended realities, the affordances available for adapting embodied actions will likely influence emerging cultures and social norms in these environments. Designers have a responsibility to critically reflect on the interactions they enable, as these can quietly codify the norms of future collaborative environments. Future work should continue to test adapting embodied actions across diverse contexts, ensuring the design of these spaces are well-informed.

6.2 Strategies for Navigating Heterogeneous Contexts

The Spatial Heterogeneity Framework and extended Expectancy Violations Theory (EVT) (Burgoon 1993) aims to help researchers discuss and compare solutions for blending heterogeneous contexts.

This thesis presents a snapshot of these frameworks in action. However, I note that heterogeneity between contexts evolves as the technology, collaborative activity or physical space changes.

Not all heterogeneity challenges need to be resolved with technology. people might employ different strategies to change their environment and overcome barriers caused by their heterogeneous contexts. Recognising the dynamic nature of collaborative work (Gutwin & Greenberg 2002), there are a range of ways people might opt to control their environment (**Figure 6.1**). These can be grouped into the following categories, where collaborators might choose to:

- **Appropriate the available technology** and use a different solution or re-calibrate how their spaces are blended. For example, Grønbæk et al.'s (2023) *RealityBlender* assumes a non-linear discontinuous level of heterogeneity across the blended zones. Instead of trying to blend the whole active zone, they instead reduce the blended zones to smaller areas at a rigid level of heterogeneity. By splicing up the spaces, they aim to reduce the degree of heterogeneity and therefore gain a number of blended proxemics for free within the blended zones. Other technical solutions could offer different ways to overcome this level of heterogeneity. For example, Suma et al. (2012) use non-linear mapping to remap a smaller physical space to a larger one. Although their system focuses on virtual reality in a single space, it is possible to imagine how this technical solution could be applied across distributed locations.

- **Adapt their collaborative activity** restricting it to more compatible blended zones or compromising the range of supported proxemic cues. For example, instead of remapping the space, Yang et al.'s (2024) system prompts the user to adapt where they stand in order to enable collaboration across the distributed locations. This effectively restricts the active zone by asking people to adapt their behaviour.

- **Rearrange their physical spaces** so that the spatial layouts are more similar across the locations. Although originally created to *find* the optimal free space between distributed spaces, Lehment et al. (2014) and Yang et al.'s (2024) systems could also be used to show how physical objects in the space can be moved by participants to *create* an optimal collaboration space.

Researchers should be aware of the different choices people have when modifying their environment. By understanding these strategies, it is possible to explore a wider range of solutions that extend beyond simply layering additional blending techniques into a system. Such an approach may help reduce perceptual distortions, which can arise when multiple blending techniques conflict or reach their practical limits. In some cases, collaborators may benefit more from adjusting their physical environment or collaborative activity rather than relying solely on system adaptations. However, these alternative modification strategies place the burden for modification on people, requiring careful consideration of the costs and benefits involved. Future work should therefore examine how people can collaborate *with* blended realities systems to co-construct

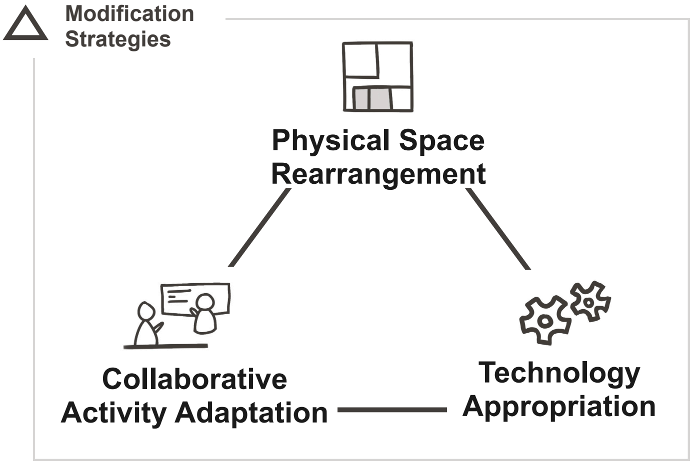


Fig. 6.1 Different modification strategies can be used to overcome heterogeneity caused by spatial differences. These include changing the technology used, the physical space layout and collaborative activity requirements. Future work might take these different strategies into consideration, to support a more comprehensive perspective of how people interact with their environment.

places for distributed interaction, and identify when human versus system adaptation best supports the collaboration context.

6.3 Limitations and Future Work

The Spatial Heterogeneity Framework presented in this thesis was generated and evaluated based on current state of the art. These blended realities solutions are not exhaustive, however, the framework does not aim to categorise every possible system perfectly. Instead, it provides a unifying language to discuss and compare approaches. As the field evolves, new forms of blended proxemics or blending techniques will emerge to address spatial heterogeneity, and certain asymmetries (e.g. groups working co-located with headsets alongside a remote participant on a mobile phone) may require the framework to be extended. Nevertheless, advancing blended proxemics requires a starting point. The Spatial Heterogeneity Framework establishes this foundation by offering a unifying language to analyse differences between spaces, the types of actions being blended, and how these are supported by the spatial layouts or technical solution. This shared language enables researchers to compare existing systems and deliberately design new approaches that extend beyond the current perspectives.

A key challenge of this research is that fully functioning blended realities systems are rare and time intensive to develop. This makes it difficult to rapidly test a wide range of interaction possibilities on the fly. I therefore employed low-fidelity visualisations and scenario-based simulations in Chapter 5 to probe participants' preferences. This speculative approach offered flexibility and efficiency, producing valuable insights with minimal resources. However, the lack of a real immersive system limits the findings, and they should not be overgeneralised. Instead, the study identifies key factors that may influence how participants prioritise spatial information or sociocultural norms in different collaborative contexts. Future research might build on these factors, using these qualitative findings as a foundation for more rigorous system design and hypothesis development than would be possible without prior data.

The open-ended, participant-led nature of the semi-structured interviews was another trade-off. While it sacrificed experimental control and did not generate effect sizes, it surfaced what participants considered significant in the moment, providing rich qualitative insights for theory-building. A natural next step is to test the generalisability of these themes through larger-scale quantitative studies, enabling stronger claims about how people navigate tensions between spatial information and sociocultural norms in blended realities.

Furthermore, the study presented in Chapter 5 examines sociocultural norms, yet it was not designed to compare these across cultural contexts. The expectations participants voiced, around personal space, hierarchy and behavioural appropriateness, are likely to vary significantly across geographic and cultural boundaries (Burgoon, Floyd & Guerrero 2016, Kendon 2004). Comparative cross-cultural studies could show that blended realities need different adaptations to respect and accommodate diverse sociocultural norms.

Finally, the findings from the user study suggest that over time, people may become accustomed to certain glitches in blended environments. While this acceptance might reduce friction, it raises concerns, especially if persistent misalignments begin to obscure intent or normalise problematic behaviours. This is particularly salient in translation mode, where social signals are transformed and may no longer accurately reflect people's actions. Although not the central focus of this study, I highlight the need for future research to explore how designers might balance seamless interaction with transparency, particularly in situations where masking intent could cause harm.

Chapter 7

Conclusion

In this thesis I examine how blended realities overcome heterogeneous contexts to coherently transmit embodied actions from one location to another. In this chapter I revisit the aims and research questions, synthesising the final insights in a concise and structured manner.

7.1 Synthesised Contributions

This thesis aims to *investigate how embodied actions are transmitted across heterogeneous contexts in blended realities and the design trade-offs this involves.*

To achieve this aim, I developed a taxonomy that identifies three modes that blended realities used to transmit embodied actions across distributed spaces: (1) *transcription*, where actions are directly sent from one space to another without any adaptations; (2) *transformation*, where the system constructs models of each space and remaps actions to overcome spatial heterogeneity; and (3) *translation*, where additional embodiment rules are deliberately encoded to uphold sociocultural norms.

Using this taxonomy as a foundation for my work, I show how spatial heterogeneity is a key challenge for coherently transmitting embodied actions across distribute locations. I show that if the blended zones are dissimilar, then *transcribing* embodied actions without any adaptation distorts or changes the meaning of these actions when perceived in the remote space. Instead, actions need to be *transformed* from one space to the other, to preserve spatially dependent information.

While there has been some blended realities systems and blending techniques developed to overcome spatial heterogeneity, these lack a unified language and theoretical construct to discuss, compare and predict when these solutions work best. In this thesis I present the Spatial Heterogeneity Framework to clearly define the challenge of spatial dissimilarities and explain the assumptions blended realities systems make when they transmit interactions from one space to another using different blending techniques. This framework demonstrates how systems that assume a higher degree of spatial heterogeneity between the blended zones require more blending techniques since less embodied actions are enabled “for free” by the spatial layout. It also sets out strategies for supporting the transmission of embodied actions by reducing the degree of heterogeneity between the blended zones.

As outlined in the taxonomy for transmitting embodied actions across distributed spaces, blending heterogeneous contexts involves more than preserving spatial information. It requires a holistic approach to context, recognising both Dey’s definition of context as an external configuration of entities and their relationships (Dey 2001), and Dourish’s account of context as socially enacted and culturally situated (Dourish 2004). These epistemological views of context require researchers to carefully consider how adaptations across distributed spaces transform actions to preserve spatial accuracy and translate them to uphold sociocultural norms.

