

Dual Routes of Chunking Social Interaction: Insights From Grouping Two Agent Actions in Working Memory

Jinglan Wu¹, Yang Guo^{1, 2}, Zhiyun Chen¹, Mowei Shen¹, and Zaifeng Gao¹

¹Department of Psychology and Behavioral Sciences, Zhejiang University

²Zhejiang Lab, Hang Zhou, People's Republic of China

Humans have evolved the sophisticated ability to extract social relations embedded in interactive entities. One typical demonstration is a social chunking phenomenon wherein the cognitive system chunks individual actions into a unified episode basing on perceived interactive actions. However, the mechanisms underlying social chunking remain to be elucidated. Most studies have adopted static images and manipulated interactions through agents' facingness (face-to-face vs. back-to-back). Connecting agents via directed contingent actions is crucial in forming real-life social interaction. Hence, we employed dynamic actions as stimuli, separated physical- and communicative-contingency interactive actions, and predicted that domain-general physical regularities and domain-specific social relationships are crucial in social interactions, respectively. We tested this prediction by using an involuntary chunking effect in working memory, wherein two individual actions are involuntarily chunked when containing task-irrelevant interactive information. We found that involuntary chunking occurred for both types of upright interactive actions (Experiments 1, 3, 5, and 6). Inverting actions erased the chunking of communicative- but not physical-contingency actions (Experiments 2, 4, and 5). The facingness of dyads did not participate in chunking physical-contingency actions but was a prerequisite for chunking communicative-contingency actions (Experiments 3 and 6). These results reveal the dual routes of chunking interactive actions. Moreover, they suggest that the chunking mechanisms of dynamic social interaction are distinct from those of static images, highlighting the importance of using dynamic stimuli to explore the mechanisms of social interaction in emerging people-watching interdisciplinarity.

Public Significance Statement

This study for the first time demonstrates that our mind involuntarily chunks the physical- and communicative-contingency interactive actions via domain-general physical regularities and domain-specific social relationships, respectively. The dual routes of chunking interactive actions in working memory indicate that social cues are only involved in grouping communicative-contingency interactive actions. The finding strongly suggests that the chunking mechanisms of dynamic social interactions are distinct from those of static images, underscoring the importance of using dynamic stimuli in interdisciplinary people-watching.

Keywords: social interaction, social chunking, cognitive chunking, working memory

Supplemental materials: <https://doi.org/10.1037/xge0001539.supp>

Social success requires the ability to process not only individual agents and their actions, but also pairs of agents and directed relationships (i.e., social interaction). Processing social interaction from

a third-person perspective—also called third-party encounters or people-watching—provides observers with social insight (Quadflieg & Koldewyn, 2017; Quadflieg & Westmoreland, 2019; Quadflieg

This article was published Online First January 22, 2024.

Jinglan Wu  <https://orcid.org/0000-0003-3056-1127>

Zaifeng Gao  <https://orcid.org/0000-0001-9727-8524>

This research was supported by the National Natural Science Foundation of China (32271090 and 32071044), Key Program of Natural Science Foundation of Zhejiang Province (LZ20C090001), and Fundamental Research Funds for the Central Universities. There is no financial interest. All data and study materials have been made publicly available at the Open Science Framework and can be accessed at <https://osf.io/6pyed/>. This study was not preregistered. The data and ideas included in this article have not been published or presented previously.

Jinglan Wu served as lead for data curation, formal analysis, validation, and visualization and contributed equally to conceptualization. Yang Guo served as lead for conceptualization, data curation, formal analysis,

investigation, methodology, and software. Zhiyun Chen contributed equally to data curation and served in a supporting role for writing—review and editing. Mowei Shen served as lead for project administration and supervision and contributed equally to writing—review and editing. Zaifeng Gao served as lead for conceptualization, funding acquisition, methodology, project administration, supervision, and writing—original draft. Jinglan Wu and Zhiyun Chen contributed equally to investigation and software. Jinglan Wu and Yang Guo contributed equally to writing—original draft.

Correspondence concerning this article should be addressed to Yang Guo or Mowei Shen or Zaifeng Gao, Department of Psychology and Behavioral Science, Zhejiang University, No. 3 Hainayuan Building, Zijingang Campus, 866 Yuhangtang Road, Hangzhou, Zhejiang, 310058, People's Republic of China. Email: guoyang_psy@zju.edu.cn or mwshen@zju.edu.cn or zaifengg@zju.edu.cn

et al., 2015), for instance, discovering cooperative agents interested in forming new alliances and avoiding agents prone to dangerous, uncooperative, unfair, or immoral social behavior. This fundamental capability may be facilitated by evolutionary pressure (Bryant et al., 2016), as humans must forge close bonds to survive adversity. Given the significance of social interaction in human life, interdisciplinary research has burgeoned in the past decade (for reviews see Bietti & Sutton, 2015; Capozzi & Ristic, 2018; Cheng, Yuan, & Jiang, 2021; Feng et al., 2021; Hari et al., 2015; Quadflieg & Koldewyn, 2017; Quadflieg & Penton-Voak, 2017; Quadflieg & Westmoreland, 2019).

One important finding is the social chunking phenomenon, wherein the cognitive system is posited to automatically chunk individual actions into a unified episode based on perceived social interactions (e.g., Abassi & Papeo, 2020; Adibpour et al., 2021; Ding et al., 2017; Ji et al., 2020; Lu et al., 2022; Paparella & Papeo, 2022; Papeo, 2020; Papeo et al., 2017, 2019; Vestner et al., 2019; Vestner, Gray, & Cook, 2021; Vestner, Over, Gray, & Cook, 2021). However, underlying mechanisms governing social chunking remain a fundamental issue yet to be fully elucidated. The present study aimed to address this by demonstrating that two distinct routes contribute to the chunking of social interaction.

Chunking of Social Interaction: A Domain-Specific or Domain-General Mechanism?

Exploration of the social chunking phenomenon is motivated by a domain-specific account of social interaction processing in cognition. That is, domain-specific visual processing is selectively or preferentially engaged by social interactions, and helps interacting individuals compete more effectively for limited resource (Papeo et al., 2019; Vestner et al., 2019, 2020). Akin to face and body perception, accumulating evidence suggests domain-specific cerebral cortices for perceiving social interaction (Isik et al., 2017; McMahon et al., 2023; Walbrin et al., 2018; see also Wurm & Caramazza, 2019; Wurm et al., 2017), possibly with a homolog in the nonhuman primate brain (Sliwa & Freiwald, 2017). Hence, human vision may be especially sensitive to spatial relations that reliably correlate with social interactions. For instance, physically interacting agents are often physically close to and facing each other.

