Spatial Reference Frame Integration Performance in Collaborative Tasks

Authors: Qi Pei, Li Jing, Li Jing

Date: 2025-11-29T00:00:00+00:00

Abstract

This study conducted three experiments to investigate the patterns of individuals' spatial reference frame representation in spatial collaborative tasks and the conditions that facilitate reference frame integration. In each experiment, participants first performed collaborative learning of spatial scenes, then individually completed relative position judgment tasks. In Experiment 1, the angular separation between participants and their collaborative partners during the learning phase was 45°, whereas in Experiment 2 it was 135°. In Experiment 3, the angular separation was identical to that in Experiment 2, but the partner departed prematurely midway through the task. The results revealed: (1) Multiple forms of reference frame representation exist in spatial collaborative tasks, with both single and multiple reference frame representations being possible; (2) Reference frame integration in spatial collaboration is modulated by the partner' s position; compared to standing side-by-side, an approximately face-to-face angle with the partner better facilitates participants' completion of reference frame integration; (3) When the collaborative partner' s presence time is reduced, participants can still complete reference frame integration; however, compared to situations where the partner remains present throughout, the partner' s departure after the reconstruction phase prompted participants to form a more profound representation of the partner' s perspective.

Full Text

Integration of Spatial Reference Frames in Collaborative Tasks

QI Pei, LI Jing
(School of Psychology, Nanjing Normal University, Nanjing 210097, China)

Abstract

This study investigated the patterns of individual spatial reference frame representation in spatial collaborative tasks and the conditions that facilitate reference frame integration through three experiments. In each experiment, participants first engaged in collaborative learning of a spatial scene, then completed relative position judgment tasks individually. In Experiment 1, the angle between participant and collaborator positions was 45°; in Experiment 2, this angle increased to 135°; and in Experiment 3, the angle remained the same as in Experiment 2, but the partner left midway. Results revealed: (1) Multiple forms of reference frame representation emerged in spatial collaborative tasks, with both single and multiple reference frame representations possible; (2) Reference frame integration in spatial collaboration was modulated by the collaborator' s position, with an approximately face-to-face angle being more conducive to integration than a side-by-side arrangement; (3) When the collaborator' s presence time was reduced, participants could still complete reference frame integration, but the partner' s departure after the reconstruction phase prompted participants to form a more profound representation of the partner' s perspective compared to when the partner remained present throughout.

Keywords: spatial frame of reference, spatial perspective-taking, spatial collaboration, reference frame integration
Classification Code: B842

---

1. Introduction

1.1 Spatial Frame of Reference

In daily life, to better navigate and orient ourselves in space, people often need to establish a spatial coordinate system in their minds—known as a spatial frame of reference—to represent objects and others' locations in the environment (Shelton & McNamara, 2001). Since this concept was proposed, it has attracted extensive research from scholars. These studies can be broadly categorized into two directions: one focuses on the dominance of egocentric versus allocentric reference frames in cognition (Klatzky, 1998; Mou et al., 2008; Xie et al., 2009), while the other examines how individuals select and construct spatial reference frames under different conditions (Hatzipanayioti & Avraamides, 2021). The former has enriched the theoretical framework of spatial reference frames and laid a solid foundation for building comprehensive spatial reference models; the latter has deepened our understanding of how various factors shape spatial reference frame representation, providing scientific basis and effective guidance for practical applications such as voice navigation and human-computer interaction. These influencing conditions can be divided into subject-related and object-related factors. Among them, individuals' spatial reference preferences and scene-related experiences (e.g., the first perspective used when observing a scene) significantly impact reference frame establishment (Christou & Bülthoff, 1999; Mou et al., 2007; Simons & Wang, 1998; Valiquette et al., 2003). Additionally, external factors such as scene intrinsic axes, object orientation, and other people in the scene also affect individuals' spatial reference frame construction (Freundlieb et al., 2017; Li & Zhang, 2011; Zhao & Mou, 2005). However, most of these studies focused on single participants learning a scene individually in static conditions without others present. In real life, many of our behaviors are dynamic and require interaction with others, sometimes even relying on collaboration to complete tasks. Therefore, understanding how individuals establish spatial reference frames when collaborating with others is crucial for successful task completion.

Schafer and Bowman (2004) have confirmed that reference frame selection affects spatial collaboration efficiency, and this influence is closely related to task requirements. Currently, research on the mechanisms underlying individual spatial reference frame establishment in collaborative tasks remains limited. Furthermore, which factors in collaborative tasks can promote individuals to represent others' perspectives and integrate their reference frames into their own spatial reference systems constitutes the core content this study aims to explore further.

When others are present in a spatial scene, participants' spatial reference frame representations are also affected, though certain conditions must be met. Freundlieb et al. (2017) noted that spontaneous perspective-taking (imagining oneself standing in another' s perspective to view the spatial scene) only occurs when participants perceive that the other' s gaze is also directed at the scene. Xie (2020) also found that when others in the learning scene were merely present, participants could not represent their perspective; however, when others simply described scene objects verbally (naming objects) without any other communication, participants showed response advantages to the other' s perspective in some scene tasks (compared to other perspectives). These findings collectively reveal the prerequisites and complexity of how others influence individual spatial representation. The underlying mechanism of this influence falls within the research domain of spatial collaboration.

1.2 Reference Frame Establishment in Spatial Collaboration

Living in complex and diverse spatial environments, individuals frequently need to engage in spatial interactions with others. Spatial interaction refers to a dynamic process where individuals understand, transmit, and share spatial information based on specific goals, involving multiple spatial cognitive mechanisms such as spatial perspective-taking and spatial reference frames (Xiao et al., 2019; Xiao et al., 2021). In spatial interactions, representing and understanding the other' s perspective through spatial perspective-taking is particularly crucial for smooth interaction. As a core form of spatial interaction, efficient spatial collaboration (where both parties cooperate to complete a spatial task) similarly relies on spatial perspective-taking. Spatial perspective-taking refers to an individual' s ability to flexibly shift and consider others' perspectives under specific conditions. In spatial collaborative tasks, by adopting the partner' s perspective, individuals may construct corresponding spatial reference frames based on that perspective, thereby providing cognitive support for successful task completion. This study aims to systematically verify this hypothesis and explore its underlying mechanisms.

As previously discussed, the "other" in the scene is an important influencing factor for both spatial reference frame establishment and spatial perspective-taking processes. However, spatial collaborative tasks are more complex as they involve not only individual factors but also multiple aspects of bilateral interaction. Therefore, when exploring individuals' spatial reference direction selection in collaborative tasks, more diverse influencing factors must be considered. First, spatial collaborative tasks have clear goal orientation, requiring both parties to work together to achieve specific objectives. In this context, both parties' spatial abilities become important factors influencing participants' spatial reference direction selection. Schober' s (2009) early research using spatial language tasks found that when participants knew their partner had lower spatial abilities, they were more inclined to describe from the partner' s perspective even when describing from their own perspective would be more convenient. Second, researchers have also found that the partner' s performance in the task influences participants' reference perspective selection. For example, Furlanetto et al. (2013) used short videos in spatial perspective-taking tasks and found that others' gaze toward objects in the video and their action tendencies could prompt participants to adopt the other' s perspective for spatial judgment, and this facilitation effect remained significant even when participants could only see the other' s action tendencies (with the other' s gaze blurred in the video). Additionally, Freundlieb et al. (2016) used joint spatial tasks to explore the conditions for participants' spontaneous perspective-taking of their partner, noting that participants only spontaneously adopted the other' s perspective when they fully believed the partner was actively participating in the spatial task. Besides partner characteristics, task instructions and other task-related cues also influence participants' spatial reference direction selection. Multiple studies have shown that if participants know beforehand that they will collaborate with someone and must describe space to a partner with a different perspective, they are more inclined to describe from the other' s perspective or represent space accordingly (Galati et al., 2013; Kelly et al., 2018; Shelton & McNamara, 2004).

Some researchers have simultaneously examined social and physical factors in spatial interaction tasks (Galati & Avraamides, 2015). They combined social cues in spatial tasks with scene characteristics to explore how these two types of cues interactively influence individuals' spatial representation and language description. In their experiment, participants were paired, with one serving as the instruction giver responsible for scene learning, description, and subsequent relative direction judgment tasks, while the paired partner served as the instruction receiver, responsible for scene reconstruction while listening to descriptions. The between-group factors included consistency between both parties' perspectives and scene intrinsic axes (scene intrinsic axis aligned with instruction giver' s perspective/receiver' s perspective/neither) and whether the instruction giver knew the partner' s location (know/don' t know). Results showed that when the partner' s location was known, participants tended to use the scene intrinsic axis direction as their spatial reference frame representation direction regardless of whether their own or the partner' s perspective aligned with the scene intrinsic axis. When participants didn' t know the partner' s location, they tended to represent space from their own observation perspective. Additionally, participants showed flexibility in spatial description tasks: they described more from their own perspective when the scene intrinsic axis aligned with themselves, but could adopt the partner' s perspective when it aligned with the partner. These findings further reveal that factors influencing participants' spatial reference direction selection are more diverse in complex spatial interaction tasks. When multiple reference cues coexist, participants make choices after weighing them. However, how various factors specifically influence individuals' spatial reference direction selection in complex spatial interaction tasks requires extensive follow-up research.