Adaptations that transform or translate embodied interactions can come into conflict. Particularly in highly heterogeneous contexts where multiple blending techniques are needed to sustain blended proxemics. Such conflicts introduce design trade-offs where blending techniques that preserve spatial accuracy risk violating social norms, and techniques that uphold social norms can distort spatial meaning. The user study in Chapter 5 showed that participants evaluate these trade-offs situationally, sometimes prioritising spatial information, other times sociocultural norms, depending on the collaborative context. Participants’ awareness of the mediating system, combined with concern that their actions might be perceived differently, raises ethical questions about when and how actions should be adapted.

7.1.1 Research Questions

(RQ1) How do current canonical blended realities systems overcome spatial heterogeneity challenges?

By analysing 14 canonical blended realities systems, I demonstrate how each addresses spatial heterogeneity by categorising blended zones into six levels of difference. These levels, ordered by increasing heterogeneity, are: causal where the spaces are inherently connected spaces (Sutherland et al. 1965); to rigid spaces that require a rotational transformation (Pejsa et al. 2016, Orts-Escolano et al. 2016); affine differences that require scaling or mirroring (Grønbæk et al. 2024, Herskovitz et al. 2022); non-linear continuous differences that require point-to-point warping (Congdon et al. 2018, Yoon et al. 2021), non-linear discontinuous spaces that can be spliced into partially mapped zones (Fink et al. 2022, Wang et al. 2022, Wang, Kim, Panda, Ofek, Franco & Won 2024); and categorical spaces that are so different that only individual objects can be blended and not the space between them (Yang et al. 2024, Johnson et al. 2021).

Across these systems, blended proxemics, such as manipulating physical objects, maintaining one-to-one scale, spatial deictic cues, congruent body position and trajectory, consistent vicinity, and enabling verbal referencing, are either afforded directly by the physical layout or achieved through blending techniques, such as deixis warping (Yoon et al. 2022), teleportation (Wang, Kim, Panda, Ofek, Franco & Won 2024), transition animations (Grønbæk et al. 2017), or redirected walking (Wang et al. 2022).

(RQ2) How might current approaches for overcoming spatial heterogeneity be characterised in a theoretical framework?

In Chapter 4, I presented the Spatial Heterogeneity Framework. This was developed iteratively based on 14 canonical examples and validated against 32 additional blended realities systems identified in a literature review. The framework unifies approaches to overcoming spatial heterogeneity and consists of four components: (1) *activity zones*, which define the functional roles of different areas in each collaborator's physical environment, including the "blended zones" that are mapped to each other; (2) the *heterogeneity ladder*, which characterises how similar or dissimilar these blended zones are; (3) *blended proxemics*, which uses the degree of heterogeneity to determine how proxemics (e.g., deictic cues) are preserved "for free" by the spatial layout or enabled through a blending technique; and (4) the *solutions matrix*, which compares blended realities solutions and shows whether they afford proxemics through blending techniques or assume they are available "for free" via the spatial layout.

The framework demonstrates that as systems assume greater heterogeneity between spaces, they require more blending techniques to coherently transmit spatially dependent actions across contexts. These techniques make interaction possible across highly heterogeneous spaces but also increase the risk of perceptual distortions, since each technique loosely couples actions from one spatial context to another.

(RQ3) How do people weigh the trade-offs between upholding spatial information and sociocultural norms during remote collaboration?

In Chapter 5, I demonstrated how people weigh up the trade-offs between upholding spatial information and sociocultural norms in a user study with 20 participants. Using counterfactual cards, I probed how factors in different collaborative contexts influenced whether participants chose to preserve spatial information (transform) or uphold sociocultural norms (translate). My findings showed that these decisions closely align with Burgoon's expectancy violations theory (EVT). Participants first drew on communicator characteristics such as their partner's communication style and perceived competence with technology, which shaped their baseline expectations. They also considered relational characteristics, including personal stakes, hierarchy, and power dynamics: for example, participants were less forgiving of proxemic violations when the interaction involved a supervisor than when it involved a peer. Finally, context characteristics such as time pressure or public visibility played a critical role: in urgent scenarios, participants often prioritised efficiency and spatial information, while in formal settings they preferred translation to avoid breaching sociocultural norms.

Consistent with EVT, participants then compared actual behaviours against these expectancies, evaluating the violation valence, whether a breach was acceptable, unacceptable, or even beneficial. Crucially, they also assessed the communicator reward valence, judging the violation positively or negatively based on their impression of their collaborator. Here, awareness of system mediation softened negative judgments: participants frequently attributed violations to technical glitches, which led them to excuse behaviours that might otherwise have been seen as inappropriate. However, this tolerance also introduced ambiguity, as participants were uncertain whether they should attribute the perceived proxemics to their partner or the system.

Taken together, these findings show that participants evaluated trade-offs between spatial information and sociocultural norms through the same expectancy–violation–reward logic that EVT describes, although their final judgement is influenced by their awareness of the mediating system. This suggests EVT can be extended to account for how people flexibly recalibrate expectancies in light of technical ambiguity, and how they selectively uphold either transformation or translation rules depending on communicator, relational, and contextual factors.

(RQ4) What should designers and developers consider when incorporating transformation and translation modes into blended reality systems?

In Chapter 6 I discuss how it is important for blended realities systems to clarify who or what performs the adaptation. When the system automatically transforms or translates actions, participants may struggle to distinguish between system-driven recalibration and their partner's intent. This ambiguity can obscure accountability, sometimes leading participants to excuse norm violations as "glitches" but at other times undermining trust or masking harmful behaviour. Designers must consider how to balance automation with transparency, perhaps by selectively signalling whether an interaction stems from human intent or machine mediation without overwhelming users with constant cues.

In addition, ethical concerns arise when users cannot perceive how they are represented in another's space. Transformation and translation may preserve spatial coherence or sociocultural norms locally but they also risk misrepresenting intent remotely, raising issues of deception. Designers and developers must remain aware of this risk and perform rigorous research to responsibly design these collaboration spaces.

Finally, there are other options for adapting embodied actions that do not require blending techniques. Blended realities should flexibly help people to decide when to auto-adapt embodied actions and when to employ alternative strategies, such as re-zoning spaces, rearranging layouts, or adjusting the activity requirements. This might mean designing and developing systems where people actively collaborate with the blended realities system to adapt actions across heterogeneous contexts, researching the costs and benefits this has for collaborative outcomes.

7.2 Final Remarks

Blended realities systems offer new opportunities for people to work across distributed spaces, while still accessing the tangible affordances of their physical environment. These systems create a sense of co-location even when participants are geographically distributed.

However, when collaborators' environments differ, blended realities systems must transform the spatial information or translate social norms to maintain coherence across *heterogeneous contexts*. Importantly, this definition of heterogeneity across distributed collaborative environments takes into account both the spatial mapping of entities (Dey 2001), as well as the sociocultural elements of place-making (Dourish & Bellotti 1992, Harrison & Dourish 1996). As demonstrated in this thesis, these different modes of blending embodied interactions can at times conflict, generating "social noise" that disrupts the illusion of co-location. When evaluating such trade-offs, collaborators implicitly weigh environmental cues and interpersonal expectations in line with Expectancy Violations Theory (Burgoon 1993), while also taking into account their awareness of the mediating system. Similar to Putnam's (1975) "twin world" thought experiment, loose coupling between objects and actions across sites can complicate communication, leaving participants uncertain whether a breakdown originates from their partner's intent or the system's intervention. Designers of blended reality systems have a responsibility to consider what these "twin worlds" mean for cognition and transparency of intent.

Returning to Hollan and Stornetta's (1992) *Beyond Being There*, blended realities' goal to recreate co-located embodied interactions calls into question whether it is preferable to remain situated in one's own physical space and bring collaborators into it, or to "go beyond" into a shared virtual environment that sacrifices tangible affordances for more consistent virtual interactions. These ideas need not be two different approaches. Instead future research should seek to understand when and to what degree these outlooks should be applied, depending on the collaborative context. For example, as the heterogeneity between distributed environments increases, perhaps it is preferable to go more "beyond".

As blended realities systems continue to develop, it will be crucial to compare how these strategies for being situated in the external world, and going beyond, support cognition and social coordination in different contexts. This thesis presents foundational theoretical ideas for this inquiry, problematising how blended realities should strategically blend spaces to preserve physical affordances, when embodied actions might be adapted, as well as who or what adapts these. By situating these design challenges in theories of externalism (Clark & Chalmers 1998), contextual awareness (Dey 2001, Dourish 2004), embodied interactions (Dourish 2001), and place-making (Harrison & Dourish 1996), I make a case for blended realities to enable the best of being *here and* beyond across heterogeneous contexts.