Researchers have recently focused on whether static face-to-face dyads exhibit a processing advantage over nonrelated dyads (e.g., back-to-back) in visual search, working memory (WM), and long-term memory tasks. A processing advantage for facing dyads was robustly revealed in visual search although the dyads stimuli were static postures (Papeo et al., 2019; Vestner et al., 2019, 2020; Vestner, Gray, & Cook, 2021). This advantage diminished when inverting the stimuli, analogous to a robust marker of the highly specialized face perception process. Further, the face-to-face dyads were maintained better in both WM and long-term memory relative to the back-to-back dyads (Paparella & Papeo, 2022; Vestner et al., 2019). These findings suggest that face-to-face dyads are chunked as holistic units (see also Ji et al., 2020) and engage in domain-specific social interaction processing that aids in competition for limited attention. Additional research revealed that chunking facing dyads as one unit improved the discrimination of individual agents (Manera et al., 2011; Neri et al., 2006) and shortened the subjective duration of interactive actions (Liu et al., 2018). Meanwhile, the holistic unit impaired the maintenance of precise distance between face-to-face dyads in WM (Vestner et al., 2019), and led to more binding errors in the constituent agents of face-to-face dyads in

long-term memory (Vestner, Flavell, et al., 2021). Together, facingness appears to be a crucial, reliable signal to evoke domain-specific processing, and the facing dyad in the visual cortex may be an early rudimentary visual representation of social interaction (Abassi & Papeo, 2020).

However, recent studies have challenged the domain-specific account of chunking facing dyads. These findings support a domain-general account, suggesting that the chunking effect can be attributed to visuospatial attentional mechanisms rather than the social nature of the dyads. Specifically, Vestner et al. (2020) proposed a directional cueing account, suggesting that the previous search advantage for facing dyads was actually a by-product of individual heads and bodies directing observers' visuospatial attention. The face-to-face arrangement may create a distinct "hotspot," which is a relatively small region of space that attracts attention through multiple cues. These hotspots may guide observers' attention to the target location relatively early in a serial visual search (Vestner et al., 2020; Vestner, Flavell, et al., 2021). For "nonsocial" stimuli (e.g., pairs of arrows and cameras) that could direct observers' visuospatial attention, a similar search advantage could be demonstrated when they are displayed face-to-face (Vestner et al., 2020; Vestner, Over, Gray, & Cook, 2021; Vestner, Over, Gray, Tipper, & Cook, 2021). Although the domain-general explanation was initially raised in the context of the visual search task, it may explain the chunking results of face-to-face static dyads in WM and long-term memory. For instance, Paparella and Papeo (2022) found that meaningless face-to-face static dyads (i.e., containing only domain-general cues) were memorized better than back-to-back ones in a WM task, when no concurrent digit articulatory rehearsal task was accompanied. Therefore, the factor driving social chunking might not be related to social interaction or social relationship.

Real-World Social Interactions Are Dynamic

Most previous studies about the chunking of social interaction used static images only, which is a sensible first approach to demonstrate the basic phenomenon. At the same time, it must not be ignored that real-world social interactions are highly dynamic. The human brain has evolved to address dynamic information (Dunbar & Shultz, 2007; Hari et al., 2015) and is equipped with domain-specific neural hubs to tackle social interaction (Centelles et al., 2011; Isik et al., 2020; Okruszek et al., 2018; Quadflieg et al., 2015; Walbrin & Koldewyn, 2019; Walbrin et al., 2018). Therefore, using minimalist displays of static images may obscure or distort the chunking mechanisms of dynamic social interaction (Isik et al., 2020; Schweinberger & Döbel, 2021). Indeed, two recent WM studies demonstrated that facingness is likely one of several prerequisites but not sufficient in chunking interactive actions (Ding et al., 2017; Lu et al., 2022) for dynamic actions. Without meaningful social interaction, facingness alone did not lead to the chunking phenomenon in WM. Hence, we argue that incorporating the dynamic essence of social interaction will provide insights for a more comprehensive understanding of social interaction processing (including chunking) and will establish a solid foundation for interdisciplinary people-watching.

Crucial in forming real-life social interaction is connecting distinct agents via directed contingent actions between them. Researchers have posited that two distinct types of contingencies are embedded in interactive actions: physical and communicative contingency (Fedorov et al., 2018; Hafri & Firestone, 2021; Manera et al., 2011).

In physical-contingency interactions (e.g., dyads fighting, dancing, and jumping), the actions of interactive agents are often similar because of action coupling (or synchronization in phase), similarity, and symmetry in individuals' movement patterns (Bernieri & Rosenthal, 1991; Nessler & Gilliland, 2009), hence the action of one agent can be used to predict the action of the other (Manera et al., 2011). In communicative-contingency interactions (e.g., A asks B to sit down, and then B sits down), the two actions aimed at communicating certain information with distinct actions, and the performance of an agent's action is not physically contingent upon the performance of the other's action (Manera et al., 2011).

Most of the previous chunking studies of human actions did not distinguish the physical- and communicative-contingency-dominated interactions. Certain chunking-related studies imply that the two types of interaction share processing properties (e.g., Manera et al., 2011; Neri et al., 2006; Su et al., 2016). However, recent neuroimaging research has brought to light an intriguing revelation: communicative-contingency interactions, in contrast to joint actions, which largely overlap with the physical-contingency interactions, are uniquely responsible for eliciting responses in the superior temporal sulcus. This finding implies the importance of distinguishing between the two categories of interaction (McMahon et al., 2023). We predict that distinct chunking mechanisms may exist between physical- and communicative-contingency interactions. Particularly, the former may predominantly rely on physical regularities (e.g., gestalt law similarity), which belongs to a typical domain-general "visual chunking." Conversely, the latter may rely on social relationship embedded in social interaction, which is the real domain-specific "social chunking."

The Current Study

The current study presented participants with either communicative-contingency actions, or physical-contingency actions, aiming to examine whether two distinct routes contribute to the chunking of social interaction in the mind. We achieved this aim by taking advantage of the social chunking phenomenon in WM, which refers to the process that WM involuntarily chunks interactive actions when only the individual actions have to be retained (Ding et al., 2017; Lu et al., 2022). The WM social chunking enables the examination of a default chunking mechanism while avoiding potential top-down processing strategies, and allows a direct examination of the role of facingness in chunking social interaction in a dynamic interaction scenario.

To examine whether chunking occurs involuntarily and further test the role of facingness in chunking dyads, we presented participants with three conditions in Experiments 1–4: two face-to-face dyads with social relationship (interactive pair), two face-to-face dyads without social relationship (noninteractive pair), and four individual actions (individual four). Moreover, to evaluate the holistic processing property of social structures, we presented upright (Experiments 1 and 3) or inverted actions (Experiments 2 and 4). We predict that, for upright stimuli, involuntary chunking occurs for both types of social interaction, yielding higher performance in the interactive pair than that in the individual four (Experiments 1 and 3). Further, involuntary chunking is predicted to remain for inverted physical-contingency interactions, because physical regularities do not change (Experiment 2); yet the involuntary chunking will disappear for inverted communicative-contingency interactions, as stimuli-inversion disrupts the social relationship (Experiment 4). Additionally, if facingness is determinant in chunking social interaction,

the interactive and noninteractive pairs should not differ for upright stimuli (Experiments 1 and 3).