1.3 Research Questions and Hypotheses

In summary, researchers have made many explorations into spatial reference frames, spatial perspective-taking, and spatial interaction (Xiao et al., 2019), but existing research has two main limitations. First, previous studies lacked sufficient social context and collaborative depth. For example, in Xie' s (2020) study, participants either didn' t communicate or, although they communicated, their learning performance wasn' t affected by collaborative outcomes, resulting in insufficient collaboration. In Galati and Avraamides' (2015) experiment, participants passively received information from their collaborative partner, leaving room for improvement in interactivity. Therefore, building on previous research and focusing on human interaction characteristics, this study upgraded the single-participant individual scene learning scenario to a situation where participants collaborate with a partner to complete tasks, aiming to explore spatial reference frame establishment in richer collaborative contexts.

Second, in spatial collaborative tasks, participants have more diverse reference cues, requiring consideration of social interaction factors beyond traditional physical cues. Although numerous studies have confirmed this viewpoint as previously described, discrepancies remain. For instance, Sjolund et al. (2014) found that regardless of whether a collaborative partner was present, participants tended to establish egocentric reference frame representations. This suggests that the specific manifestations and underlying mechanisms of spatial reference frame integration in collaborative tasks remain unclear. Therefore, this study aims to provide more data support for this field to further clarify related issues.

This study designed three experiments to investigate participants' spatial reference frame representation forms and influencing mechanisms in spatial collaborative tasks by constructing spatial collaboration scenarios, introducing others to participate in simplified spatial collaborative tasks with participants, and using the relative direction judgment paradigm based on previous research findings. Experiment 1 aimed to explore how individuals establish spatial reference frames in spatial collaborative tasks, specifically whether participants could integrate spatial reference frames through their partner's perspective. The hypothesis was that after collaborating with the partner to reconstruct the full scene, participants could integrate both parties' reference frames, manifested as no significant difference in pointing errors between the two reference directions. Experiment 2 increased the angular difference between participant and partner to create an approximate face-to-face situation, exploring whether participants could integrate reference frames. The hypothesis was that after changing the partner' s position, participants could still integrate both parties' reference frames, showing no significant difference in pointing errors between the two reference systems. Experiment 3 built on Experiment 2 by further adjusting experimental conditions and manipulating the timing of the partner' s departure to explore this factor' s influence on reference frame integration in collaborative tasks. The hypothesis was that even if the partner left after completing scene reconstruction, participants could still complete reference frame integration, specifically showing no significant difference in pointing errors between the two reference systems. This study expects that by integrating interpersonal collaboration factors into spatial reference frame theory, it can not only deepen the theory' s connotation but also extend its application to human-computer interaction, demonstrating certain practical value.

2. Experiment 1: Spatial Reference Frame Selection in Spatial Collaboration Tasks

2.1 Research Purpose and Background

This study employed several new experimental design considerations, primarily addressing two questions. First, how to achieve spatial collaboration? Spatial collaboration specifically means cooperating to complete a spatial task. This study improved upon the classic research paradigm of Mou and McNamara (2002) and referenced Holmes et al.' s (2018) multi-viewpoint task paradigm to create a "spatial jigsaw puzzle task" that required collaboration to complete. In this task, participants and their collaborative partners separately learned parts of the full scene (equivalent to each receiving several "jigsaw pieces"), then successively reconstructed their learned parts in the same area to ultimately complete full scene restoration (the "jigsaw puzzle" task). Second, what scene should be used? Previous spatial reference frame research mostly adopted "T-shaped" scenes (Li & Zhang, 2011; Mou et al., 2008; Street & Wang, 2014), which have a prominent vertical axis forming an obvious left-right symmetrical structure. To avoid the influence of scene internal structure on participants' spatial reference direction selection, we adopted a fully symmetrical circular scene

(Greenauer & Waller, 2008; Xie, 2020) with no prominent axis, ensuring that both parties' learning and memory of the full scene wouldn' t be biased by scene structure and facilitating examination of participants' selection between the two reference directions. Based on these two changes, we designed Experiment 1 to examine the feasibility of this experimental paradigm. If participants showed significant response advantages for their own perspective and its associated axis or coordinate system, it would indicate they still chose their own perspective to form the reference frame. If the advantage of their own perspective and its associated axis or coordinate system decreased, it would suggest they could notice and even process the collaborative partner' s perspective, attempting to incorporate it into their own spatial reference frame representation, forming an integration effect.

2.2 Methods

2.2.1 Participants Using G*Power 3.1.9.7 to estimate sample size and referencing Song et al.' s (2024) standards, we set $\alpha = 0.05$, 1-$\beta = 0.95$, and medium effect size f = 0.25, which indicated that 23 participants were needed for a single-factor eight-level within-subjects design. After balancing gender, scene, and trial presentation order, the experiment ultimately recruited 30 university student participants with a mean age of 22.97 years (SD = 0.90 years), including 15 males and 15 females, meeting the requirements. All participants had normal or corrected-to-normal vision, participated voluntarily, had not participated in similar experiments before, and received corresponding compensation upon completion. All experiments were approved by the Nanjing Normal University Biomedical Ethics Review Committee (Approval No.: NJU2022060041).

2.2.2 Experimental Design Experiment 1 used a single-factor within-subjects design. The independent variable was the imagined orientation in the relative direction judgment task, with eight levels: 0°/45°/90°/135°/180°/225°/270°/315°. Referencing Wang et al. (2021), we grouped these imagined orientations in two ways. The first grouping divided them into two coordinate systems: one being the participant' s coordinate system, including directions aligned with the participant' s perspective (0°/90°/180°/270°), and the other being the collaborative partner' s coordinate system, including directions aligned with the partner' s perspective (45°/135°/225°/315°). The second grouping divided them into four axes, combining the two orientations on each axis into one category, resulting in four axes: the 0°-180° axis (aligned with participant' s perspective, referred to as participant' s view axis), the 315°-135° axis (aligned with partner' s perspective, referred to as partner' s view axis), the 90°-270° axis (perpendicular to participant' s perspective, referred to as participant' s orthogonal axis), and the 45°-225° axis (perpendicular to partner' s perspective, referred to as partner' s orthogonal axis).

The dependent variables were pointing error (the angle between the participant's pointing direction and the object's actual direction) and reaction time. Following

Mou and McNamara (2002), we used pointing error as the primary measure of interest, with reaction time as a supplementary variable (Wang et al., 2021).

2.2.3 Apparatus and Materials **Experimental Apparatus:** The test program was compiled using E-Prime 2.0 and presented on a Dell Optiplex 7020 desktop computer with a 27-inch monitor (refresh rate: 60 Hz, resolution: 1920$\times$1080).

Questionnaire: Referencing previous research (Wang et al., 2021), this experiment used the Santa Barbara Sense of Direction Scale (SBSOD; Hegarty et al., 2002) as a tool to measure participants' spatial abilities. Questionnaire scores were used as covariates in subsequent ANOVA analyses.

Learning Scene: Both learning and collaboration phases took place in a cylindrical tent with a diameter of 3m. A round table (height: 1m, tabletop diameter: 1.2m) stood at the center for placing experimental materials, and an incandescent lamp at the top provided lighting. The experimental materials included nine common objects numbered 1-9: pen holder, jar, vase, toothpick, trash bag, cotton swab, candle, glass, and glue (see Figure 1

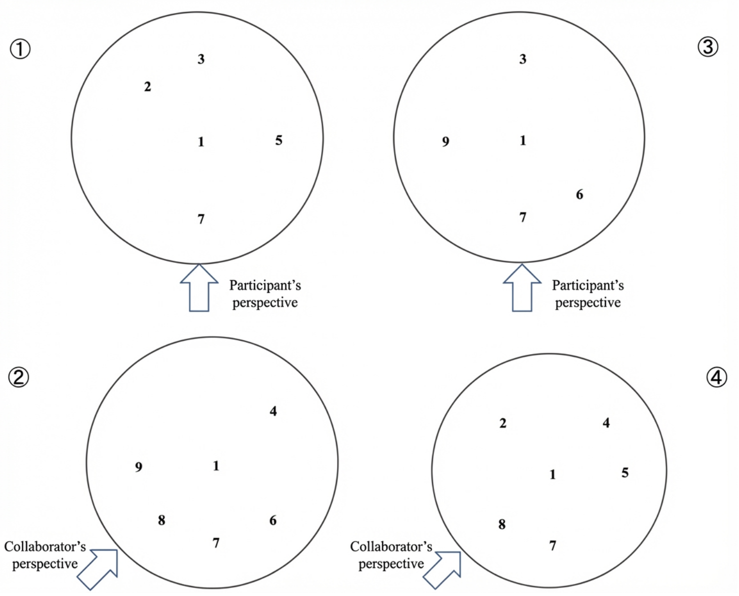

Figure 1: Figure 1

). All objects had no obvious orientation and no clear semantic connections.