References

Abdullah, A., Kolkmeier, J., Lo, V. & Neff, M. (2021), 'Videoconference and Embodied VR: Communication Patterns Across Task and Medium', *Proceedings of the ACM on Human-Computer Interaction* **5**(CSCW2), 1–29. Publisher: Association for Computing Machinery (ACM).
URL: *https://dl.acm.org/doi/10.1145/3479597*

Abramczuk, K., Bohdanowicz, Z., Muczyński, B., Skorupska, K. H. & Cnotkowski, D. (2023), 'Meet me in vr! can vr space help remote teams connect: a seven-week study with horizon workrooms', *International Journal of Human-Computer Studies* **179**, 103104.

Akselrad, D., DeVeaux, C., Han, E., Miller, M. R. & Bailenson, J. N. (2023), Body Crumple, Sound Intrusion, and Embodiment Violation: Toward a Framework for Miscommunication in VR, *in* 'Computer Supported Cooperative Work and Social Computing', ACM, Minneapolis MN USA, pp. 122–125.
URL: *https://dl.acm.org/doi/10.1145/3584931.3606968*

Auda, J., Busse, L., Pfeuffer, K., Gruenefeld, U., Rivu, R., Alt, F. & Schneegass, S. (2021), I'm in control! transferring object ownership between remote users with haptic props in virtual reality, *in* 'Proceedings of the 2021 ACM Symposium on Spatial User Interaction', SUI '21, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3485279.3485287*

Bailenson, J. N. (2021), 'Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue', *Technology, Mind, and Behavior* **2**(1). https://tmb.apaopen.org/pub/nonverbal-overload.

Bailenson, J. N., Blascovich, J., Beall, A. C. & Loomis, J. M. (2001), 'Equilibrium theory revisited: Mutual gaze and personal space in virtual environments', *Presence: Teleoperators & Virtual Environments* **10**(6), 583–598.

Baldis, J. J. (2001), Effects of spatial audio on memory, comprehension, and preference during desktop conferences, *in* 'Proceedings of the SIGCHI conference on Human factors in computing systems', pp. 166–173.

Barden, P., Comber, R., Green, D., Jackson, D., Ladha, C., Bartindale, T., Bryan-Kinns, N., Stockman, T. & Olivier, P. (2012), Telematic dinner party: designing for togetherness through play and performance, *in* 'Proceedings of the Designing Interactive Systems Conference', DIS '12, Association for Computing Machinery, New York, NY, USA, p. 38–47.
URL: *https://doi.org/10.1145/2317956.2317964*

Barua, H. B., Mg, T. H., Pramanick, P. & Sarkar, C. (2024), 'Enabling social robots to perceive and join socially interacting groups using f-formation: A comprehensive overview', *J. Hum.-Robot Interact.* **13**(4).
URL: *https://doi.org/10.1145/3682072*

Benford, S., Greenhalgh, C., Reynard, G., Brown, C. & Koleva, B. (1998), 'Understanding and constructing shared spaces with mixed-reality boundaries', *ACM Transactions on computer-human interaction (TOCHI)* **5**(3), 185–223.

Benford, S., Greenhalgh, C., Rodden, T. & Pycock, J. (2001), 'Collaborative virtual environments', *Communications of the ACM* **44**(7), 79–85.

Bente, G., Rüggenberg, S., Krämer, N. C. & Eschenburg, F. (2008), 'Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations', *Human Communication Research* **34**(2), 287–318.
URL: *https://academic.oup.com/hcr/article/34/2/287-318/4210796*

Billinghurst, M. & Kato, H. (1999), Real world teleconferencing, *in* 'CHI '99 Extended Abstracts on Human Factors in Computing Systems', CHI EA '99, Association for Computing Machinery, New York, NY, USA, p. 194–195.
URL: *https://doi.org/10.1145/632716.632838*

Blackwell, L., Ellison, N., Elliott-Deflo, N. & Schwartz, R. (2019), 'Harassment in social virtual reality: Challenges for platform governance', *Proc. ACM Hum.-Comput. Interact.* **3**(CSCW).
URL: *https://doi.org/10.1145/3359202*

Braun, V. & Clarke, V. (2006), 'Using thematic analysis in psychology', *Qualitative research in psychology* **3**(2), 77–101.

Brave, S., Ishii, H. & Dahley, A. (1998), Tangible interfaces for remote collaboration and communication, *in* 'Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work', CSCW '98, Association for Computing Machinery, New York, NY, USA, p. 169–178.
URL: *https://doi.org/10.1145/289444.289491*

Brown, B., MacColl, I., Chalmers, M., Galani, A., Randell, C. & Steed, A. (2003), Lessons from the lighthouse: collaboration in a shared mixed reality system, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', CHI '03, Association for Computing Machinery, New York, NY, USA, p. 577–584.
URL: *https://doi.org/10.1145/642611.642711*

Burgoon, J. & Dunbar, N. (2006), *Nonverbal Expressions of Dominance and Power in Human Relationships*, pp. 279–297.

Burgoon, J. K. (1993), 'Interpersonal expectations, expectancy violations, and emotional communication', *Journal of language and social psychology* **12**(1-2), 30–48.

Burgoon, J. K., Bonito, J. A., Lowry, P. B., Humpherys, S. L., Moody, G. D., Gaskin, J. E. & Giboney, J. S. (2016), 'Application of Expectancy Violations Theory to communication with and judgments about embodied agents during a decision-making task', *International Journal of Human-Computer Studies* **91**, 24–36.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1071581916000227*

Burgoon, J. K., Floyd, K. & Guerrero, L. (2016), *Nonverbal communication*, Routledge, Taylor & Francis Group, London New York.

Buxton, B. (2009), *Mediaspace – Meaningspace – Meetingspace*, Springer London, London, UK.

Buxton, W. (2006), 'Space-function integration and ubiquitous media', *Towards an Interactive and Integrative Design Process. Linz, Austria: DOM Publications* pp. 248–271.

Casanueva, J. & Blake, E. (2000), The effects of group collaboration on presence in a collaborative virtual environment, *in* 'Virtual Environments 2000: Proceedings of the Eurographics Workshop in Amsterdam, The Netherlands, June 1–2, 2000', Springer, pp. 85–94.

Clark, A. (2010), *Minds in Space*, Oxford University Press, United Kingdom, pp. 7–15.

Clark, A. & Chalmers, D. (1998), 'The extended mind', *analysis* **58**(1), 7–19.

Cochet, H. & Vauclair, J. (2010), 'Pointing gesture in young children: Hand preference and language development', *Gesture* **10**(2-3), 129–149.

Cone, J. & Rand, D. G. (2014), 'Time pressure increases cooperation in competitively framed social dilemmas', *PLoS one* **9**(12), e115756.

Congdon, B. J., Wang, T. & Steed, A. (2018), Merging environments for shared spaces in mixed reality, *in* 'Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology', VRST '18, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3281505.3281544*

Dey, A. K. (2001), 'Understanding and using context', *Personal and ubiquitous computing* **5**(1), 4–7.

Dourish, P. (2001), *Where the Action Is: The Foundations of Embodied Interaction*, The MIT Press.
URL: *https://doi.org/10.7551/mitpress/7221.001.0001*

Dourish, P. (2004), 'What we talk about when we talk about context', **8**(1), 19–30.
URL: *https://doi.org/10.1007/s00779-003-0253-8*

Dourish, P. (2006), Re-space-ing place: "place" and "space" ten years on, *in* 'Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work', CSCW '06, Association for Computing Machinery, New York, NY, USA, p. 299–308.
URL: *https://doi.org/10.1145/1180875.1180921*

Dourish, P. & Bellotti, V. (1992), Awareness and coordination in shared workspaces, *in* 'Proceedings of the 1992 ACM conference on Computer-supported cooperative work', pp. 107–114.

Ens, B., Lanir, J., Tang, A., Bateman, S., Lee, G., Piumsomboon, T. & Billinghurst, M. (2019), 'Revisiting collaboration through mixed reality: The evolution of groupware', *International Journal of Human-Computer Studies* **131**, 81–98.

Fincher, S. & Tenenberg, J. (2005), 'Making sense of card sorting data', *Expert Systems* **22**(3), 89–93.

Fink, D. I., Zagermann, J., Reiterer, H. & Jetter, H.-C. (2022), 'Re-locations: Augmenting personal and shared workspaces to support remote collaboration in incongruent spaces', *Proc. ACM Hum.-Comput. Interact.* **6**(ISS).
URL: *https://doi.org/10.1145/3567709*

Follmer, S., Leithinger, D., Olwal, A., Hogge, A. & Ishii, H. (2013), inform: dynamic physical affordances and constraints through shape and object actuation, *in* 'Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology', UIST '13, Association for Computing Machinery, New York, NY, USA, p. 417–426.
URL: *https://doi.org/10.1145/2501988.2502032*

Freiwald, J. P., Schenke, J., Lehmann-Willenbrock, N. & Steinicke, F. (2021), Effects of avatar appearance and locomotion on co-presence in virtual reality collaborations, *in* 'Proceedings of Mensch Und Computer 2021', MuC '21, Association for Computing Machinery, New York, NY, USA, p. 393–401.
URL: *https://doi.org/10.1145/3473856.3473870*

Fussell, S. R., Kraut, R. E. & Siegel, J. (2000), Coordination of communication: effects of shared visual context on collaborative work, *in* 'Proceedings of the 2000 ACM conference on Computer supported cooperative work', CSCW '00, Association for Computing Machinery, New York, NY, USA, pp. 21–30.
URL: *https://dl.acm.org/doi/10.1145/358916.358947*

Gaver, W. W., Sellen, A., Heath, C. & Luff, P. (1993), One is not enough: Multiple views in a media space, *in* 'Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems', pp. 335–341.