To confirm the distinct mechanisms involved when retaining upright (Experiment 3) and inverted (Experiment 4) communicative-contingency interactions in WM, Experiment 5 required participants to retain the interactive-pair and individual-four conditions of communicative-contingency interactions, while manipulating the action orientation (upright vs. inverted) in different blocks. We predict a similar pattern of results as in Experiments 3 and 4. Finally, to further distinguish the distinct cues involved in chunking social interaction and the role of facingness, Experiment 6 transformed the two types of interactive pairs from face-to-face to back-to-back, while maintaining the noninteractive pairs as face-to-face. We predict that the involuntary chunking will occur for back-to-back physical-contingency interactions as physical regularities remain intact. However, the involuntary chunking will disappear for back-to-back communicative-contingency interactions, as the manipulation of back-to-back positioning impairs social relationships.

General Method

Although the two types of contingencies often intertwine in real-life social interactions, we could largely manipulate them independently, for instance with the technique of point light displays (PLDs; e.g., Cheng, Liu, et al., 2021; Ding et al., 2017; Ji et al., 2020; Lu et al., 2022). All six experiments used PLDs as stimuli to avoid contaminations from faces, clothes, etc., and shared similar designs and procedures. We reported the common elements together.

Participants

Participants completed the experiments online. All participants were students from different universities in China. They provided signed informed consent and self-reported gender and age information, had normal color vision and normal or corrected-to-normal visual acuity, and received payment for their participation. This study was approved by the Research Ethics Board of Zhejiang University.

The sample size of the current study was determined based on the sequential Bayes factor design (Schönbrodt et al., 2017). The Bayes factor (BF_{10}) value indicates the ratio of the likelihood of supporting an alternative versus a null hypothesis (for the main effect of memory condition in Experiments 1–4; for the interaction between memory condition and action orientation in Experiment 5; and for the interaction between memory condition and action type in Experiment 6; Jeffreys, 1961). Following the sequential Bayes factor design, the BF_{10} was repeatedly calculated using JASP 0.16.3.0 after reaching the minimum sample size. If the predetermined evidence threshold was not reached, more participants were recruited until the Bayes factors reached this threshold. The minimum sample size was set to 40 before data collection, and the threshold for stopping sampling was set to $BF_{10} > 5$ or $BF_{10} < 1/5$.

Apparatus and Stimuli

Participants completed the experiment through their laptops with a screen 13–16 inches wide and in a quiet environment. The experimental program was written in JavaScript and jsPsych 7.1 and run by the Chrome browser.

The dynamic actions were conveyed by PLDs (Johansson, 1973). This technique depicts human actions via a set of light points that

isolate human kinetic information from other sources (e.g., texture, hair, and clothes). PLDs were selected from the Social Perception and Interaction Database (SoPID) recorded by Okruszek and Chrustowicz (2020), which offers both physical- and communicative-contingency interactions. The materials in the original database were skeletal animations in fbx format. To transform these into PLDs, we extracted the 3D coordinates of each joint point from these skeletal animations using the BioMotion Toolbox (Su et al., 2016) and then normalized each animation to approximately 4.2° (height) \times 1.8° (width). Each animation had 24 white points drawn according to their coordinates on a black background, contained 180 frames of images, and was played at a frequency of 60 Hz for 3 s.

Both physical- and communicative-contingency interactions contained nine distinct pairs (see Table 1). No touching of bodies occurred during the action, and any two face-to-face actions were displayed in distinct locations, with the center distance between the two actions being approximately 2° . The spatial distances between physical- and communicative-contingency interactions were comparable. It is of note that we trimmed and adjusted the interaction animations to ensure they all had a uniform duration of 3 s, which resulted in certain deviations from the original stimuli. Additionally, SoPID's assessment of these interactions was carried out with European participants. Thus, it is imperative to ascertain the suitability of these selected interactions for Chinese participants, as detailed in [Evaluation of Interactive Actions section in the online supplemental materials](#). We began by investigating whether the stimuli effectively conveyed clear and meaningful social interactions. To do this, we had 42 participants classify whether the animations contained meaningful social interactions. Notably, all the interactive actions were unanimously categorized as meaningful interactions. We then examined participants' categorization of interactive actions, determining whether they belonged to communicative- or physical-contingency interactions. The results from 50 participants were consistent with our expectations. Finally, one of the features that distinguishes physical- and communicative-contingency interactions is the similarity of their constituent actions as per their operational definitions. To examine whether the current set of interactions met this feature, we engaged another 100 new participants. The results revealed that physical-contingency interactions exhibited a higher degree of similarity compared to communicative-contingency interactions. In sum, these findings substantiate the suitability of the materials currently employed in this study.¹

In Experiments 1–4, the memory array contained three memory conditions with the same set of actions: two interactive pairs, two noninteractive pairs, and four individual actions (see Figure 1A). The noninteractive action pairs were mismatched interactive pairs and conveyed the least degree of interactive information. For instance, if AB and CD were the original interactive pairs, AD and CB were noninteractive pairs. Critically, interactive and noninteractive pairs were presented face-to-face, with equal spatial proximity. Hence, a gradient of chunking cues was observed across the three memory conditions. Specifically, the individual-four condition lacked any chunking cue, the noninteractive-pair condition solely featured a facingness cue (come with spatial proximity, without social relationship), while the interactive-pair condition encompassed both facingness cue and social relationship. Stimuli were displayed on an invisible circle with a radius of 5.5° from the screen center. For the interactive- and noninteractive-pair conditions, the centers of the action pairs were randomly set at two of the five

positions of the circle; for the individual-four condition, the centers of each action were randomly selected at four positions. There were 18 trials for each memory condition and each action was repeated 4 times in the memory array.

In Experiment 1, the stimuli consisted of upright physical-contingency interactions, whereas in Experiment 2, the stimuli were inverted. Experiment 3 employed upright communicative-contingency interactions as stimuli, while Experiment 4 utilized inverted versions. In Experiment 5, a direct comparison was made between inverted and upright communicative-contingency interactions, examining the interactive-pair and individual-four conditions. Experiment 6 investigated four distinct memory conditions, including back-to-back physical- or communicative-contingency interactions, as well as face-to-face noninteractive pairs formed from physical- or communicative-contingency interactions.

Design and Procedure

Experiments 1–4 included the memory condition (interactive pair, noninteractive pair, and individual four) as a within-subject factor. Experiment 5 included the within-subject factors of memory condition (interactive pair vs. individual four) and action orientation (upright vs. inverted). In Experiment 6, the within-subject factors were memory condition (interactive vs. noninteractive pair) and action type (physical- vs. communicative-contingency interactions).

Participants were explicitly instructed to memorize the four individual actions without implying any relationship between them. They completed eight practice trials before the formal experiment to ensure they understood the task. A 2- to 3-min break every 18 trials was allowed during the formal experiment. Experiments 1–4 lasted about 20 min (54 trials) and Experiments 5–6 lasted about 25 min (72 trials).