Figure 1. Experimental scene structure and actual scene photograph

The complete scene used in the experiment consisted of the nine objects arranged in a circular layout. The pen holder was positioned at the center, with the remaining objects evenly spaced around it at 45° intervals (see Figure 1), each approximately 0.3m from the table center. Participants and their collaborative partners separately learned parts of this scene: the participant' s learning scene contained five objects, while the partner' s scene (which partially overlapped with the participant' s) contained six objects.

To reduce the influence of scene arrangement on participants, Experiment 1 set up two partial scenes for participants and collaborative partners (see Figure 2

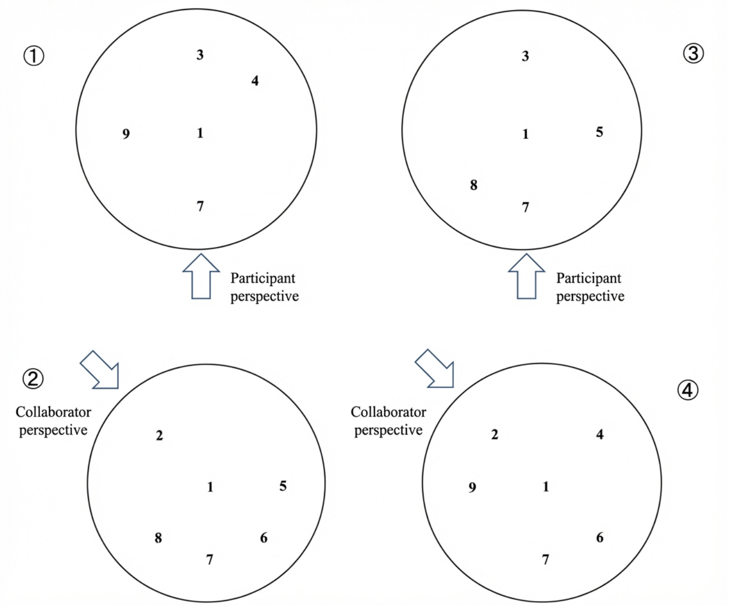

Figure 2: Figure 2

), with the left side as one group and the right side as another. That is, half of the participants learned Scene while their collaborative partner saw Scene ; the other half learned Scene while their partner saw Scene .

Scenes and shared a common principal axis (aligned with the participant' s viewing direction, the 0°-180° axis), formed by objects 3-1-7. Scenes and shared a common principal axis (aligned with the collaborative partner' s viewing direction, the 135°-315° axis), formed by objects 4-1-8. The participant' s and partner' s viewing perspectives did not possess each other' s advantageous principal axis directions—for example, Scenes and lacked objects 4 and 8, while Scenes and lacked object 3.

Figure 2. Schematic diagram of learning scenes in Experiment 1

Test Items: The test used the classic relative direction judgment task (Mou & McNamara, 2002), specifically phrased as "Imagine you are standing at the location of object A, facing object B. Please point to the direction where object

C is located"(in the actual experiment, A/B/C were replaced with object names). If the direction from A to B constituted the imagined orientation and it coincided with the participant' s selected reference direction, judgment performance would be better than for other items. After balancing the frequency of each object' s appearance in the statements and the frequency of each imagined orientation, we set 80 formal trials, 6 practice trials, and 8 filler trials. In the formal trials, each imagined orientation appeared 10 times, with 8 imagined orientations totaling 8 × 10 = 80 trials. Among the eight objects excluding the central pen holder, each object served as the locating object A and the orientation object B 10 times each (for A or B: 8 × 10 = 80 total), and as the test object C nine times. The pen holder only appeared as test object C eight times (for C: 8 × 9 + 8 = 80 total). Considering that the pen holder was the central object and that orientations including it (as object B) were relatively easy to imagine and less difficult, we designed eight such trials as filler trials, which were excluded from final data analysis. To reduce participant fatigue, based on pilot study reaction times and feedback, the formal experiment was divided into four blocks (with one filler trial before and after each block), and participants could rest between blocks. Additionally, to reduce familiarity and learning effects across trials, the trial order in each block was predetermined to avoid the same object or orientation appearing consecutively, and the order of the four blocks was balanced across participants.

2.2.4 Experimental Procedure The experimental procedure was based on the following assumption: Both collaborators learn separately, then reconstruct the scene through a "jigsaw puzzle" process. During this process, since the scene has no explicit geometric features (Li, Mou, & McNamara, 2009), both parties learn based on their own observation perspectives, thus adopting different reference frames. Participants establish scene representation using their own observation perspective direction (0°), then observe the composition of the partner' s objects through scene reconstruction operations, and deepen their impression and reflectively learn the partner' s intrinsic reference direction (i.e., the collaborative partner' s perspective at 315°) during the joint learning process. It is possible that due to the prerequisite of collaboration, participants need to incorporate the other' s reference frame into their own representation, manifested as no significant difference in reaction times between items containing their own reference frame direction (i.e., the 7$→1→3*direction)anditemscontainingthepartner′sreferenceframedirection*(*i.e.*,*the*8→1→$4 direction) in the relative position judgment task. Conversely, if integration does not occur, participants would respond faster to items containing their own reference direction.

The experimental procedure referenced Sjolund et al.' s (2014) research process, including six stages: preparation, learning, scene reconstruction, joint learning, scene labeling, and testing, with the entire process lasting approximately 60 minutes.

1. **Preparation Phase:** The experimenter introduced the experimental procedure and specific requirements to participants, ensuring they fully understood the task. Participants were informed that they and a partner would separately learn parts of the full scene, that the partner had already completed learning and was waiting to reconstruct the scene together, but that they wouldn' t know the partner' s location until the reconstruction phase. Participants were also told they wouldn' t know that the collaborative partner was actually a confederate. After the introduction, participants had approximately one minute to familiarize themselves with all nine objects and their corresponding names. Once participants confirmed their understanding, they completed the Santa Barbara Sense of Direction Scale on a laptop computer.
2. **Learning Phase:** After completing the questionnaire, participants put on a blindfold and entered the tent. Once confirmed to have moved to the preset position (participant' s viewpoint in Figure 2), they removed the blindfold (participants remained stationary during the first five phases) and received the instruction: "You will see several objects on the table. Please try to accurately remember each object' s location so you can reconstruct the scene with your partner in the next collaboration phase. You have unlimited learning time. Please inform me when you are confident you have remembered the scene." After participants completed learning from the "participant' s perspective" shown in Figure 2, they put on the blindfold again, and the experimenter removed all objects from the table into an object box.
3. **Scene Reconstruction Phase:** The experimenter left the tent, asked the collaborative partner (confederate) to put on a blindfold and guided them into the tent to the designated position (collaborative partner's viewpoint in Figure 2). Both parties then received the instruction: "Please remove your blindfolds together. You have separately learned parts of the full scene. Next, you need to work together to reconstruct the entire scene. I will randomly designate one of you to first arrange your learned scene, and the other will supplement based on their own learning. Now, please arrange your learned scene first." The experimenter always pointed to the partner first. Across all three experiments, the partner' s object arrangement order during the reconstruction phase was always top-to-bottom, left-to-right. For example, in the left scene of Figure 2, the reconstruction order was: 4$→1→8→9→6→$7 (the first three objects' order indicated the partner' s intrinsic reference direction; Mou, Liu, & McNamara, 2009). After the partner completed their task, they handed the object box to the participant, who continued to supplement the remaining object positions. Once all objects were placed on the table, the experimenter gave the instruction: "Now all objects are on the table. You still have opportunities to take turns adjusting object positions until both of you are satisfied. Then I will check the result. If there are no problems, the collaboration is successful; if there are serious deviations, I will ask you to collaborate again."

During this process, the experimenter provided no guidance or hints, but the partner would adjust object positions to approximate the preset full scene as closely as possible. After both parties finished adjusting, the experimenter photographed the result as a record of their collaborative performance.