Gerhard, M., Moore, D. & Hobbs, D. (2004), 'Embodiment and copresence in collaborative interfaces', *International Journal of Human-Computer Studies* **61**(4), 453–480.

Goffman, E. (2008), *Behavior in public places*, Simon and Schuster.

Goodwin, C. (2000), 'Action and embodiment within situated human interaction', *Journal of pragmatics* **32**(10), 1489–1522.

Gorski, P. S. (2013), 'What is critical realism? and why should you care?'.

Gower, J., Baumeister, J., Knight, M., Thomas, B. & Zucco, J. (2022), Augmented reality support for performance and decision-making in collaborative time-critical tasks, *in* '2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)', IEEE, pp. 958–963.

Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R. & Wang, M. (2011), 'Proxemic interactions: the new ubicomp?', *Interactions* **18**(1), 42–50.
URL: *https://doi.org/10.1145/1897239.1897250*

Grønbæk, J. E., Knudsen, M. S., O'Hara, K., Krogh, P. G., Vermeulen, J. & Petersen, M. G. (2020), Proxemics beyond proximity: Designing for flexible social interaction through cross-device interaction, *in* 'Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems', CHI '20, Association for Computing Machinery, New York, NY, USA, p. 1–14.
URL: *https://doi.org/10.1145/3313831.3376379*

Grønbæk, J. E., Korsgaard, H., Petersen, M. G., Birk, M. H. & Krogh, P. G. (2017), Proxemic Transitions: Designing Shape-Changing Furniture for Informal Meetings, *in* 'Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems', CHI '17, Association for Computing Machinery, New York, NY, USA, pp. 7029–7041.
URL: *https://doi.org/10.1145/3025453.3025487*

Grønbæk, J. E. S., Pfeuffer, K., Velloso, E., Astrup, M., Pedersen, M. I. S., Kjær, M., Leiva, G. & Gellersen, H. (2023), Partially blended realities: Aligning dissimilar spaces for distributed mixed reality meetings, *in* 'Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems', CHI '23, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3544548.3581515*

Grønbæk, J. E. S., Sánchez Esquivel, J., Leiva, G., Velloso, E., Gellersen, H. & Pfeuffer, K. (2024), Blended whiteboard: Physicality and reconfigurability in remote mixed reality collaboration, *in* 'Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems', CHI '24, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3613904.3642293*

Gutwin, C. & Greenberg, S. (2002), 'A descriptive framework of workspace awareness for real-time groupware', *Computer Supported Cooperative Work (CSCW)* **11**, 411–446.

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

Hall, J. A. & Knapp, M. L., eds (2013), *Nonverbal communication*, number 2 *in* 'Handbooks of communication science', De Gruyter Mouton, Berlin ; Boston.

Hanks, W. F. (2009), 'Fieldwork on deixis', *Journal of Pragmatics* **41**(1), 10–24. Towards an Emancipatory Pragmatics.
URL: *https://www.sciencedirect.com/science/article/pii/S0378216608001975*

Harrison, S. & Dourish, P. (1996), Re-place-ing space: the roles of place and space in collaborative systems, *in* 'Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work', CSCW '96, Association for Computing Machinery, New York, NY, USA, p. 67–76.
URL: *https://doi.org/10.1145/240080.240193*

He, Z., Du, R. & Perlin, K. (2020), Collabovr: A reconfigurable framework for creative collaboration in virtual reality, *in* '2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)', IEEE, New York, NY, USA, pp. 542–554.

Heath, C. & Luff, P. (1993), 'Disembodied conduct: Interactional asymmetries in video-mediated communication', *Technology in working order: Studies of work, interaction, and technology* pp. 35–54.

Heath, C. & Luff, P. (2000), *Technology in action*, Cambridge university press.

Hedayati, H., Szafir, D. & Kennedy, J. (2020), Comparing f-formations between humans and on-screen agents, *in* 'Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems', CHI EA '20, Association for Computing Machinery, New York, NY, USA, p. 1–9.
URL: *https://doi.org/10.1145/3334480.3383015*

Herskovitz, J., Cheng, Y. F., Guo, A., Sample, A. P. & Nebeling, M. (2022), 'Xspace: An augmented reality toolkit for enabling spatially-aware distributed collaboration', *Proc. ACM Hum.-Comput. Interact.* **6**(ISS).
URL: *https://doi.org/10.1145/3567721*

Higuchi, K., Chen, Y., Chou, P. A., Zhang, Z. & Liu, Z. (2015), Immerseboard: Immersive telepresence experience using a digital whiteboard, *in* 'Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems', CHI '15, Association for Computing Machinery, New York, NY, USA, p. 2383–2392.
URL: *https://doi.org/10.1145/2702123.2702160*

Hollan, J. & Stornetta, S. (1992), Beyond being there, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', CHI '92, Association for Computing Machinery, New York, NY, USA, p. 119–125.
URL: *https://doi.org/10.1145/142750.142769*

Hoppe, A. H., van de Camp, F. & Stiefelhagen, R. (2021), 'Shisha: Enabling shared perspective with face-to-face collaboration using redirected avatars in virtual reality', *Proc. ACM Hum.-Comput. Interact.* **4**(CSCW3).
URL: *https://doi.org/10.1145/3432950*

Huang, X. & Xiao, R. (2024), 'Surfshare: Lightweight spatially consistent physical surface and virtual replica sharing with head-mounted mixed-reality', *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **7**(4).
URL: *https://doi.org/10.1145/3631418*

Irlitti, A., Latifoglu, M., Zhou, Q., Reinoso, M. N., Hoang, T., Velloso, E. & Vetere, F. (2023), Volumetric mixed reality telepresence for real-time cross modality collaboration, *in* 'Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems', CHI '23, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3544548.3581277*

Irlitti, A., Smith, R. T., Von Itzstein, S., Billinghurst, M. & Thomas, B. H. (2016), Challenges for asynchronous collaboration in augmented reality, *in* '2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)', IEEE, New York, NY, USA, pp. 31–35.

Ishii, H. & Kobayashi, M. (1992), Clearboard: A seamless medium for shared drawing and conversation with eye contact, *in* 'Proceedings of the SIGCHI conference on Human factors in computing systems', pp. 525–532.

Jo, D., Kim, K.-H. & Kim, G. J. (2015), 'Spacetime: adaptive control of the teleported avatar for improved ar tele-conference experience', *Computer Animation and Virtual Worlds* **26**(3-4), 259–269.

Johnson, J. G., Gasques, D., Sharkey, T., Schmitz, E. & Weibel, N. (2021), Do you really need to know where "that" is? enhancing support for referencing in collaborative mixed reality environments, *in* 'Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems', CHI '21, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3411764.3445246*

Johnson, J. G., Sharkey, T., Butarbutar, I. C., Xiong, D., Huang, R., Sy, L. & Weibel, N. (2023), Unmapped: Leveraging experts' situated experiences to ease remote guidance in collaborative mixed reality, *in* 'Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems', CHI '23, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3544548.3581444*

Johnson-Laird, P. N. (1983), *Mental models: Towards a cognitive science of language, inference, and consciousness*, Harvard University Press, Cambridge, Massachusetts.

Junuzovic, S., Inkpen, K., Blank, T. & Gupta, A. (2012), Illumishare: sharing any surface, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', CHI '12, Association for Computing Machinery, New York, NY, USA, p. 1919–1928.
URL: *https://doi.org/10.1145/2207676.2208333*

Kang, J., Yang, D., Kim, T., Lee, Y. & Lee, S.-H. (2023), Real-time retargeting of deictic motion to virtual avatars for augmented reality telepresence, *in* '2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)', pp. 885–893.

Kendon, A. (1990), *Conducting interaction: Patterns of behavior in focused encounters*, Vol. 7, CUP Archive, Cambridge, UK.

Kendon, A. (2004), *Gesture: Visible action as utterance*, Cambridge University Press.

Kendon, A. (2010), 'Spacing and orientation in co-present interaction', *Development of Multimodal Interfaces: Active Listening and Synchrony: Second COST 2102 International Training School, Dublin, Ireland, March 23-27, 2009, Revised Selected Papers* pp. 1–15.

Kim, D., Kim, H.-I. & Woo, W. (2022), Mutual space generation with relative translation gains in redirected walking for asymmetric remote collaboration, *in* '2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)', pp. 674–677.

Kim, D., Kim, S., Choi, S. & Woo, W. (2024), Spatial affordance-aware interactable subspace allocation for mixed reality telepresence, *in* '2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)', pp. 1256–1265.

Kim, S., Kim, D., Shin, J.-E. & Woo, W. (2024), Object cluster registration of dissimilar rooms using geometric spatial affordance graph to generate shared virtual spaces, *in* '2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)', pp. 796–805.