Each trial (see Figure 1B) began with two white digits presented on the screen center for 1,000 ms. Participants were asked to repeat the two digits aloud continuously throughout a trial at a rate of approximately two digits per second to inhibit the verbal labeling of actions² (Baddeley, 1986; Gao et al., 2015; Vogel et al., 2001). Chen and Cowan (2009) suggested that reliable chunks can be best observed when articulatory rehearsal of phonological representations is prevented by making participants engage in concurrent articulation during encoding. Next, a fixation cross was presented for 1,000 ms to remind the participants of the beginning of the memory task. The memory array was presented for 3,000 ms, followed by a 1,000 ms fixation interval. Then, a single action probe was

¹ Owing to technical constraints, our choice of interactive stimuli was limited to SoPID. However, the interactions in SoPID were not consistently designed in accordance with strict physical- and communicative-contingency criteria. As a result, certain communicative-contingency interactions included elements of physical contingency, and some physical-contingency interactions had aspects of communicative contingency. To be precise, we should refer to the present communicative-contingency interactions as “rich communicative-contingency interactions,” while the physical-contingency interactions should be denoted as “rich physical-contingency interactions.”

² In studies investigating the WM mechanisms of visual stimuli, it has become customary to incorporate an articulatory task as a means of impeding the utilization of a verbal labeling strategy. While the introduction of an articulatory task effectively hinders the verbal labeling of these stimuli, it does not interfere with the unintentional retrieval of semantic information intertwined within the stimuli. This semantic information can originate from long-term memory or be directly extracted during the perceptual process.

Table 1*Description of the Action Content of the Memory Materials From SoPID*

Physical-contingency actions	Communicative-contingency actions
A throws the ball to B, B catches and throws it back to A	A asks B to stand up, B stands up
A and B play football	A asks B to sit down, B sits down
A and B practice fencing	A points something in the sky, B looks at it
A and B throw punches at each other while boxing	A points something on the floor to B, B picks it up and hides in a pocket
A and B perform a kickboxing routine	A asks B to squat down, B squats down
A and B perform a dance routine	A asks B to come closer, B comes closer
A waves to B, B waves back to A	A accuses B of something, B denies it
A points something on the floor to B and squats down to look at it, B squats down next to A to look at it	A and B quarrel with each other
A and B are excited and they give each other high five	A tells some good news to B, B starts jumping

Note. SoPID = social perception and interaction database.

presented at the screen center. Participants were required to judge whether the probe had appeared in the memory array within 4,000 ms, by pressing “F” or “J” on the keyboard for “yes” and “no,” respectively. The probe action had a 50% chance of appearing in the memory array. After a key press or timeout, a red digit appeared, and participants were required to judge whether the digit was one of the repeated digits within 2,000 ms. If participants did not respond within the time window, the corresponding trial was considered incorrect. The intertrial interval was randomly selected between 1 and 1.5 s. Of both tasks, accuracy instead of reaction time (RT) was emphasized.

Analysis

Data in the WM task were analyzed regardless of the performance of the digit task, as the overall performance of the digit task was high. Memory accuracy was transformed into detection sensitivity (d') based on the signal detection theory (Stanislaw & Todorov, 1999).³ Participants with accuracies lower or equal to 0.8 in the digit task, or chance level in the memory task were excluded. We performed a one-way repeated-measures analysis of variance (ANOVA) on d' for Experiments 1–4, with the memory condition as a within-subject factor. In cases where the memory condition produced a significant main effect, post hoc Bonferroni correction was applied. For Experiments 5 and 6, a two-way repeated-measures ANOVA was conducted on d' , with within-subject factors of memory condition and action orientation in Experiment 5, and memory condition and action type in Experiment 6.

The BF_{10} was calculated to compare evidence for the alternative (H_1) and null (H_0) hypotheses (Jeffreys, 1961). $BF_{10} > 3$ or $< 1/3$ indicated moderate evidence for the presence or absence, respectively, of a main effect or interaction under consideration. Similarly, $BF_{10} > 10$ and $BF_{10} < 1/10$ indicated strong evidence for the presence or absence, respectively, of a main effect or interaction under consideration. The BF_{10} -inclusion values were reported for main and interaction effects in the models, which indicate the likelihood of the data under models that included rather than stripped an effect.

Transparency and Openness

We reported how we determined the sample size, all data exclusions, all manipulations, and all measures in the study. All data, analysis codes, and research materials are available at <https://osf.io/6pyed/>. Data were analyzed using JASP 0.16.3.0. This study's design and its analysis were not preregistered.

Results

Experiment 1: Retaining Upright Physical-Contingency Interactions

In Experiment 1, a total of 42 participants (16 male, 26 female, 18–34 years old) were asked to memorize upright physical-contingency actions. Two male participants were removed because of low accuracy in the digit task.

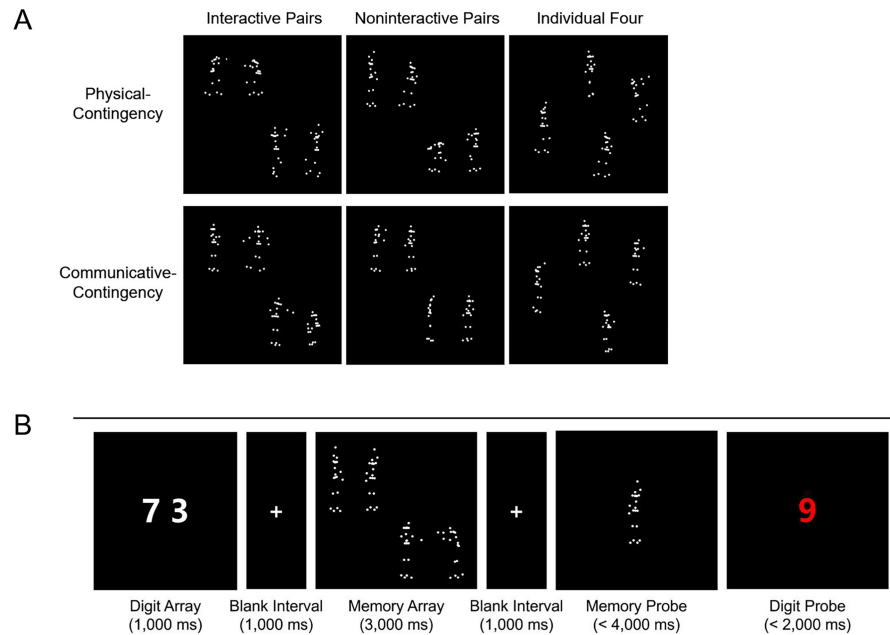
The mean accuracy on the digit task was 0.97. The d' values for each memory condition are shown in Figure 2A. ANOVA revealed a significant main effect of memory condition, $F(2, 78) = 18.101$, $p < .001$, $\eta_p^2 = .317$, $BF_{10} = 4.104e + 5$. Post hoc comparisons revealed that the d' of the interactive pair was significantly higher than that of the noninteractive pair, $t(39) = 4.576$, $p < .001$, Cohen's $d = 1.006$, $BF_{10} = 3.771e + 3$, and individual four, $t(39) = 5.672$, $p < .001$, Cohen's $d = 1.247$, $BF_{10} = 2.681e + 3$, whereas there was no significant difference between the latter two, $t(39) = 1.096$, $p = .829$, Cohen's $d = 0.241$, $BF_{10} = 0.291$.