4. **Joint Learning Phase:** Both parties remained in their positions and received the following instruction: "The reconstructed scene is very successful. All objects are in their correct positions. You now need to continue learning the locations of all objects in the scene to complete the final memory test. The test score will also be recorded as your collaborative performance. You still have unlimited learning time. Please inform me when you are confident you have remembered the scene."
5. **Scene Labeling Phase:** After participants indicated they had finished learning (the partner always indicated completion shortly after the participant), the experimenter removed all objects from the table and distributed a circular white paper and pen to both parties. Following a participant-first then partner-second order, both parties were asked to label the names of the nine objects from their own perspectives on separate sheets. The experimenter provided accuracy feedback, and participants had to complete three consecutive error-free labelings to proceed to the final test. While one party was labeling, the other rested nearby with a blindfold.
6. **Testing Phase:** After completing all scene labeling tasks, the experimenter first guided the participant out of the tent to complete the relative direction judgment task, then dismissed the collaborative partner. Before testing, the experimenter explained the task: The task required participants to use the interface-indicated imagined standing position and orientation as reference ("Imagine you are standing at the location of object A, facing object B. Please point to the direction where object C is located"), rotate the pointer by clicking the left mouse button, and click again to confirm the target object's direction. The program automatically recorded participants' response coordinates and reaction times. Participants first completed practice trials with immediate feedback from the experimenter until they fully understood the task and responded correctly three times. After practice, the formal experiment began. Participants were instructed to respond as quickly as possible while maintaining accuracy. The formal experimental procedure is shown in Figure 3

.

Figure 3. Task flow in the testing phase

2.2.5 Data Analysis The collected coordinate data were first converted to pointing error (transformed to absolute values, always positive). Following Wang et al.' s (2021) processing method, trials with pointing error ≥ 45° were defined as error trials, and data from participants with accuracy below 60% were

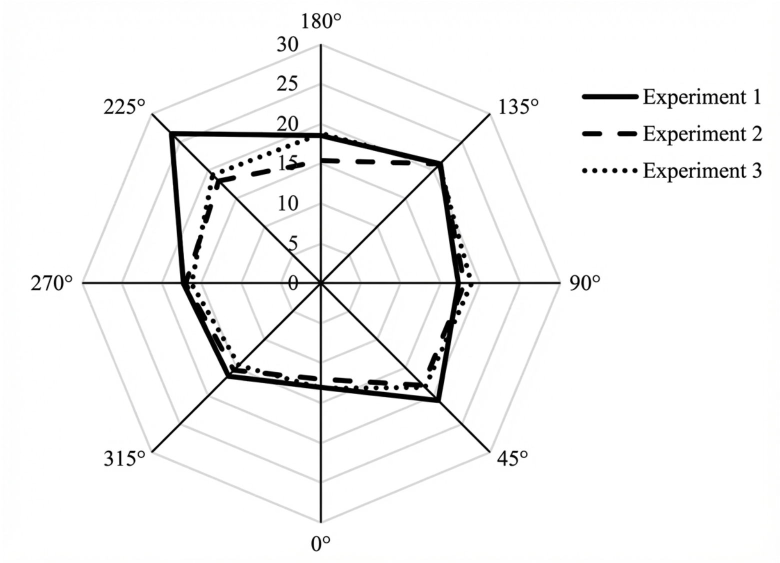

Figure 3: Figure 3

excluded. Subsequently, scores from the Santa Barbara Sense of Direction Scale were used as covariates, and SPSS was used for repeated measures ANOVA.

2.3 Results

In Experiment 1, participants' average response accuracy in the relative direction judgment task was 90.8%, and no participants were excluded due to low accuracy. Correlation analysis of all participants' reaction times and pointing errors yielded a Pearson correlation coefficient of -0.07 ($p = 0.71$), indicating no speed-accuracy trade-off. Detailed data are shown in Table 1 .

Table 1. Means (M) and Standard Deviations (SD) of Pointing Errors (degrees) and Reaction Times (seconds) for Each Imagined Orientation in Experiment 1

Using participants' scores on the Santa Barbara Sense of Direction Scale as covariates, ANOVA on pointing errors (see Figure 4) revealed a significant main effect of imagined orientation when divided into eight levels (Greenhouse-Geisser corrected), $F(4.99, 139.77) = 2.98$, $p = 0.01$, ${}^{2}p = 0.10$. Post-hoc comparisons showed that pointing errors at 0° orientation were significantly lower than the other seven orientations ($ps < 0.02$), while errors at 225° orientation were significantly higher than the other seven orientations ($ps < 0.01$). Additionally, $45° > 315°$ ($p = 0.04$) and $135° > 270°/315°$ ($ps < 0.03$). All other pairwise comparisons between orientations were non-significant (">" or "<" indicate significant differences, "=" indicates no significant difference, same below).

Figure 4 [FIGURE:4]. Pointing errors under different imagined orientation conditions in Experiment 1 (error bars represent $±$1SE, same below)

We conducted two types of grouping analyses to further explore participants' performance patterns. First, when dividing imagined orientations into two coordinate systems for ANCOVA, results showed a significant main effect, F(1, 28) = 5.39, p = 0.03, ^{2}p = 0.16, 95%CI = [-6.51, -2.50]. Participants' pointing errors for their own coordinate system (M = 16.66, SD = 5.80) were significantly lower than for the collaborative partner' s coordinate system (M = 21.40, SD = 7.86), indicating that participants did not incorporate the partner' s entire reference frame into their spatial representation. However, does this mean the partner' s perspective was completely ignored? To further investigate, we divided imagined orientations into four axes for analysis. Results showed a significant main effect, F(3, 84) = 4.13, p = 0.01, ^{2}p = 0.13. Post-hoc comparisons revealed that pointing errors for the 0°-180° axis (M = 15.67, SD = 7.53) were significantly lower than for the 45°-225° axis (M = 23.57, SD = 10.38, 95%CI = [-10.81, -4.05], $p < 0.001$) and the 135°-315° axis (M = 19.24, SD = 7.03, 95%CI = [-5.75, -0.11], p = 0.04). The 45°-225° axis showed significantly higher pointing errors than the 90°-270° axis (M = 17.24, SD = 6.30, 95%CI = [3.33, 8.82], $p < 0.001$) and the 135°-315° axis (95%CI = [1.46, 7.55], p = 0.01). No significant differences were found between the 0°-180° and 90°-270° axes (p = 0.32) or between the 90°-270° and 135°-315° axes (p = 0.20). These results suggest that although participants in Experiment 1 did not integrate both parties' reference frames, their response errors for the axis aligned with the collaborative partner' s perspective were significantly lower than for the axis perpendicular to the partner' s perspective, demonstrating that participants could already consider the partner' s perspective when representing spatial reference frames. However, this advantage did not extend to directions perpendicular to the partner' s perspective.

ANCOVA on reaction times showed that when imagined orientations were divided into eight levels, the main effect of imagined orientation was non-significant (Greenhouse-Geisser corrected), F(2.30, 64.37) = 2.30, p = 0.09. However, the advantage of 0° over other orientations ($ps < 0.01$) and the disadvantage of 225° over other orientations except 135° ($ps < 0.31$) were present. When imagined orientations were divided into two categories, the main effect was non-significant (F(1, 28) = 2.44, p = 0.13), but there was a trend for the participant' s perspective coordinate system (M = 16.25, SD = 7.67) to be faster than the partner' s perspective coordinate system (M = 19.29, SD = 10.15). When divided into four axes, the main effect was also non-significant, F(1.84, 51.56) = 1.76, p = 0.16. However, the 0°-180° axis (M = 15.46, SD = 6.40) showed advantages over the 45°-225° axis (M = 19.11, SD = 9.29, 95%CI = [-5.23, -2.07], $p < 0.001$) and the 135°-315° axis (M = 19.47, SD = 11.25, 95%CI = [-6.37, -1.64], $p < 0.01$). Although no significant main effects were found in reaction time analyses, the general trend was consistent with pointing errors.

2.4 Discussion

Experiment 1' s results differed somewhat from Sjolund et al.' s (2014) findings. In their study, some participants were assigned to individual learning groups and learned scenes from 0° or 45° directions; other participants were assigned to paired groups where one learned from 0° and the other from 45°, then collaborated to reconstruct the scene. After scene reconstruction, all participants completed relative direction judgment tasks individually. Results showed that regardless of whether participants reconstructed scenes individually or collaboratively, they all chose directions consistent with their own observation perspectives for spatial representation, and the partner' s perspective showed no advantage whatsoever. The presence of a collaborative partner did not significantly influence their spatial reference frame selection; participants established a single egocentric reference frame, demonstrating the important role of egocentric experience in spatial representation.

In this experiment, we attempted to improve the collaborative nature of the experimental context. Compared to previous research, we found that participants' data trends from the partner' s perspective showed some changes in this experiment: in the eight-level comparison, performance at 315° was second only to 0°; in the four-axis comparison, performance on the 135°-315° axis showed no difference from the 90°-270° axis. This indicates that the collaborative improvements in this experiment had some effect in promoting participants' memory representation from the partner' s perspective.