Kim, S., Lee, G., Sakata, N. & Billinghurst, M. (2014), Improving co-presence with augmented visual communication cues for sharing experience through video conference, *in* '2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)', pp. 83–92.

Kiyokawa, K., Billinghurst, M., Hayes, S., Gupta, A., Sannohe, Y. & Kato, H. (2002), Communication behaviors of co-located users in collaborative ar interfaces, *in* 'Proceedings. International Symposium on Mixed and Augmented Reality', Institute of Electrical and Electronic Engineers, New York, NY, USA, pp. 139–148.

Klemmer, S. R., Hartmann, B. & Takayama, L. (2006), How bodies matter: five themes for interaction design, *in* 'Proceedings of the 6th conference on Designing Interactive systems', ACM, pp. 140–149.
URL: *https://dl.acm.org/doi/10.1145/1142405.1142429*

Lawson, B. (2007), *Language of space*, Architectural Press, Woburn, MA, USA.

Lehment, N. H., Merget, D. & Rigoll, G. (2014), Creating automatically aligned consensus realities for ar videoconferencing, *in* '2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)', IEEE, New York, NY, USA, pp. 201–206.

Levinson, S. C. (2006), 'Deixis', *The handbook of pragmatics* pp. 97–121.

Llobera, J., Spanlang, B., Ruffini, G. & Slater, M. (2010), 'Proxemics with multiple dynamic characters in an immersive virtual environment', *ACM Transactions on Applied Perception* **8**(1), 1–12. Publisher: Association for Computing Machinery (ACM).
URL: *https://dl.acm.org/doi/10.1145/1857893.1857896*

Luff, P. & Heath, C. (1998), Mobility in collaboration, *in* 'Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work', CSCW '98, Association for Computing Machinery, New York, NY, USA, pp. 305–314.
URL: *https://doi.org/10.1145/289444.289505*

Mackay, W. E. & McGrenere, J. (2025), 'Comparative structured observation', *ACM Transactions on Computer-Human Interaction* **32**(2), 1–27.

Maloney, D., Freeman, G. & Wohn, D. Y. (2020), '" talking without a voice" understanding non-verbal communication in social virtual reality', *Proceedings of the ACM on Human-Computer Interaction* **4**(CSCW2), 1–25.

Marquardt, N., Hinckley, K. & Greenberg, S. (2012), Cross-device interaction via micro-mobility and f-formations, *in* 'Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology', UIST '12, Association for Computing Machinery, New York, NY, USA, p. 13–22.
URL: *https://doi.org/10.1145/2380116.2380121*

Marshall, P. & Necker, E. H. (2013), Theories of Embodiment in HCI, *in* 'The SAGE Handbook of Digital Technology Research', SAGE Publications Ltd, pp. 144–158.
URL: *https://methods.sagepub.com/book/the-sage-handbook-of-digital-technology-research/n11.xml*

McGill, M., Boland, D., Murray-Smith, R. & Brewster, S. (2015), A dose of reality: Overcoming usability challenges in vr head-mounted displays, *in* 'Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems', CHI '15, Association for Computing Machinery, New York, NY, USA, p. 2143–2152.
URL: *https://doi.org/10.1145/2702123.2702382*

McVeigh-Schultz, J. & Isbister, K. (2022), 'A "beyond being there" for vr meetings: envisioning the future of remote work', *Human–Computer Interaction* **37**(5), 433–453.
URL: *https://doi.org/10.1080/07370024.2021.1994860*

Mentis, H. M., O'Hara, K., Sellen, A. & Trivedi, R. (2012), Interaction proxemics and image use in neurosurgery, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', CHI '12, Association for Computing Machinery, New York, NY, USA, p. 927–936.
URL: *https://doi.org/10.1145/2207676.2208536*

Milgram, P. & Kishino, F. (1994), 'A taxonomy of mixed reality visual displays', *IEICE TRANSACTIONS on Information and Systems* **77**(12), 1321–1329.

Newen, A., Gallagher, S. & De Bruin, L. (2018), 4e cognition: Historical roots, key concepts, and central issues, *in* A. Newen, L. De Bruin & S. Gallagher, eds, 'The Oxford Handbook of 4E Cognition', 1 edn, Oxford University Press, pp. 3–16.
URL: *https://academic.oup.com/edited-volume/28083/chapter/212118698*

Norman, D. A. (2014), Some observations on mental models, *in* 'Mental models', Psychology Press, New York, NY, USA, pp. 15–22.

Oh Kruzic, C., Kruzic, D., Herrera, F. & Bailenson, J. (2020), 'Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments', *Scientific reports* **10**(1), 20626.

O'Hara, K., Kjeldskov, J. & Paay, J. (2011), 'Blended interaction spaces for distributed team collaboration', *ACM transactions on computer-human interaction* **18**, 1–28.

Onishi, Y., Takashima, K., Higashiyama, S., Fujita, K. & Kitamura, Y. (2022), Waddlewalls: Room-scale interactive partitioning system using a swarm of robotic partitions, *in* 'Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology', UIST '22, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3526113.3545615*

Oprean, D., Simpson, M. & Klippel, A. (2018), 'Collaborating remotely: An evaluation of immersive capabilities on spatial experiences and team membership', *International journal of digital earth* **11**(4), 420–436.

Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P. L., Khamis, S., Dou, M., Tankovich, V., Loop, C., Cai, Q., Chou, P. A., Mennicken, S., Valentin, J., Pradeep, V., Wang, S., Kang, S. B., Kohli, P., Lutchyn, Y., Keskin, C. & Izadi, S. (2016), Holoportation: Virtual 3d teleportation in real-time, *in* 'Proceedings of the 29th Annual Symposium on User Interface Software and Technology', UIST '16, Association for Computing Machinery, New York, NY, USA, p. 741–754.
URL: *https://doi.org/10.1145/2984511.2984517*

Oulasvirta, A. & Hornbæk, K. (2016), Hci research as problem-solving, *in* 'Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems', CHI '16, Association for Computing Machinery, New York, NY, USA, p. 4956–4967.
URL: *https://doi.org/10.1145/2858036.2858283*

Oulasvirta, A. & Hornbæk, K. (2022), ‘Counterfactual thinking: What theories do in design’, *International Journal of Human–Computer Interaction* **38**(1), 78–92.

Pechmann, T. & Deutsch, W. (1982), ‘The development of verbal and nonverbal devices for reference’, *Journal of experimental child psychology* **34**(2), 330–341.

Pejsa, T., Kantor, J., Benko, H., Ofek, E. & Wilson, A. (2016), Room2room: Enabling life-size telepresence in a projected augmented reality environment, *in* ‘Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing’, CSCW ’16, Association for Computing Machinery, New York, NY, USA, p. 1716–1725.
URL: *https://doi.org/10.1145/2818048.2819965*

Pooladvand, S., Hasanzadeh, S., Takahashi, G., Park, K. J. & Marroquin, J. (2025), ‘Simulating social pressure: Evaluating risk behaviors in construction using augmented virtuality’, *IEEE Transactions on Visualization and Computer Graphics* .

Poretski, L., Lanir, J. & Arazy, O. (2018), ‘Normative tensions in shared augmented reality’, *Proc. ACM Hum.-Comput. Interact.* **2**(CSCW).
URL: *https://doi.org/10.1145/3274411*

Putnam, H. (1975), ‘The meaning of" meaning"’.

Riedl, R., Mohr, P. N., Kenning, P. H., Davis, F. D. & Heekeren, H. R. (2014), ‘Trusting humans and avatars: A brain imaging study based on evolution theory’, *Journal of Management Information Systems* **30**(4), 83–114.

Robertson, T. (1997), Cooperative Work and Lived Cognition: A Taxonomy of Embodied Actions, *in* ‘Proceedings of the Fifth European Conference on Computer Supported Cooperative Work’, Springer Netherlands, Dordrecht, pp. 205–220. Noted, maybe read.
URL: *http://link.springer.com/10.1007/978-94-015-7372-6*$_1$4

Rogers, Y. (2004), ‘New theoretical approaches for hci’, *Annual review of information science and technology* **38**(1), 87–143.

Sakashita, M., Zhang, R., Li, X., Kim, H., Russo, M., Zhang, C., Jung, M. F. & Guimbretière, F. (2023), Remotion: Supporting remote collaboration in open space with automatic robotic embodiment, *in* ‘Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems’, CHI ’23, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3544548.3580699*

Sanaei, M., Machacek, M., Gilbert, S. B., Wu, P. & Oliver, J. (2023), Comparing perceptions of performance across virtual reality, video conferencing, and face-to-face collaborations, *in* ‘2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)’, pp. 4556–4561.

Schäfer, A., Reis, G. & Stricker, D. (2022), ‘A survey on synchronous augmented, virtual, andmixed reality remote collaboration systems’, *ACM Comput. Surv.* **55**(6).
URL: *https://doi.org/10.1145/3533376*

Sereno, M., Wang, X., Besançon, L., McGuffin, M. J. & Isenberg, T. (2022), ‘Collaborative work in augmented reality: A survey’, *IEEE Transactions on Visualization and Computer Graphics* **28**(6), 2530–2549.