Consistent with previous studies (Ding et al., 2017; Lu et al., 2022), Experiment 1 showed that the memory performance of interactive pair was better than that of noninteractive pair and individual four. This indicates that physical-contingency interactions can be stored in WM as chunks. Noninteractive pairs contained spatial proximity cues and were displayed face-to-face compared to individual four. However, no significant difference existed between them, and a Bayesian value of less than one-fifth favored the null hypothesis. This finding implies that only facingness or spatial proximity cues embedded in individual actions cannot lead to chunking in WM.

Experiment 2: Retaining Inverted Physical-Contingency Interactions

In Experiment 2, participants memorized inverted physical-contingency actions, which effectively eliminated the contribution of social relationship in the interactive set while retaining physical cues (e.g., Grossman & Blake, 2001; Gu et al., 2019;

³For raw accuracy, similar results as d' were achieved. These are reported in the online supplemental materials. In addition, RT of the WM task were recorded. However, due to the emphasis on accuracy and the online mode of experimentation, the RT results showed a large variance. No clear pattern across experiments existed, hence these results are not reported.

Figure 1*Example Memory Arrays for Each Condition (A) and the General Trial Procedure (B)*

Note. Each trial began with two white digits presented on the screen center for 1,000 ms, followed by a 1,000 ms fixation interval. Next, the memory array was presented for 3,000 ms, followed by a 1,000 ms fixation interval. Participants were instructed to remember four individual actions while repeating the digits. They were required to judge whether the probe had appeared in the memory array, and whether the digit was one of the repeated digits, within 4,000 ms and 2,000 ms, respectively. In this example trial, the two pairs of actions in the memory array are interactive. The probed action appears in the memory array but the digit does not. ms = milliseconds. See the online article for the color version of this figure.

Neri et al., 2006; Troje & Westhoff, 2006). If the chunking effect is driven by low-level physical cues, it should not be affected by stimuli inversion. Otherwise, the chunking effect will be eliminated. Experiment 2 had 42 total participants (14 male, 28 female; 19–26 years old). One male and one female participant were removed because of low accuracy in the digit task.

The mean accuracy on the digit task was 0.96. The d' values for each memory condition are shown in Figure 2B. ANOVA revealed a significant main effect of memory condition, $F(2, 78) = 18.236$, $p < .001$, $\eta_p^2 = .319$, $BF_{10} = 1.836e + 6$. Post hoc comparisons revealed that the d' of the interactive pair was significantly higher than that of the noninteractive pair, $t(39) = 5.257$, $p < .001$, Cohen's $d = 1.250$, $BF_{10} = 238.0$, and individual four, $t(39) = 5.203$, $p < .001$, Cohen's $d = 1.237$, $BF_{10} = 2.067e + 4$. There was no significant difference between the latter two, $t(39) = -0.054$, $p = 1.000$, Cohen's $d = -0.013$, $BF_{10} = 0.171$.

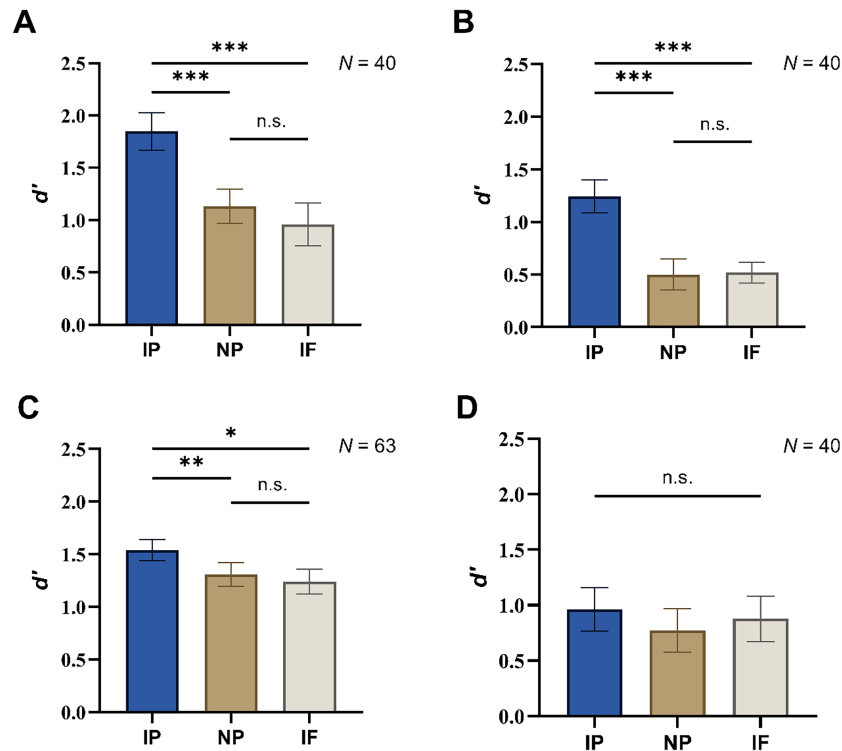
Experiment 2 found a similar chunking effect as Experiment 1 after stimuli inversion, implying that participants chunked individual actions basing on physical cues (e.g., similarity) embedded in physical-contingency interactions. We further compared the chunking effect (interactive pair minus noninteractive pair) between Experiments 1 and 2. Independent t test found that the chunking effect was comparable, $t(78) = -0.053$, $p = .958$, Cohen's $d = -0.012$, $BF_{10} = 0.171$, implying that the chunking effect in Experiment 1 was not a sum effect of physical and social relationship cues.

Experiment 3: Retaining Upright Communicative-Contingency Interactions

In Experiment 3, a total of 67 participants (23 male, 44 female; 17–27 years old) were asked to remember upright communicative-contingency actions. Two male participants were removed because of low accuracy in the digit task. One male and one female participant were removed because of chance-level memory accuracy.

The mean accuracy on the digit task was 0.97. The d' values for each memory condition are shown in Figure 2C. ANOVA revealed a significant main effect of memory condition, $F(2, 124) = 5.377$, $p = .006$, $\eta_p^2 = .080$, $BF_{10} = 5.542$. Post hoc comparisons revealed that the d' of the interactive pair was significantly higher than that of the noninteractive pair, $t(62) = 2.441$, $p = .048$, Cohen's $d = 0.354$, $BF_{10} = 3.464$, and individual four, $t(62) = 3.117$, $p = .007$, Cohen's $d = 0.452$, $BF_{10} = 11.744$. The latter two did not differ significantly, $t(62) = 0.676$, $p = 1.000$, Cohen's $d = 0.098$, $BF_{10} = 0.166$.

In line with Experiment 1, Experiment 3 showed that communicative-contingency interactions could be chunked in WM. Moreover, the memory performance between the noninteractive pair and individual four did not differ significantly, suggesting that facingness or spatial proximity cues alone cannot lead to the chunking of distinct actions in WM.