Could participants achieve complete reference frame integration? At this point, we considered the importance of gaze in perspective-taking mentioned in the literature review. Some scholars have pointed out that spontaneous perspective-taking only occurs when participants perceive that their partner can also visually receive scene information (Freundlieb et al., 2016; Ward et al., 2020). However, in Experiment 1, participants and partners stood relatively close (45° angular difference), possibly resulting in weak perception of the partner' s perspective and a continued tendency to establish reference frames from their own perspective. Additionally, participants might have believed the partner' s angle was too similar to warrant perspective-switching to understand the scene from the other' s viewpoint, reducing the likelihood of establishing a reference frame from that perspective. Furthermore, the 315° orientation (compared to 45°) has shown special judgment advantages in some studies (e.g., Zhou & Zhang, 2005). To address these issues, Experiment 2 changed the partner' s position based on Experiment 1 to create an approximate face-to-face situation while increasing the angular difference between participant and partner, further exploring the conditions that promote spatial reference frame integration.

3. Experiment 2: Spatial Reference Frame Selection in Spatial Collaboration Tasks—The Role of Partner Position

3.1 Introduction

Experiment 2 further explored the conditions that promote reference frame integration. Based on previous research, participants' attention to their partner' s gaze might affect perception of that direction. Therefore, Experiment 2 altered previous conditions to seek evidence of participants' reference frame integration.

3.2 Methods

3.2.1 Participants Based on G*Power calculations, the planned sample size remained 23, and the experiment ultimately recruited 30 university student participants with a mean age of 23.78 years (SD = 1.39 years), including 15 males and 15 females.

3.2.2 Experimental Design The apparatus, questionnaire, and data processing methods were the same as in Experiment 1. The difference was that during the learning phase, the collaborative partner' s perspective shifted from 315° to 225°. Under the premise of an unchanged full scene, Experiment 2 also set up two partial scenes for participants and collaborative partners: Participant Scene consisted of objects 1, 3, 4, 7, and 9, with the corresponding partner scene consisting of objects 1, 2, 5, 6, 7, and 8. Participant Scene consisted of objects 1, 3, 5, 7, and 8, with the corresponding partner scene consisting of objects 1, 2, 4, 6, 7, and 9. Both participant perspective scenes shared a common principal axis formed by objects 3-1-7, while both collaborative partner scenes shared a common principal axis formed by objects 6-1-2 (Figure 5 [FIGURE:5]).

Figure 5. Schematic diagram of learning scenes in Experiment 2

After this modification, the independent variable classification changed when divided into four axes: the 0°-180° and 90°-270° axes remained unchanged, the 45°-225° axis became the partner' s view axis, and the 315°-135° axis became the partner' s orthogonal axis. The experimental procedure remained the same as Experiment 1 except for the change in partner position.

3.3 Results

In Experiment 2, participants'average response accuracy in the relative direction judgment task was 93.1%, and no participants were excluded due to low accuracy. Correlation analysis of all participants'reaction times and pointing errors yielded a Pearson correlation coefficient of 0.24 ($p = 0.18$), with no significant correlation between accuracy and reaction time, indicating no speed-accuracy trade-off in Experiment 2. Details are shown in Table 2 .

Table 2. Means (M) and Standard Deviations (SD) of Pointing Errors (degrees) and Reaction Times (seconds) for Each Imagined Orientation in Experiment 2

3.3.1 Single-Experiment Analysis ANCOVA on pointing errors showed that when imagined orientations were divided into eight levels, the main effect of imagined orientation was non-significant (Greenhouse-Geisser corrected), F(3.85, 107.89) = 0.28, p = 0.96 (see Figure 6 [FIGURE:6]). When divided into two categories (participant' s coordinate system/partner' s coordinate system), the main effect was non-significant, F(1, 28) < 0.01, p = 0.96. When divided into four axes (0°-180° axis/90°-270° axis/45°-225° axis/135°-315° axis), the main effect was also non-significant (Greenhouse-Geisser corrected), F(2.22, 62.26) = 0.04, p = 0.99.

Figure 6. Pointing errors under different imagined orientation conditions in Experiment 2

ANCOVA on reaction times showed that when imagined orientations were divided into eight levels, the main effect was non-significant (Greenhouse-Geisser corrected), F(3.94, 110.28) = 2.06, p = 0.09. When divided into two categories, the main effect was significant, F(1, 28) = 5.03, p = 0.03, ^{2}p = 0.15, 95%CI = [-3.86, -1.48], with the participant' s coordinate system (M = 16.15, SD = 4.26) being faster than the partner' s coordinate system (M = 18.81, SD = 6.19). When divided into four axes, the main effect was significant, F(3, 87) = 3.04, p = 0.03, ^{2}p = 0.10. Post-hoc comparisons revealed that reaction times for the 0°-180° axis (M = 15.07, SD = 4.22) were significantly shorter than for the 45°-225° axis (M = 18.95, SD = 6.07, 95%CI = [-5.39, -2.37], p < 0.001), the 90°-270° axis (M = 17.22, SD = 4.84, 95%CI = [-3.35, -0.94], p = 0.001), and the 135°-315° axis (M = 18.67, SD = 12.78, 95%CI = [-5.26, -1.94], p < 0.001). The 45°-225° axis showed significantly longer reaction times than the 90°-270° axis (p = 0.01, 95%CI = [0.39, 3.08]). All other pairwise comparisons were non-significant (ps > 0.08). The reaction time results, though not showing significant main effects, showed the same general trend as pointing errors.

3.3.2 Joint Analysis To better compare results across experiments, we conducted a joint analysis of pointing errors from Experiments 1 and 2. When using eight imagined orientation categories, the interaction between imagined orientation and experiment was significant (Greenhouse-Geisser corrected), F(5.27, 300.43) = 2.44, p = 0.03, ^{2}p = 0.04. Based on previous analysis results, the data trend for imagined orientations in Experiment 1 (p = 0.01) differed significantly from that in Experiment 2 (p = 0.96). To better interpret the results, we focused on the other two independent variable classifications.

When imagined orientations were divided into two categories, the interaction between experiment and the two coordinate systems was non-significant, F(1, 57) = 2.43, p = 0.12. When divided into four axes, it should be noted that the status of the 45°-225° and 315°-135° axes differed between the two experiments. Therefore, we analyzed four levels: participant' s view axis/participant' s orthogonal axis/partner' s view axis/partner' s orthogonal axis. Results showed no significant main effect of between-experiment differences, F(1, 57) = 1.74, p = 0.19, but a significant interaction between experiment and the four axes,

$F(3, 171) = 2.90$, $p = 0.04$, $^{2}p = 0.05$. Based on previous analysis results, the difference among the four axes was significant in Experiment 1 ($p = 0.01$) but non-significant in Experiment 2 ($p = 0.99$). Combined with the data trend in Figure 7 and statistical results from Experiments 1 and 2, clear differences existed between the experiments. In Experiment 1, the participant' s view axis showed significant advantages, while the partner' s orthogonal axis was at a clear disadvantage. However, in Experiment 2, the disadvantage of the partner' s orthogonal axis decreased substantially, and the data line flattened, leading to reduced overall significance.

Figure 7 [FIGURE:7]. Pointing errors under four imagined orientation axis conditions in Experiments 1 and 2

Joint analysis of reaction times found that regardless of whether using eight imagined orientations, four axes, or two coordinate systems, their interactions with experiment were all non-significant ($Fs < 0.71$, $ps > 0.58$).

3.4 Discussion

Combined pointing error and reaction time results showed that after increasing the angle between participant and partner in Experiment 2 (placing the collaborative partner at the 225° position), evidence of participants' reference frame integration was found, supporting the previous hypothesis. That is, changing the partner's position from beside the participant to an approximate face-to-face arrangement promoted more advanced spatial reference frame integration. Reaction time results, however, showed that participants' response speed for their own orthogonal direction remained significantly faster than for the partner' s orthogonal direction, with similar trends across Experiments 1 and 2. Although spatial reference frame integration was clearly demonstrated in pointing errors, participants might have formed "primary-secondary" dual reference frames. The "primary" reference frame from their own perspective and the "secondary" reference frame from the partner' s perspective showed no significant difference in pointing error (response accuracy), but the "secondary" reference frame was slightly slower in reaction speed than the "primary" one.