Setti, F., Russell, C., Bassetti, C. & Cristani, M. (2015), ‘F-formation detection: Individuating free-standing conversational groups in images’, *PloS one* **10**(5), e0123783.

Shannon, C. E. (1948), 'A mathematical theory of communication', *The Bell system technical journal* **27**(3), 379–423.

Simeone, A. L., Velloso, E. & Gellersen, H. (2015), Substitutional reality: Using the physical environment to design virtual reality experiences, *in* 'Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems', CHI '15, Association for Computing Machinery, New York, NY, USA, p. 3307–3316.
URL: *https://doi.org/10.1145/2702123.2702389*

Skarbez, R., Smith, M. & Whitton, M. C. (2021), 'Revisiting milgram and kishino's reality-virtuality continuum', *Frontiers in Virtual Reality* **2**.
URL: *https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2021.647997*

Slater, M., Gonzalez-Liencres, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., Watson, Z., Breen, G., Schwarz, R., Steptoe, W., Szostak, D., Halan, S., Fox, D. & Silver, J. (2020), 'The Ethics of Realism in Virtual and Augmented Reality', *Frontiers in Virtual Reality* **1**. Publisher: Frontiers.
URL: *https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2020.00001/full*

Smith, H. J. & Neff, M. (2018), Communication behavior in embodied virtual reality, *in* 'Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems', CHI '18, Association for Computing Machinery, New York, NY, USA, p. 1–12.
URL: *https://doi.org/10.1145/3173574.3173863*

Sousa, M., dos Anjos, R. K., Mendes, D., Billinghurst, M. & Jorge, J. (2019), Warping deixis: Distorting gestures to enhance collaboration, *in* 'Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems', CHI '19, Association for Computing Machinery, New York, NY, USA, p. 1–12.
URL: *https://doi.org/10.1145/3290605.3300838*

Speicher, M., Hall, B. D. & Nebeling, M. (2019), What is mixed reality?, *in* 'Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems', CHI '19, Association for Computing Machinery, New York, NY, USA, p. 1–15.
URL: *https://doi.org/10.1145/3290605.3300767*

Sra, M., Mottelson, A. & Maes, P. (2018), Your place and mine: Designing a shared vr experience for remotely located users, *in* 'Proceedings of the 2018 Designing Interactive Systems Conference', DIS '18, Association for Computing Machinery, New York, NY, USA, p. 85–97.
URL: *https://doi.org/10.1145/3196709.3196788*

Steinicke, F., Lehmann-Willenbrock, N. & Meinecke, A. L. (2020), A first pilot study to compare virtual group meetings using video conferences and (immersive) virtual reality, *in* 'Proceedings of the 2020 ACM Symposium on Spatial User Interaction', SUI '20, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3385959.3422699*

Suchman, L. A. (1987), *Plans and situated actions: The problem of human-machine communication*, Cambridge university press.

Suma, E. A., Lipps, Z., Finkelstein, S., Krum, D. M. & Bolas, M. (2012), 'Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture', *IEEE Transactions on Visualization and Computer Graphics* **18**(4), 555–564.

Sutherland, I. E. et al. (1965), The ultimate display, *in* 'Proceedings of the IFIP Congress', Vol. 2, New York, pp. 506–508.

Suzuki, R., Hedayati, H., Zheng, C., Bohn, J. L., Szafir, D., Do, E. Y.-L., Gross, M. D. & Leithinger, D. (2020), Roomshift: Room-scale dynamic haptics for vr with furniture-moving swarm robots, *in* 'Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems', CHI '20, Association for Computing Machinery, New York, NY, USA, p. 1–11.
URL: *https://doi.org/10.1145/3313831.3376523*

Thoravi Kumaravel, B., Anderson, F., Fitzmaurice, G., Hartmann, B. & Grossman, T. (2019), Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence, *in* 'Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology', UIST '19, Association for Computing Machinery, New York, NY, USA, p. 161–174.
URL: *https://doi.org/10.1145/3332165.3347872*

Tian, Y., Liu, S. & Wang, J. (2023), 'A corpus study on the difference of turn-taking in online audio, online video, and face-to-face conversation', *Language and Speech* **67**(3), 00238309231176768.

Tong, L., Serna, A., Pageaud, S., George, S. & Tabard, A. (2016), It's not how you stand, it's how you move: F-formations and collaboration dynamics in a mobile learning game, *in* 'Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services', MobileHCI '16, Association for Computing Machinery, New York, NY, USA, p. 318–329.
URL: *https://doi.org/10.1145/2935334.2935343*

Tran, T. Q., Langlotz, T. & Regenbrecht, H. (2024), A survey on measuring presence in mixed reality, *in* 'Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems', CHI '24, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3613904.3642383*

Tsui, K. M., Desai, M., Yanco, H. A. & Uhlik, C. (2011), Exploring use cases for telepresence robots, *in* 'Proceedings of the 6th international conference on Human-robot interaction', pp. 11–18.

Velloso, E. & Hornbæk, K. (2025), Theorising in hci using causal models, *in* 'Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems', CHI '25, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3706598.3713789*

Virilio, P. (1989), 'The museum of accidents', *public* .

Wang, C. Y., Kim, H., Panda, P., Ofek, E., Franco, M. G. & Won, A. S. (2024), Mrtransformer: Transforming avatar non-verbal behavior for remote mr collaboration in incongruent spaces, *in* '2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)', pp. 545–548.

Wang, P., Miller, M. R., Queiroz, A. C. & Bailenson, J. N. (2024), Socially Late, Virtually Present: The Effects of Transforming Asynchronous Social Interactions in Virtual Reality, *in* 'Proceedings of the CHI Conference on Human Factors in Computing Systems', ACM, Honolulu HI USA, pp. 1–19.
URL: *https://dl.acm.org/doi/10.1145/3613904.3642244*

Wang, X., Ye, H., Sandor, C., Zhang, W. & Fu, H. (2022), 'Predict-and-drive: Avatar motion adaption in room-scale augmented reality telepresence with heterogeneous spaces', *IEEE Transactions on Visualization and Computer Graphics* **28**(11), 1–10.

Webb, S. (2001), 'Avatarculture: Narrative, power and identity in virtual world environments', *Information, Communication & Society* **4**(4), 560–594.

White, C. H. (2008), Expectancy violations theory and interation adaption theory, *in* 'Engaging theories in interpersonal communication : multiple perspectives', Sage Publications.

Williamson, J., Li, J., Vinayagamoorthy, V., Shamma, D. A. & Cesar, P. (2021), Proxemics and social interactions in an instrumented virtual reality workshop, *in* 'Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems', CHI '21, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3411764.3445729*

Williamson, J. R., O'Hagan, J., Guerra-Gomez, J. A., Williamson, J. H., Cesar, P. & Shamma, D. A. (2022), Digital proxemics: Designing social and collaborative interaction in virtual environments, *in* 'Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems', CHI '22, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3491102.3517594*

Wimmer, H. & Perner, J. (1983), 'Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception', *Cognition* **13**(1), 103–128.

Wong, E., Genay, A., Grønbæk, J. E. S. & Velloso, E. (2025), Spatial heterogeneity in distributed mixed reality collaboration, *in* 'Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems', CHI '25, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3706598.3714033*

Wong, E., Grønbæk, J. E. S. & Velloso, E. (2024), The jamais vu effect: Understanding the fragile illusion of co-presence in mixed reality, *in* 'Proceedings of the 2024 ACM Designing Interactive Systems Conference', DIS '24, Association for Computing Machinery, New York, NY, USA, p. 2227–2246.
URL: *https://doi.org/10.1145/3643834.3661574*

Wong, E., Sánchez Esquivel, J., Leiva, G., Grønbæk, J. E. S. & Velloso, E. (2024), Practice-informed patterns for organising large groups in distributed mixed reality collaboration, *in* 'Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems', CHI '24, Association for Computing Machinery, New York, NY, USA.
URL: *https://doi.org/10.1145/3613904.3642502*

Wong, N. & Gutwin, C. (2010), Where are you pointing?: the accuracy of deictic pointing in CVEs, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', ACM, Atlanta Georgia USA, pp. 1029–1038.
URL: *https://dl.acm.org/doi/10.1145/1753326.1753480*

Wu, C. M., Schulz, E., Pleskac, T. J. & Speekenbrink, M. (2022), 'Time pressure changes how people explore and respond to uncertainty', *Scientific reports* **12**(1), 4122.

Yang, D., Kang, J., Kim, T. & Lee, S.-H. (2024), 'Visual guidance for user placement in avatar-mediated telepresence between dissimilar spaces', *IEEE Transactions on Visualization and Computer Graphics* **30**(12), 7558–7570.