Figure 2*Results of Experiment 1 (A), Experiment 2 (B), Experiment 3 (C), and Experiment 4 (D)*

Note. Participants were shown upright and inverted physical-contingency interactions in Experiments 1 and 2, respectively. Participants were shown upright and inverted communicative-contingency interactions in Experiments 3 and 4, respectively. d' is presented for each condition. IP = interactive pair; NP = noninteractive pair; IF = individual four; CI = confidence interval. Error bars indicate 95% CI (Cousineau & O'Brien, 2014). Each line segment identifies the significance of post hoc contrasts. N represents the number of participants in each experiment. n.s. See the online article for the color version of this figure.

* $p < .05$. ** $p < .01$. *** $p < .001$. $p > .10$.

Experiment 4: Retaining Inverted Communicative-Contingency Interactions

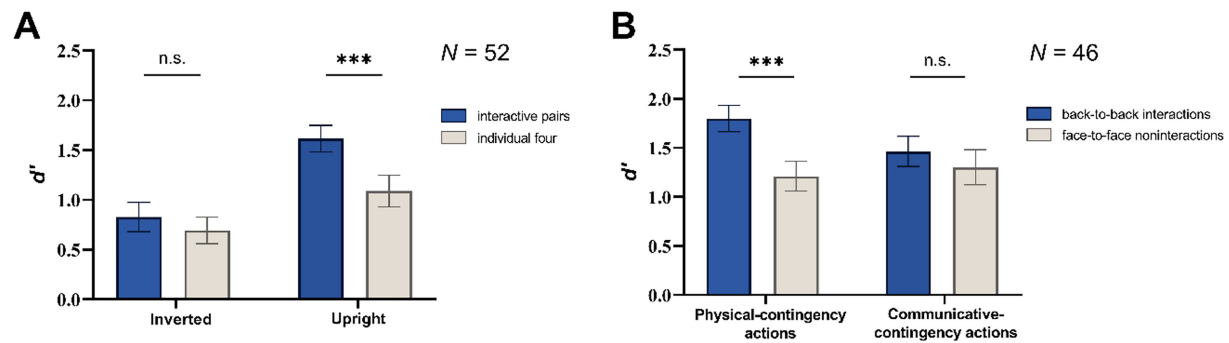
Experiment 4 inverted communicative-contingency actions to impair social relationship information while keeping physical cues intact. A total of 42 participants (14 male, 28 female; 18–29 years old) completed Experiment 4. One male participant was removed because of low accuracy in the digit task. One male participant was removed because of chance-level memory accuracy.

The mean accuracy on the digit task was 0.97. The d' values for each condition are shown in Figure 2D. ANOVA showed that the main effect of memory condition was not significant, $F(2, 78) = 0.631$, $p = .535$, $\eta_p^2 = .016$, $BF_{10} = 0.151$, suggesting that performance among three conditions was comparable. Experiment 4 found that the chunking effect was erased by inverting communicative-contingency interactions, with Bayesian values of less than one-fifth. This finding suggests that the chunking of communicative-contingency actions relies on social relationship cue.

Experiment 5: Validating the Findings of Experiments 3 and 4

To further substantiate the distinct mechanisms underlying Experiments 3 and 4, Experiment 5 examined the impact of action orientation on participants' ability to memorize interactive-pair and individual-four conditions with communicative-contingency actions. In order to minimize the potential influence of preexisting expectations or conceptualizations of social interaction that may arise from observing upright actions, participants were instructed to first memorize the inverted actions followed by the upright ones. A total of 58 participants (41 male, 17 female; 18–25 years old) completed Experiment 5. Two male participants were removed due to low accuracy in the digit task, and four male participants were removed due to below chance-level memory accuracy.

The mean accuracy on digit task was 0.95. The d' values for each condition are shown in Figure 3A. ANOVA showed that the main effect of action orientation was significant, $F(1, 51) = 48.757$, $p < .001$, $\eta_p^2 = .489$, $BF_{10} = 1.043e + 6$, indicating that the upright actions had a higher memory performance relative to

Figure 3*Results of Experiment 5 (A) and Experiment 6 (B)*

Note. d' is presented for each condition. Error bars indicate 95% CI (Cousineau & O'Brien, 2014). N represents the number of participants in each experiment. CI = confidence interval. n.s. See the online article for the color version of this figure.

*** $p < .001$. $p > .10$.

inverted ones. The main effect of memory condition was significant, $F(1, 51) = 13.431$, $p < .001$, $\eta_p^2 = .208$, $BF_{10} = 37.870$, suggesting that d' values for interaction pairs were better than that for individual four. The interaction between action orientation and memory condition was significant, $F(1, 51) = 7.400$, $p = .009$, $\eta_p^2 = .127$, $BF_{10} = 8.921$. Further tests revealed higher performance in the interactive pair than in the individual four condition for upright actions, $t(51) = 4.409$, $p < .001$, Cohen's $d = 0.611$, $BF_{10} = 401.710$. However, there was no significant difference between interactive pair and individual four for inverted actions, $t(51) = 1.208$, $p = .232$, Cohen's $d = 0.168$, $BF_{10} = 0.300$. These results replicate the key findings of Experiments 3 and 4 and suggest that inverting actions in Experiment 4 indeed disrupted the chunking process in Experiment 3.

Experiment 6: Retaining Back-to-Back Physical- and Communicative-Contingency Interactions

Physical gestalt cues cannot be removed while leaving social relationship cue intact for physical-contingency actions, complicating the testing of the role of physical cues alone. We therefore tested a direct prediction of chunking based on physical cues in Experiment 6. To reiterate, the involuntary chunking will take place for back-to-back physical-contingency interactions but vanish for back-to-back communicative-contingency interactions, because back-to-back manipulation can erase the social relationship but not physical cues. There were four randomly displayed memory conditions: back-to-back physical- or communicative-contingency interactions, and face-to-face noninteractive pairs constructed from physical- or communicative-contingency interactions. A total of 49 participants (25 male, 24 female; 18–29 years old) completed Experiment 6. Two male and one female participant were removed because of low accuracy in the digit task.

The mean accuracy on the digit task was 0.96. The d' values for each condition are shown in Figure 3B. Two-way ANOVA showed a significant main effect of memory condition, $F(1, 45) = 18.359$, $p < .001$, $\eta_p^2 = .290$, $BF_{10} = 88.498$, suggesting that d' for the back-to-back interactive pair was better than for the face-to-face noninteractive pair. The main effect of action type was not significant, $F(1, 45) = 1.825$, $p = .183$, $\eta_p^2 = .039$, $BF_{10} = 0.379$, suggesting that accuracy was comparable between the two action types. The interaction between

memory condition and action type was significant, $F(1, 45) = 5.641$, $p = .022$, $\eta_p^2 = .111$, $BF_{10} = 5.905$. Further tests revealed higher performance in the back-to-back interactive pair than in the face-to-face noninteractive pair for physical-contingency interactions, $t(45) = 5.303$, $p < .001$, Cohen's $d = 0.782$, $BF_{10} = 5,414.570$. However, there was no significant difference between back-to-back interactive pair and face-to-face noninteractive pair for communicative-contingency actions, $t(45) = 1.167$, $p = .249$, Cohen's $d = 0.172$, $BF_{10} = 0.302$. Additionally, there was no significant difference between face-to-face noninteractions under physical-contingency actions and the two conditions under the communicative-contingency actions, $F(2, 90) = 1.832$, $p = .166$, $\eta_p^2 = .039$, $BF_{10} = 0.358$. This finding further suggests that no chunking occurred under the communicative-contingency interaction conditions.