Although Experiment 2 showed that participants could integrate spatial reference frames to some extent, both experiments involved multiple phases, particularly the post-collaborative reconstruction phase, leaving questions about how integration occurs and whether it depends on the partner' s continuous presence. Zhou et al. (2022) found that even when an object first appeared in a scene and then left, participants could still spontaneously adopt that object' s perspective. In our research, if participants continued subsequent tasks alone after collaboratively completing scene reconstruction with their partner, could they maintain or re-establish effective spatial reference frame integration? In other words, is the partner' s presence a necessary condition for maintaining this integrated state? This study planned to test whether this phenomenon applies to the current research context through Experiment 3. The experiment set the partner to leave after joint learning ended, providing preliminary exploration of integration timing, with future research extending the departure condition to other phases. This will help us better understand how individuals process spatial information during complex interactions and explore the underlying spatial cognitive mechanisms.

---

4. Experiment 3: Spatial Reference Frame Selection in Spatial Collaboration Tasks—The Role of Partner Departure Timing

Building on Experiment 2, Experiment 3 further manipulated the partner' s departure timing to alter experimental conditions, aiming to examine whether participants' performance in relative direction judgment tasks could replicate Experiment 2' s results.

4.1 Methods

4.1.1 Participants Based on G*Power calculations, the planned sample size remained 23. The experiment ultimately recruited 32 university student participants, with two excluded due to accuracy below 60%, leaving 30 participants with a mean age of 23.16 years (SD = 1.46 years), including 14 males and 16 females.

4.1.2 Experimental Design The experimental design, apparatus, materials, and data analysis were the same as in Experiment 2. The difference from the previous two experiments was that in this experiment, after participants completed the scene reconstruction task with their collaborative partner (i.e., after completing Phase 3), the experimenter used instructions to ask the partner to leave the tent first, while the participant remained inside to learn the full scene and complete the scene labeling phase alone. When participants indicated they had completed full scene memory, the experimenter asked them to continue with the scene labeling phase. After scene labeling, participants were asked to put on a blindfold and wait, while the experimenter left the tent to have the partner complete the scene labeling task (which took about one minute, the average time from the previous two experiments). Finally, the experimenter brought the participant out of the tent for the final testing phase.

4.2 Results

After excluding two participants, the remaining 30 participants in Experiment 3 had an average response accuracy of 91.5% in the relative direction judgment task. Correlation analysis of participants' reaction times and pointing errors yielded a Pearson correlation coefficient < 0.01 ($p = 0.99$), with no significant correlation between accuracy and reaction time, indicating no speed-accuracy trade-off in Experiment 3. Details are shown in Table 3 .

Table 3. Means (M) and Standard Deviations (SD) of Pointing Errors (degrees) and Reaction Times (seconds) for Each Imagined Orientation in Experiment 3

4.2.1 Single-Experiment Analysis ANCOVA on pointing errors showed that when imagined orientations were divided into eight levels, the main effect was non-significant (Greenhouse-Geisser corrected), F(3.48, 97.61) = 0.40, $p = 0.90$ (see Figure 8 [FIGURE:8]). When divided into two categories (participant' s coordinate system/partner' s coordinate system), the main effect was non-significant, F(1, 28) < 0.01, $p = 0.99$. When divided into four axes (0°-180° axis/90°-270° axis/45°-225° axis/135°-315° axis), the main effect was also non-significant, F(3, 84) = 0.09, $p = 0.97$.

Figure 8 [FIGURE:8]. Pointing errors under different imagined orientation conditions in Experiment 3

ANCOVA on reaction times showed that when imagined orientations were divided into eight levels, the main effect was significant, F(7, 196) = 2.75, $p = 0.01$, ^{2}p = 0.09. Post-hoc comparisons revealed that judgments at 0° orientation were significantly faster than the other seven orientations (ps < 0.03), while judgments at 225° orientation were significantly slower than the other seven orientations (ps < 0.03). Additionally, 45° > 270°/315° (ps < 0.02), 135° > 45°/90°/270°/315° (ps < 0.02), and 180° > 90°/270°/315° (ps < 0.02). All other pairwise comparisons were non-significant. When imagined orientations were divided into two categories, the main effect was non-significant, F(1, 28) = 3.10, $p = 0.09$. When divided into four axes, the main effect was also non-significant, F(3, 84) = 1.53, $p = 0.21$.

4.2.2 Joint Analysis We further conducted joint analysis of pointing errors from Experiments 2 and 3. Regardless of whether using eight imagined orientations, four axes, or two coordinate systems, their interactions with experiment were all non-significant (Fs < 1.33, ps > 0.26). Therefore, we conducted joint analysis of reaction times as a supplement.

Joint analysis of reaction times showed that with the main effect of between-experiment differences being non-significant (F(1, 84) = 0.29, $p = 0.59$), the interaction between the four-axis analysis and experiment was significant, F(3, 171) = 3.49, $p = 0.01$, ^{2}p = 0.06, indicating significant differences in reaction time performance across the four axes between the two experiments. Based on previous analysis results, the axis effect was significant in Experiment 2 ($p < 0.01$) but non-significant in Experiment 3 ($p = 0.21$). As shown in Figure 9 [FIGURE:9], the 0°-180° axis (participant' s view axis) in Experiment 2 was significantly faster than other axes, while this advantage diminished in Experiment 3. When using eight imagined orientations and two coordinate systems, their interactions with experiment were both non-significant (Fs < 1.45, ps >

0.18).

Figure 9 [FIGURE:9]. Reaction times under four imagined orientation axis conditions in Experiments 2 and 3

4.3 Discussion

The previous joint analysis results showed differences between Experiments 1 and 2 in pointing errors, specifically manifested as a significant decrease in pointing errors for the collaborative partner’ s orthogonal axis. This indicates that the design of changing the partner’ s position to approximate face-to-face significantly improved participants’ learning of the partner’ s perspective reference frame, thereby changing the form of participants’ spatial reference frame representation. However, Experiments 2 and 3 showed no significant differences in pointing errors, indicating that manipulating the partner’ s departure timing did not significantly alter participants’ spatial reference frame representation form. Both Experiment 3 and Experiment 2 showed evidence of spatial reference frame integration. Overall, this suggests that although the partner left before participants’ second representation, it did not affect participants’ ability to integrate reference frames. Additionally, we noted that participants’ self-view axis advantage was further affected in this process. The dynamic information of the partner’ s appearance and departure formed a cue that strengthened participants’ attention to the partner’ s perspective, deepening memory traces. In contrast, interference with their own perspective also increased, thereby narrowing the reaction time difference between their own perspective axis and the other’ s perspective axis. This was manifested as significant differences in reaction times across the four axes in Experiment 2, while these differences disappeared in Experiment 3, with the self-perspective axis advantage decreasing. Similar patterns have been found in perspective-taking test tasks during perceptual representation processes (Dou & Li, 2024), suggesting that tendencies in perceptual processes extend into memory representation.

5. General Discussion

This study took spatial reference frames as its starting point, combined with research experience in spatial perspective-taking and spatial collaboration, created a jigsaw-style spatial collaboration scenario where participants completed spatial tasks with partners, and used the relative direction judgment paradigm to design three experiments that progressively explored participants’ reference frame representation characteristics in spatial collaboration and the psychological mechanisms influencing spatial reference frame integration. From the comparison of pointing errors under the most typical eight imagined orientation conditions across the three experiments, we examined participants’ reference frame establishment strategies under three conditions: partner standing at 315°, partner at 225°, and partner at 225° with mid-experiment departure. As shown in Figure 10 [FIGURE:10], pointing errors at 225° decreased significantly from Experiment 1 to Experiments 2 and 3 (post-hoc tests showed ps < 0.02), with 45° (opposite to 225°) also showing a decreasing trend. This reveals that people adopt their partner' s perspective and integrate the partner' s reference frame into their own, though integration effectiveness is influenced by factors such as angular difference and partner departure timing.

Figure 10 [FIGURE:10]. Comparison of pointing errors under different imagined orientation conditions across three experiments (unit: degrees)

5.1 Reference Frame Integration: Establishment of Multiple Reference Systems Under Collaborative Conditions

Previous research using the same scene layout found that when no others were present in the scene, or when others were present but didn' t communicate or collaborate with participants, participants only represented the scene egocentrically from their own perspective (Li et al., under review). In our Experiment 1, participants collaborated with a partner at the 315° direction to complete spatial tasks, yet results showed participants still established spatial reference frames only from their own perspective, demonstrating advantages in both reaction time and pointing error for the self-perspective reference frame. No sufficient evidence of spatial reference frame integration was found in the results, similar to Sjolund et al.' s (2014) findings. In their experiments, regardless of whether participants reconstructed scenes individually or collaboratively, they all chose directions consistent with their own observation perspectives for spatial representation. The collaborative partner' s presence did not significantly influence their spatial reference frame selection; participants established a single egocentric reference frame.