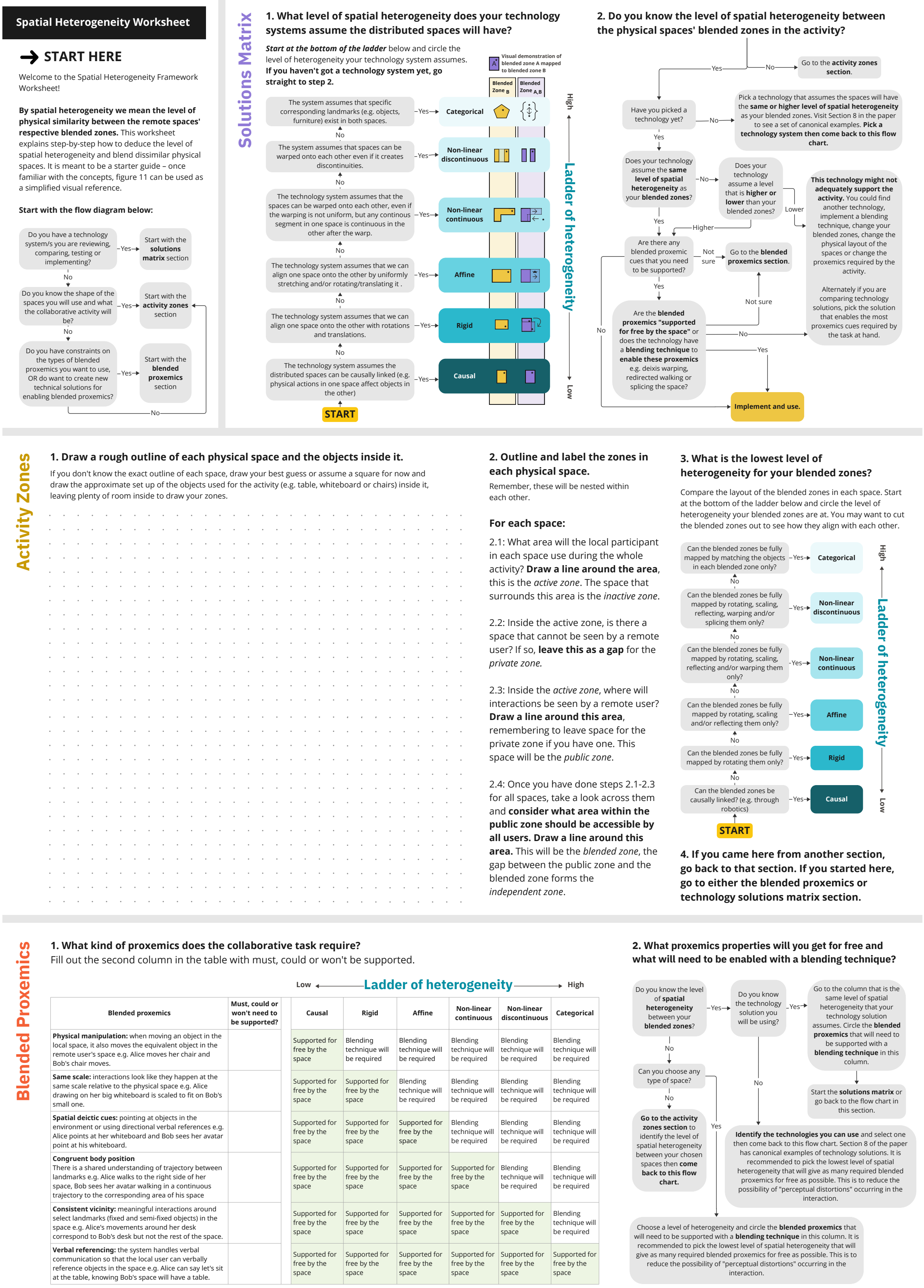
Yoon, L., Yang, D., Chung, C. & Lee, S.-H. (2021), 'A full body avatar-based telepresence system for dissimilar spaces'.

Yoon, L., Yang, D., Kim, J., Chung, C. & Lee, S.-H. (2022), 'Placement retargeting of virtual avatars to dissimilar indoor environments', *IEEE Transactions on Visualization and Computer Graphics* **28**(3), 1619–1633.

Zubek, J., Nagórska, E., Komorowska-Mach, J., Skowrońska, K., Zieliński, K. & Rączaszek-Leonardi, J. (2022), 'Dynamics of remote communication: Movement coordination in video-mediated and face-to-face conversations', *Entropy* **24**(4), 559.

Appendix A

Spatial Heterogeneity Framework Worksheet



Spatial Heterogeneity Worksheet

START HERE

Welcome to the Spatial Heterogeneity Framework Worksheet!

By spatial heterogeneity we mean the level of physical similarity between the remote spaces' respective blended zones. This worksheet explains step-by-step how to deduce the level of spatial heterogeneity and blend dissimilar physical spaces. It is meant to be a starter guide – once familiar with the concepts, figure 11 can be used as a simplified visual reference.

Start with the flow diagram below:

Do you have a technology system/s you are reviewing, comparing, testing or implementing? –Yes→ Start with the **solutions matrix** section

No

Do you know the shape of the spaces you will use and what the collaborative activity will be? –Yes→ Start with the **activity zones** section

No

Do you have constraints on the types of blended proxemics you want to use, OR do want to create new technical solutions for enabling blended proxemics? –Yes→ Start with the **blended proxemics** section

No

Solutions Matrix

1. What level of spatial heterogeneity does your technology systems assume the distributed spaces will have?

Start at the bottom of the ladder below and circle the level of heterogeneity your technology system assumes. **If you haven't got a technology system yet, go straight to step 2.**

Visual demonstration of blended zone A mapped to blended zone B

Blended Zone B

Blended Zone A,B

The system assumes that specific corresponding landmarks (e.g. objects, furniture) exist in both spaces. –Yes→ **Categorical**

No

The system assumes that spaces can be warped onto each other even if it creates discontinuities. –Yes→ **Non-linear discontinuous**

No

The technology system assumes that the spaces can be warped onto each other, even if the warping is not uniform, but any continous segment in one space is continuous in the other after the warp. –Yes→ **Non-linear continuous**

No

The technology system assumes that we can align one space onto the other by uniformly stretching and/or rotating/translating it . –Yes→ **Affine**

No

The technology system assumes that we can align one space onto the other with rotations and translations. –Yes→ **Rigid**

No

The technology system assumes the distributed spaces can be causally linked (e.g. physical actions in one space affect objects in the other) –Yes→ **Causal**

START

High ← **Ladder of heterogeneity** → Low

2. Do you know the level of spatial heterogeneity between the physical spaces' blended zones in the activity?

No → Go to the **activity zones section**.

Yes → Have you picked a technology yet?

No → Pick a technology that assumes the spaces will have the **same or higher level of spatial heterogeneity** as your blended zones. Visit Section 8 in the paper to see a set of canonical examples. **Pick a technology system then come back to this flow chart.**

Yes → Does your technology assume the **same level of spatial heterogeneity** as your **blended zones**?

No → Does your technology assume a level that is **higher or lower** than your blended zones?

Lower → **This technology might not adequately support the activity.** You could find another technology, implement a blending technique, change your blended zones, change the physical layout of the spaces or change the proxemics required by the activity.

Alternately if you are comparing technology solutions, pick the solution that enables the most proxemics cues required by the task at hand.

Higher / Yes → Are there any blended proxemic cues that you need to be supported?

Not sure → Go to the **blended proxemics section**.

No → **Implement and use.**

Yes → Are the **blended proxemics "supported for free by the space"** or does the technology have a **blending technique** to **enable these proxemics** e.g. deixis warping, redirected walking or splicing the space?

Not sure → Go to the **blended proxemics section**.

No → **This technology might not adequately support the activity.**

Yes → **Implement and use.**

Activity Zones

1. Draw a rough outline of each physical space and the objects inside it.

If you don't know the exact outline of each space, draw your best guess or assume a square for now and draw the approximate set up of the objects used for the activity (e.g. table, whiteboard or chairs) inside it, leaving plenty of room inside to draw your zones.

2. Outline and label the zones in each physical space.

Remember, these will be nested within each other.

For each space:

2.1: What area will the local participant in each space use during the whole activity? **Draw a line around the area**, this is the *active zone*. The space that surrounds this area is the *inactive zone*.

2.2: Inside the active zone, is there a space that cannot be seen by a remote user? If so, **leave this as a gap** for the *private zone.*

2.3: Inside the *active zone*, where will interactions be seen by a remote user? **Draw a line around this area**, remembering to leave space for the private zone if you have one. This space will be the *public zone*.

2.4: Once you have done steps 2.1-2.3 for all spaces, take a look across them and **consider what area within the public zone should be accessible by all users. Draw a line around this area.** This will be the *blended zone*, the gap between the public zone and the blended zone forms the *independent zone*.

3. What is the lowest level of heterogeneity for your blended zones?

Compare the layout of the blended zones in each space. Start at the bottom of the ladder below and circle the level of heterogeneity your blended zones are at. You may want to cut the blended zones out to see how they align with each other.

Can the blended zones be fully mapped by matching the objects in each blended zone only? –Yes→ **Categorical**

No

Can the blended zones be fully mapped by rotating, scaling, reflecting, warping and/or splicing them only? –Yes→ **Non-linear discontinuous**

No

Can the blended zones be fully mapped by rotating, scaling, reflecting and/or warping them only? –Yes→ **Non-linear continuous**

No

Can the blended zones be fully mapped by rotating, scaling and/or reflecting them only? –Yes→ **Affine**

No

Can the blended zones be fully mapped by rotating them only? –Yes→ **Rigid**

No

Can the blended zones be causally linked? (e.g. through robotics) –Yes→ **Causal**

START

High ← **Ladder of heterogeneity** → Low

4. If you came here from another section, go back to that section. If you started here, go to either the blended proxemics or technology solutions matrix section.