In line with our view that physical cues were crucial in physical-contingency interactions, we found a significant chunking effect for back-to-back physical-contingency actions compared to corresponding face-to-face noninteractive ones. Conversely, the chunking effect vanished in back-to-back communicative-contingency actions. These findings collectively imply that facingness does not play a role in physical-contingency interactions, but contributes to social relationship cues in communicative-contingency interactions (i.e., a prerequisite in chunking communicative-contingency interactions).

General Discussion

We for the first time demonstrated that both domain-general and domain-specific mechanisms contribute to the chunking of agent actions. In the context of physical-contingency interactions, physical regularities (i.e., gestalt law) function as fundamental visual chunking cues (as demonstrated in Experiments 1, 2, and 6) and exhibit accessibility both in upright and inverted interactions. Conversely, in the case of communicative-contingency interactions, social relationship plays a pivotal role as a social chunking cue (as evidenced in Experiments 3, 4, 5, and 6) and experiences impairment when the interaction is inverted. Moreover, facingness—although involved in communicative-contingency interactions (Experiment 6)—alone did not determine whether two dynamic agent actions could be chunked as a unit (Experiments 1–4 and 6). These findings suggest that the chunking mechanisms of dynamic social interactions in real life are distinct from those of static images of social interactions.

Dual Routes of Chunking Social Interactions: Visual Versus Social Chunking

A critical insight of the current study is the importance of distinguishing between physical- and communicative-contingency interactions, which is ignored in most of the previous chunking studies of social interaction. Moreover, studies tapping either physical- or communicative-contingency interaction implied that they shared similar processing properties (e.g., Manera et al., 2011; Neri et al., 2006; Su et al., 2016). The current study, for the first time, revealed distinct principles underlying the chunking process of physical- and communicative-contingency interactions. Therefore, when exploring the mechanisms of social interaction in the mind and their related influence over other cognitive activities, it is important to take this facet into consideration. Further, our study implies that the interactive actions in Ding et al. (2017) and Lu et al. (2022) largely reflected the properties of communicative contingency, as the WM chunking effect could be erased by stimuli inversion. It is worth noting that Burling and Lu (2018) suggested that global motion synchrony, spatial proximity between agents, and highly salient moments of interpersonal coordination are critical features impacting social interaction judgments. While we consider the two distinct types of social interactions meet this criterion, the three features may have distinct weights in forming physical- and communicative-contingency interactions; for instance, physical-contingency actions should have a higher weight on global motion synchrony than communicative-contingency actions.

The people-watching interdisciplinary is in its infancy, and a critical force driving its development is the social chunking phenomenon in the mind, which is considered to reflect the domain-specific processing of social interaction (Isik et al., 2017; Walbrin et al., 2018; see also Wurm & Caramazza, 2019; Wurm et al., 2017). However, most of the previous chunking effect is likely a byproduct of the domain-general orientation cueing effect rooted in static images (Adibpour et al., 2021; Paparella & Papeo, 2022; Vestner et al., 2020; Vestner, Over, Gray, & Cook, 2021; Vestner, Over, Gray, Tipper, & Cook, 2021). Therefore, revealing social chunking is crucial in future social interaction exploration. Using dynamic social interaction stimuli, we revealed clear-cut evidence favoring the existence of domain-specific social chunking for social interaction. We argue that future studies exploring the domain-specific properties of social interaction should adopt dynamic communicative-contingency interactions. On the other hand, we need to point out that Experiments 2 and 6 only confirmed the predictions of visual chunking for physical-contingency interactions. We did not falsify the visual chunking as it is difficult to remove the gestalt cue while keeping intact the social relationship cue. Future studies may consider finding indices that are sensitive only to gestalt cue, to strictly test the revealed visual chunking.

The enhanced memory performance observed in interacting dyads was not due to the alternative that participants only had to identify and remember two actions, as each dyad shared an action. Particularly, Ji et al. (2020) demonstrated empirical evidence that when participants observed interacting PLD dyads, an event-based processing advantage appeared automatically. When one action of an interacting dyad is cued, attention is automatically spread to the other one. This holistic processing of interacting dyads implied that social interaction serves as a fundamental unit of attention as early as at the perceptual stage. Therefore, if participants just picked up one action of interacting dyads to remember, extra operation

is needed to break the interactive unit, making it an inefficient approach. Moreover, the three memory conditions in Experiments 1–4 were presented randomly, with the majority of trials being non-interactive. Participants were explicitly instructed to memorize individual actions, ensuring consistency in processing strategies across trials. While our study focused on investigating the chunking mechanisms of social interaction in WM, we acknowledge that the observed effect may originate from perception. Consequently, future studies should explore whether these mechanisms extend to perception by eliminating the WM component.

Role of Facingness in Chunking Social Interaction

In recent decade, the escalating exploration of social interaction has commonly adopted face-to-face versus back-to-back manipulations in comparing the visuocognitive processing of interactive and noninteractive agents. Using dynamic actions to convey social interaction, we demonstrated that the facingness of actions was insufficient even in domain-general visual chunking (Experiments 1 and 6). Instead, participants chunked physical-contingency interactions basing on physical gestalt laws. These findings contrast with the view that facingness is critical in building the chunk of social interactions in mind. Meanwhile, we did not mean that facingness did not play a role in social interaction, since Experiment 6 suggested that facingness affected communicative-contingency interactions, and Zhou et al. (2019) supported the role of facingness in social interaction.

Instead of revealing a domain-general visual chunking, we found that both domain-general visual chunking and domain-specific social chunking occurred for interactive actions. It seems that our brain has evolved in a distinct yet elegant fashion in detecting and taking advantage of the grouping cues in dynamic stimuli. The chunking mechanisms underlying dynamic social interactions appear more complex than those underlying static interactions. The current findings emphasize the necessity of using dynamic stimuli to explore the cognitive and neural mechanisms of social interaction in real life (see also Schweinberger & Dobel, 2021). Indeed, evidence demonstrates that dynamic information is vital in chunking multimodal information from the face, body, and voice to achieve accurate recognition (see Yovel & O'Toole, 2016 for a review). Related, Thurman and Lu (2014) suggested that our early vision system detects animate agents in the environment basing on three fundamental constraints: consistency of movements with gravity (gravity constraint), congruence of limb movements in body-centered coordinates and body motion in environmental coordinates (motion congruency constraint), and the prototypical mammalian body plan of the global structural organization of the stimuli (biological structure constraint). Only by satisfying these basic constraints can stimuli be evaluated for high-level social content (e.g., meaningful social interaction, action understanding, and intentionality). The first two constraints are absent in static social interaction images, which may push the mind to adopt a different processing mechanism (e.g., domain-general orientation cueing effect).