In contrast, in Experiments 2 and 3, the pointing error advantage of participants' own coordinate system disappeared, though some differences remained in reaction times. We therefore proposed the existence of "primary-secondary"dual reference frames, where the "primary"reference frame from one's own perspective and the "secondary" reference frame from the partner' s perspective showed no significant difference in pointing error (response accuracy), but the "secondary" reference frame was slightly slower in reaction speed than the "primary" one. This suggests that from a collaborative perspective, people may form two reference coordinate systems under specific conditions: the coordinate system formed from one' s own perspective is more easily retrieved, while the coordinate system formed from the partner' s perspective requires more cognitive effort to retrieve, thus taking longer. Although similar patterns appeared in Galati et al. (2013) and Xie' s (2020) studies, the former used verbal analysis methods and task instructions to guide participants to represent the collaborative partner' s perspective, while the latter only examined reaction time performance and focused on viewpoint dependence in global representation. This study obtained this conclusion through more natural and comprehensive experimental and data analysis methods, advancing previous research findings and demonstrating the flexibility in individuals’ spatial reference frame establishment process. Thus, perspective-taking of collaborative partners can enter participants’ memory representation through perception.

5.2 Influence of Collaborative Partner Angle on Multiple Reference System Establishment

Comparing Experiments 1 with Experiments 2 and 3 reveals that reference frame integration occurs under certain conditions, likely related to attention. In Experiment 1, the collaborative partner stood beside the participant at the 315° direction, which might have resulted in weak perception of the partner’ s gaze and failure to adequately incorporate the other’ s perspective into their own spatial reference frame representation. In Experiments 2 and 3, however, the partner stood at the 225° “face-to-face” position, likely causing participants’ attention and perception of the partner to involuntarily increase, thereby promoting learning and representation of that direction and achieving spatial reference frame integration (Li & Su, 2015). Why does a face-to-face partner position better facilitate spatial reference frame integration? Several possible explanations exist:

First, the influence of others’ gaze. The spatial perspective-taking field has long focused on the role of “gaze” (Dou & Li, 2024). Researchers have used goggles or object occlusion to block others’ gaze to explore boundary conditions for spatial perspective-taking (Freundlieb et al., 2016; Kuhn et al., 2018), noting that participants only adopt others’ perspectives when they perceive their collaborative partner’ s gaze. Meanwhile, Xu (2017) pointed out that the primary condition for perspective-taking is individuals’ ability to distinguish between their own and others’ perspectives. In this study, compared to the 225° position, when the partner stood at 315°, participants had weaker perception of the partner’ s perspective, leading to insufficient distinction between the two perspectives and thus poor perspective-taking and coordinate system establishment. Notably, although the importance of others’ “gaze” in perspective-taking has been repeatedly mentioned, previous research using blind participants (Tinti et al., 2018) and manipulating perspective-taking targets’ gaze direction (Ward et al., 2020) both found that perception of others’ gaze is not a necessary condition for perspective-taking. Tinti et al. found that blind people can perspective-take through auditory stimuli, while Ward et al. noted that others’ spatial position rather than gaze direction triggers perspective-taking. Both studies demonstrate the importance of perceiving others’ positions. In our Experiment 1, the partner was in participants’ peripheral visual field, making it temporarily impossible to distinguish between gaze and position influences. Therefore, participants’ performance might also have been affected by their weaker perception of the partner’ s position.

Second, the influence of collaborative tasks. As previously mentioned, task instructions and task-related cues also affect participants’ spatial reference direction selection (Galati et al., 2013; Kelly et al., 2018; Shelton & McNamara, 2004). Salm-Hoogstraeten et al. (2021) conducted experiments with robots and found that participants only tended to adopt the robot' s perspective when they explicitly knew they needed to interact with it; conversely, in situations without explicit interaction needs, the robot' s gaze had no significant effect on participants' spatial task performance. In this study, the experimental information participants knew before the collaborative task mainly included: first, this was a learning and memory test task; second, they had a partner with whom they separately learned parts of the full scene; third, after learning, they would collaborate with their partner to complete scene reconstruction. Beyond this, participants knew nothing about their partner' s location or the spatial scene content from the partner' s perspective, and received no instructions to learn from the partner' s perspective. Thus, simply knowing there was a partner for spatial collaboration might have been sufficient to prompt participants to attend to the other' s perspective and establish a coordinate system based on it. Further analysis of Experiment 1 can support this inference. Although participants did not achieve spatial reference frame integration in Experiment 1, when dividing the partner' s perspective reference frame into the 315°-135° axis (aligned with partner' s perspective) and the 45°-225° axis (perpendicular to partner' s perspective), participants' judgments for the 315°-135° direction were significantly better than for the 45°-225° direction. This indicates that even without completing spatial reference frame integration, the partner' s presence already promoted participants' representation of that direction, though this influence was insufficient to change participants' spatial representation mode.

5.3 Influence of Collaborative Partner Departure Timing on Multiple Reference System Establishment

The main difference between Experiment 3 and Experiment 2 was that the collaborative partner left the tent after completing the scene reconstruction task, resulting in less co-presence time in the tent than in Experiment 2. Although Experiment 3, similar to Experiment 2, showed evidence of spatial reference frame integration, some differences existed in separate and joint analyses of Experiments 2 and 3. First, as previously mentioned, separate analysis of Experiment 2 found a significant main effect of reaction time, suggesting that although participants completed spatial reference frame integration, they might have formed "primary-secondary" dual reference frames. However, separate analysis of Experiment 3 found no significant main effect of pointing errors, and no significant differences in reaction times between participants' own and partner' s orthogonal directions, suggesting participants might have formed two reference frames without obvious primary-secondary distinction.

These results can be explained using Song and Dong' s (2023) multidimensional representation model of interpersonal collaboration. The model proposes that in joint tasks where individuals act together, representations are multidimensional, including spatial representations of the spatial dimension, embodied representations related to body parts, and social representations related to social interaction between individuals and others. That is, the final formation of spatial representation is influenced not only by spatial environmental features but also by individuals' social representations of others. In Experiment 2, participants jointly learned the full scene with their partner after collaborative scene reconstruction. On one hand, the partner' s continuous presence facilitated participants' learning of that direction; on the other hand, this somewhat reduced participants' memory burden—that is, a diffusion of responsibility effect occurred. These two aspects resulted in participants being able to learn the partner' s perspective well but with less processing depth than their own perspective, manifested as no significant difference in pointing errors between the two reference frames but slower reaction speeds for the partner' s perspective than their own. In Experiment 3, although the partner left after collaboration, the collaborative state before departure might still have influenced participants' spatial information processing. As Zhou et al. (2022) found, for an object that was present and then left, participants continued to be influenced by its perspective after its departure. Therefore, after the partner briefly viewed the full scene and left, participants might have felt obligated to help complete the unfinished task, and this obligation prompted them to increase attention to the partner' s perspective, thereby increasing processing depth. This ultimately resulted in no significant differences in pointing errors or reaction speeds between participants' own and partner' s orthogonal directions.

In summary, when others in the environment have social connections with participants rather than being merely irrelevant objects, the process of participants forming spatial representations is also influenced by others' behaviors. This indicates that spatial representation construction is not only the result of individuals' internal cognitive activities but is also profoundly affected by external environments, particularly social interaction factors. Future research can improve ecological validity through three approaches: first, systematically manipulate collaborative partners' "presence-absence" timing to observe how dynamic situational changes alter spatial representation in real time; second, introduce semantically rich language communication (e.g., task instructions, emotional feedback) to examine how verbal content enhances or interferes with spatial memory, thereby further verifying the influence of social contexts on spatial reference frames; third, introduce interpersonal collaboration variables such as collaboration willingness and satisfaction to systematically test how different collaboration characteristics differentially affect research results.

---

6. Conclusions

This study created spatial collaboration tasks and used the relative direction judgment paradigm to design three experiments exploring individuals' spatial reference frame representation in spatial collaboration and the psychological mechanisms influencing spatial reference frame integration. Based on comprehensive data analysis and discussion, the main conclusions are:

1. Multiple representation forms of reference frames exist in spatial collaborative tasks, with both single- and multi-frame representations possible.
2. Reference frame integration in spatial collaboration is modulated by the collaborative partner’ s position; an approximately face-to-face angle is more conducive to integration than a side-by-side position.
3. Even when the collaborative partner’ s presence time is reduced, participants can still complete reference frame integration. However, compared to when the partner remains present throughout, the partner’ s departure after the reconstruction phase prompts participants to form a more profound representation of the partner’ s perspective.

References

Christou, C. G., & Bülthoff, H. H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition, 27*(6), 996–1007.

Dou, Y., & Li, J. (2024). The spontaneity of Level-1 visual perspective taking: Under the condition of multiple avatars. *Acta Psychologica Sinica, 56*(4), 421–434. [豆艳, 李晶. (2024). 多对象情境中一级视觉视角采择的自发性表现. 心理学报, 56(4), 421–434.]