Blended Proxemics

1. What kind of proxemics does the collaborative task require?

Fill out the second column in the table with must, could or won't be supported.

Low ← **Ladder of heterogeneity** → High

Blended proxemics	Must, could or won't need to be supported?	Causal	Rigid	Affine	Non-linear continuous	Non-linear discontinuous	Categorical
Physical manipulation: when moving an object in the local space, it also moves the equivalent object in the remote user's space e.g. Alice moves her chair and Bob's chair moves.		Supported for free by the space	Blending technique will be required	Blending technique will be required	Blending technique will be required	Blending technique will be required	Blending technique will be required
Same scale: interactions look like they happen at the same scale relative to the physical space e.g. Alice drawing on her big whiteboard is scaled to fit on Bob's small one.		Supported for free by the space	Supported for free by the space	Blending technique will be required	Blending technique will be required	Blending technique will be required	Blending technique will be required
Spatial deictic cues: pointing at objects in the environment or using directional verbal references e.g. Alice points at her whiteboard and Bob sees her avatar point at his whiteboard.		Supported for free by the space	Supported for free by the space	Supported for free by the space	Blending technique will be required	Blending technique will be required	Blending technique will be required
Congruent body position There is a shared understanding of trajectory between landmarks e.g. Alice walks to the right side of her space, Bob sees her avatar walking in a continuous trajectory to the corresponding area of his space		Supported for free by the space	Supported for free by the space	Supported for free by the space	Supported for free by the space	Blending technique will be required	Blending technique will be required
Consistent vicinity: meaningful interactions around select landmarks (fixed and semi-fixed objects) in the space e.g. Alice's movements around her desk correspond to Bob's desk but not the rest of the space.		Supported for free by the space	Supported for free by the space	Supported for free by the space	Supported for free by the space	Supported for free by the space	Blending technique will be required
Verbal referencing: the system handles verbal communication so that the local user can verbally reference objects in the space e.g. Alice can say let's sit at the table, knowing Bob's space will have a table.		Supported for free by the space	Supported for free by the space	Supported for free by the space	Supported for free by the space	Supported for free by the space	Supported for free by the space

2. What proxemics properties will you get for free and what will need to be enabled with a blending technique?

Do you know the level of **spatial heterogeneity** between your **blended zones**? –Yes→ Do you know the technology solution you will be using? –Yes→ Go to the column that is the same level of spatial heterogeneity that your technology solution assumes. Circle the **blended proxemics** that will need to be supported with a **blending technique** in this column. → Start the **solutions matrix** or go back to the flow chart in this section.

Do you know the technology solution you will be using? –No→ **Identify the technologies you can use** and select one then come back to this flow chart. Section 8 of the paper has canonical examples of technology solutions. It is recommended to pick the lowest level of spatial heterogeneity that will give as many required blended proxemics for free as possible. This is to reduce the possibility of "perceptual distortions" occurring in the interaction.

Do you know the level of spatial heterogeneity between your blended zones? –No→ Can you choose any type of space?

No → **Go to the activity zones section** to identify the level of spatial heterogeneity between your chosen spaces then **come back to this flow chart.**

Yes → Choose a level of heterogeneity and circle the **blended proxemics** that will need to be supported with a **blending technique** in this column. It is recommended to pick the lowest level of spatial heterogeneity that will give as many required blended proxemics for free as possible. This is to reduce the possibility of "perceptual distortions" occurring in the interaction.

Fig. A.1 This worksheet explains step-by-step how to identify the level of spatial heterogeneity and blend dissimilar physical spaces. It is meant to be a starter guide – Figure 4.12 can be used as a simplified visual reference.

Appendix B

Counterfactual Card Study Materials

B.0.1 Semi-structured Interview Questions

1. <*interviewer places scenario (critical/assessment/game) on the table and explains it*>
2. Do you have any questions about this scenario?
3. <*interviewer places interaction option A (transform) in condition 1 on the table*> How do you feel about this interaction?
    (a) Why?
    (b) You said <*quote*> could you explain that a bit more?
4. Coming back to the scenario, where you are collaborating with (your NASA colleague/the HR rep/your new boss), how do you think this interaction option would affect the scenario?
    (a) Why?
    (b) You said <*quote*> could you explain that a bit more?
5. <*interviewer places interaction option B (translate) in condition 1 on the table*> How do you feel about this interaction?
    (a) Why?
    (b) You said <*quote*> could you explain that a bit more?
6. Given <*scenario*> which option would you choose, interaction option A or option B? Walk me through your decision process.
7. If you had to use both of these options, are there any work around you would use to overcome their challenges?
8. <*Interviewer repeats 1-7 for each scenario, probing further where necessary*>

9. What do you think the difference is between these scenarios?
10. *<interviewer places interaction option A (transform) in condition <2/3/4/5/6> on the table>* How do you feel about this interaction?
    (a) Why?
    (b) You said *<quote>* could you explain that a bit more?
11. How do you think this interaction option would affect each scenario?
    (a) Why?
    (b) You said *<quote>* could you explain that a bit more?
12. *<interviewer places interaction option B (translate) in condition <2/3/4/5/6> on the table>* How do you feel about this interaction?
    (a) Why?
    (b) You said *<quote>* could you explain that a bit more?
13. For each scenario, which option would you choose, option A or option B? Walk me through your decision process.
14. If you had to use both of these options, are there any work around you would use to overcome their challenges?
15. *<Interviewer repeats 10-14 for each condition, probing further where necessary>*
16. Before we finish, is there anything else you can think of that would impact your decisions?

B.0.2 Participant Backgrounds

ID	Age bracket	Gender	Education	Vocation	Extended reality technologies used	Frequency
P1	26-30	Woman	Bachelor of Fine Arts, Masters of IT	Developer	Augmented reality (AR) on my phone or tablet, Mixed reality (MR) headset, Virtual reality (VR) headset	Rarely (a few times a year)
P2	26-30	Man	Bachelor of Music	Music Teacher	MR headset	Rarely (a few times a year)
P3	36-40	Man	Master of Architecture, PhD Candidate	Architect and PhD Student	AR on my phone or tablet, MR headset, VR headset	Rarely (a few times a year)
P4	26-30	Man	Bachelor of Arts, Masters of Teaching	Middle-school Teacher	AR on my phone or tablet, VR headset	Rarely (a few times a year)
P5	41-45	Man	Bachelor of Science, Graduate Diploma in International Community Development	Customer Experience Researcher	AR on my phone or tablet, VR headset	Rarely (a few times a year)
P6	18-25	Woman	Bachelor of Commerce	Undergraduate Student and Financial Services Firm Intern	AR on my phone or tablet, VR headset	Rarely (a few times a year)
P7	18-25	Woman	Bachelor of Psychology	Undergraduate Student, Hospitality Team Leader	AR on my phone or tablet, VR headset	Rarely (a few times a year)
P8	18-25	Man	Bachelor of Engineering, Master of Management	Finance	AR on my phone or tablet, VR headset	Rarely (a few times a year)
P9	18-25	Woman	Bachelor of Music, Diploma of Computer Science	Undergraduate Student, Hospitality	AR headset, MR headset, VR headset	Very Often (a few times a week)
P10	18-25	Woman	Bachelor of Biomedical Science	Undergraduate Student	None	Never
P11	36-40	Woman	Bachelor of Arts, Masters of Philosophy in Classics	Masters Research Student	VR headset	Rarely (a few times a year)
P12	18-25	Woman	Bachelor of Psychological Science, Master of Policy	Masters Student, Retail	AR on my phone or tablet, AR headset, VR headset	Rarely (a few times a year)
P13	18-25	Woman	Bachelor of Science	Undergraduate Student, Retail	AR on my phone or tablet	Rarely (a few times a year)
P14	18-25	Woman	Bachelor of Psychology	Undergraduate Student	VR headset	Rarely (a few times a year)
P15	18-25	Man	Bachelor of Science	Research Scientist	AR on my phone or tablet, VR headset	Rarely (a few times a year)
P16	18-25	Woman	Bachelor of Psychology	Undergraduate Student	AR on my phone or tablet	Rarely (a few times a year)
P17	18-25	Man	Bachelor of Science	Undergraduate Student	AR on my phone or tablet, VR headset	Never
P18	18-25	Man	Bachelor of Science	Undergraduate Student	AR on my phone or tablet, VR headset	Rarely (a few times a year)
P19	18-25	Man	Bachelor of Science, Master of Environmental Engineering	Building Services Engineer	AR on my phone or tablet, MR headset	Rarely (a few times a year)
P20	26-30	Man	Bachelor of Arts, Juris Doctorate	Lawyer	AR headset	Rarely (a few times a year)

Fig. B.1 Participant backgrounds including age, gender, education, vocation and experience using extended reality technologies.