Furlanetto, T., Cavallo, A., Manera, V., Tversky, B., & Becchio, C. (2013). Through your eyes: incongruence of gaze and action increases spontaneous perspective taking. *Frontiers in human neuroscience, 7*, 455.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another’ s visuospatial perspective? *Journal of Experimental Psychology: Human Perception and Performance, 42*(3), 401–412.

Freundlieb, M., Sebanz, N., Kovács, Á. M., Gauthier, I., & Enns, J. T. (2017). Out of your sight, out of my mind: Knowledge about another person’ s visual access modulates spontaneous visuospatial perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance, 43*(6), 1065–1072.

Galati, A., & Avraamides, M. N. (2015). Social and representational cues jointly influence spatial perspective-taking. *Cognitive Science, 39*(4), 739–765.

Galati, A., Michael, C., Mello, C., Greenauer, N. M., & Avraamides, M. N. (2013). The conversational partner’ s perspective affects spatial memory and descriptions. *Journal of Memory and Language, 68*(2), 140–159.

Greenauer, N., & Waller, D. (2008). Intrinsic array structure is neither necessary nor sufficient for nonegocentric coding of spatial layouts. *Psychonomic Bulletin and Review, 15*(5), 1015-1021.

Hatzipanayioti, A., & Avraamides, M. N. (2021). Alignment effects in spatial perspective taking from an external vantage point. *Brain Sciences, 11*(2), 204.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah. I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30*, 425-447.

Holmes, C. A., Newcombe, N., S., & Shipley, T. F. (2018). Move to learn: Integrating spatial information from multiple viewpoints. *Cognition, 178*, 7-25.

Kelly, J. W., Costabile, K. A., & Cherep, L. A. (2018). Social effects on reference frame selection. *Psychonomic Bulletin & Review, 25*(6), 2339-2345.

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In: Freksa, C., Habel, C., Wender, K. F. (eds.) *Spatial Cognition. Lecture Notes in Artificial Intelligence 1404*(pp. 1-18). Berlin: Springer-Verlag.

Kuhn, G., Vacaityte, I., D' Souza, A. D. C., Millett, A. C., & Cole, G. G. (2018). Mental states modulate gaze following, but not automatically. *Cognition, 180*, 1-9.

Li, X., Mou, W., & McNamara, T. P. (2009). Intrinsic orientation and study viewpoint in recognizing spatial structure of a shape. *Psychonomic Bulletin & Review, 16*, 518-523.

Li, J., & Su, W. (2015). The limited effect of coincident orientation on the choice of intrinsic axis. *Perceptual and Motor Skills, 120*(3), 895-905.

Li, J., Wang, Q., Xie, N., & Qi, P. (no date). Multi-perspective spatial memory representation in the presence of others. *Cognitive Processing*. Manuscript under review.

Li, J. & Zhang, K. (2011). The effect of orientation coincidence of objects on intrinsic frame of reference system in symmetrical scene. *Acta Psychologica Sinica, 43*(3), 221-228. [李晶, 张侃. (2011). 对称场景中朝向一致性对内在参照系的影响. 心理学报, 43(3), 221-228.]

Mou, W., Fan, Y., McNamara, T. P., & Owen, C. B. (2008). Intrinsic frames of reference and egocentric viewpoints in scene recognition. *Cognition, 106*(2), 750-769.

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 162-170.

Mou, W., Zhao, M., & McNamara, T. P. (2007). Layout geometry in the selection of intrinsic frames of reference from multiple viewpoints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 145-

Mou, W., Liu, X., & McNamara, T. P. (2009). Layout geometry in encoding and retrieval of spatial memory. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 83-93.

Salm-Hoogstraeten, S. V., & Müsseler, J. (2021). Human cognition in interaction with robots: Taking the robot’ s perspective into account. *Human Factors, 63*(8),1396-1407.

Schafer, W., & Bowman, D. (2004). Evaluating the effects of frame of reference on spatial collaboration using desktop collaborative virtual environments. *Virtual Reality, 7*, 164-174.

Schober, M. F. (2009). Spatial dialogue between partners with mismatched abilities. In K.R. Coventry, T. Tenbrink, & J. A. Bateman (Eds.), *Spatial language and dialogue* (pp. 23-39). Oxford University Press.

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology, 43*, 274-310.

Shelton, A. L., & McNamara, T. P. (2004). Spatial memory and perspective taking. *Memory & Cognition, 32*(3),

Simons, D. J., & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychological Science, 9*(4), 315-

Sjolund, L. A., Erdman, M., & Kelly, J. W. (2014). Collaborative inhibition in spatial memory retrieval. *Memory & Cognition, 42*(6), 876-885.

Song, X., & Dong, M. (2023). A multidimensional representation model of interpersonal collaboration: From the perspective of cognitive representation. *Advances in Psychological Science, 31*(7), 1288-1302. [宋晓蕾, 董梅梅. (2023). 人际协同的多重表征模型——基于表征的视角. 心理科学进展, 31(7), 1288-1302.]

Song, X., Li, Y., Zhang, K. (2024). Regional differences of large-scale spatial orientation ability in virtual environment. *Acta Psychologica Sinica, 56*(1), 1-14. [宋晓蕾, 李宜倩, 张凯歌. (2024). 虚拟环境中大尺度空间定向能力的地域差异. 心理学报, 56(1), 1-14.]

Street, W. N., & Wang, R. F. (2014). Differentiating spatial memory from spatial transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 602-608.

Tinti, C., Chiesa, S., Cavaglià, R., Dalmasso, S., Pia, L., & Schmidt, S. (2018). On my right or on your left? Spontaneous spatial perspective taking in blind people. *Consciousness and Cognition, 62*, 1-8.

Valiquette, C. M., McNamara, T. P., & Smith, K. (2003). Locomotion, incidental learning, and the selection of spatial reference systems. *Memory & cognition, 31*(3), 479-489.

Wang, Y., Yu, X., Dou, Y., McNamara, T. P., & Li, J. (2021). Mental representations of recently learned nested environments. *Psychological Research, 85*(8), 2922-2934.

Ward, E., Ganis, G., McDonough, K. L., & Bach, P. (2020). Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze. *Cognition, 199*, 104241.

Xie, C., Liu, Q., Li, A., Tao, W., & Sun, H. (2009). Frame of reference in the mental representation of objects layout. *Acta Psychologica Sinica, 41*(5), 414-423. [谢超香, 刘强, 黎安娟, 陶维东, 孙弘进. (2009). 空间场景表征中的参照系选取. 心理学报, 41(5), 414-423.]

Xie, N. (2020). The representation of spatial scenes in perspective taking condition. Master' s thesis of Nanjing Normal University. [谢宁. (2020). 视角采择情境下的空间场景表征 (硕士学位论文). 南京师范大学.]

Xiao, C., Fan, Y., Xu, L., & Zhou, R. (2019). Human-centered human-robot natural spatial language interaction. *Chinese Journal of Applied Psychology, 25*(4), 319-331. [肖承丽, 范丫, 徐刘飞, 周仁来. (2019). 以人为中心的人-机器人自然空间语言交互. 应用心理学, 25(4), 319-331.]

Xiao, C., Sui, Y., Xiao, S., & Zhou, R. (2021). A new perspective on spatial interaction research: The effects of multiple social factors. *Advances in Psychological Science, 29*(4), 1-10. [肖承丽, 隋雨檠, 肖苏衡, 周仁来. (2021). 空间交互研究新视角: 多重社会因素的影响. 心理科学进展, 29(4), 1-10.]

Xu, Z. (2017). The effect of different visibility of perspective taking target on the automatic perspective taking. Master' s thesis of Zhejiang University. [徐阵雁. (2017). 观点采择对象的可视性对自动观点采择的影响 (硕士学位论文). 浙江大学.]

Zhao, M., Mou, W. (2005). Orientation specificity of spatial representation in multi-view learning. *Acta Psychologica Sinica, 37*(3), 308-313. [赵民涛, 牟炜民. (2005). 多视点学习条件下空间表征的朝向特异性. 心理学报,]

Zhou, J., Peng, Y., Li, Y., Deng, X., & Chen, H. (2022). Spontaneous perspective taking of an invisible person. *Journal of Experimental Psychology: Human Perception and Performance, 48*(11), 1186-1200.

Zhou, R., & Zhang, K. (2005). Direction judgments based on integrating reference frames in imagination. *Chinese Journal of Ergonomics, 11*(2), 7-9+12. [周荣刚, 张侃. (2005). 基于参照系整合的想象空间方位转化判断. 人类工效学, 11(2), 7-9+12.]

Source: ChinaXiv —Machine translation. Verify with original.