More broadly, the current study emphasizes the importance of using real-world rather than artificial stimuli, in investigating the processing mechanisms of target information in the mind. Both behavior and brain functions have been demonstrated to differ between artificial and meaningful real-world stimuli (e.g., the current study; T. F. Brady et al., 2016, for reviews see T. F. Brady et al., 2019; Snow & Culham, 2021). Our finding contributes

new evidence suggesting that real-world stimuli contain an array of high-level information that constrains the attentional mechanism (Collegio et al., 2019; Henderson & Hayes, 2017). Meanwhile, it is important to recognize that although PLDs effectively capture the fundamental aspects of dynamic actions, they lack crucial details (e.g., texture and form cues) and thus represent a more abstract representation compared to real-world actions (Snow & Culham, 2021). Future research may consider leveraging more ecologically valid approaches, such as utilizing virtual reality environments, to investigate how people process social interactions from a third-party perspective.

Chunking in WM

Accumulating evidence has suggested that WM can group visual objects according to high-level semantic knowledge (e.g., affordance and statistical regularities; T. Brady et al., 2009, 2021; Huang & Awh, 2018; O'Donnell et al., 2018). However, the main body of evidence comes from static objects, few studies have addressed the chunking effect in higher representation nodes than objects, such as events (e.g., human actions). Social relationship is a type of typical semantic knowledge according to the WM literature (e.g., Ding et al., 2017; Paparella & Papeo, 2022). The current study hence showed that WM flexibly employs gestalt laws and semantic knowledge to chunk dynamic events. Moreover, this study provides clear evidence suggesting that social chunking can occur in WM.

The findings of Experiments 1, 2, and 6 suggest that there may be a hierarchy of employing chunking cues in WM. Although social relationship information is embedded in physical-contingency interactions (the participants could observe meaningful social interactions from them; see *Evaluation of Interactive Actions in the online supplemental materials*), participants refer to physical cues rather than social relationship. One possible explanation is that physical cues can be accessed directly from external visual input, while social relationship needs to be retrieved from long-term memory (but see Hafri & Firestone, 2021 for a different view); the former is quicker than the latter. In a similar spirit of hierarchical chunking, although Experiments 1, 2, and 6 imply that physical cue is accessed first relative to semantic cue, the findings of Experiments 3, 4, and 5 imply that the semantic cue social relationship is more powerful than the facingness cue in chunking stimuli. In particular, although the actions in the noninteractive pair were closer in space than those in the individual four condition, no chunking occurred. Moreover, the spatial gaps were comparable between interactive and noninteractive pair, yet the performance was much higher in interactive pair. It is of note that the two types of interactions in the current study were not purely based on physical or communicative contingencies due to technical constraints (see Footnote 1). Future research should employ strictly constructed stimuli to further validate the present findings.

The demonstration of two chunking mechanisms underlying social interaction casts critical light on clinical assessment and human–robot collaboration (HRC). For instance, abnormal processing of social interaction has been associated with social dysfunctions seen in autism spectrum disorder (e.g., von der Lühse et al., 2016) and schizophrenia (e.g., Okruszek, 2018; Okruszek et al., 2015). Future clinical studies may consider separately checking the processing mechanisms of physical- and communicative-contingency interactions. Moreover, a deeper understanding of the developmental

trajectories of the two types of social interactions may shed light on early markers of developmental delays and disorders. Finally, with the advances in robotic technology, research in HRC has gained in importance (Bütepage & Kragic, 2017; Henschel et al., 2020; Belhassein et al., 2022; Lemaignan et al., 2017). The current study could provide insight into the HRC in at least two aspects. First, currently, most of the HRC focuses on the communicative-contingency actions, ignoring the physical-contingency actions (Bütepage & Kragic, 2017; Belhassein et al., 2022). To enhance HRC, we argue that this more basic coordination between human and robot should be paid attention to in future design. Second, the task in the current study offers a direct way to objectively measure how well the HRC is performed from a third-person perspective, which is a key issue to ensure the social robot can be smoothly involved into human life.

Constraints on Generality Statement

Our findings provide evidence for the involuntary chunking of social interactions in WM. The stimuli consisted of video clips containing interactive actions. A large number of participants were randomly recruited from different universities in China. Thus, we expect the results to generalize to situations wherein participants view dynamic actions. Unpublished studies from our laboratory found similar results despite variations in the testing context (e.g., online or offline experiments). Consequently, we do not expect such variations to matter. We believe the results will be reproducible with participants from similar participant pools. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

References

- Abassi, E., & Papeo, L. (2020). The representation of two-body shapes in the human visual cortex. *The Journal of Neuroscience*, 40(4), 852–863. <https://doi.org/10.1523/JNEUROSCI.1378-19.2019>
- Adibpour, P., Hochmann, J. R., & Papeo, L. (2021). Spatial relations trigger visual binding of people. *Journal of Cognitive Neuroscience*, 33(7), 1343–1353. https://doi.org/10.1162/jocn_a_01724
- Baddeley, A. D. (1986). *Working memory* (Vol. 11). Clarendon Press.
- Belhassein, K., Fernández-Castro, V., Mayima, A., Clodic, A., Pacherie, E., Guidetti, M., & Alami, R. (2022). Addressing joint action challenges in HRI: Insights from psychology and philosophy. *Acta Psychologica*, 222, Article 103476. <https://doi.org/10.1016/j.actpsy.2021.103476>
- Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* (pp. 401–432). Editions de la Maison des Sciences de l'Homme, Cambridge University Press.
- Bietti, L. M., & Sutton, J. (2015). Interacting to remember at multiple time-scales: Coordination, collaboration, cooperation and culture in joint remembering. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, 16(3), 419–450. <https://doi.org/10.1075/is.16.3.04bie>
- Brady, T., Allen, M., & DeStefano, I. (2021). Chunking is not all-or-none: Hierarchical representations preserve perceptual detail within chunks. *Journal of Vision* (Charlottesville, VA.), 21(9), Article 2312. <https://doi.org/10.1167/jov.21.9.2312>
- Brady, T., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, 138(4), 487–502. <https://doi.org/10.1037/a0016797>
- Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for

- simple stimuli. *Proceedings of the National Academy of Sciences*, 113(27), 7459–7464. <https://doi.org/10.1073/pnas.1520027113>
- Brady, T. F., Störmer, V. S., Shafer-Skelton, A., Williams, J. R., Chapman, A. F., & Schill, H. M. (2019). Chapter Two—Scaling up visual attention and visual working memory to the real world. In K. D. Federmeier & D. M. Beck (Eds.), *Psychology of learning and motivation* (Vol. 70, pp. 29–69). Academic Press. <https://www.sciencedirect.com/science/article/pii/S0079742119300039>
- Bryant, G. A., Fessler, D. M. T., Fusaroli, R., Clint, E., Aarpe, L., Apicella, C. L., Petersen, M. B., Bickham, S. T., Bolyanatz, A., Chavez, B., De Smet, D., Díaz, C., Fančovičová, J., Fux, M., Giraldo-Perez, P., Hu, A., Kamble, S. V., Kameda, T., Li, N. P., ... Zhou, Y. (2016). Detecting affiliation in colughter across 24 societies. *Proceedings of the National Academy of Sciences—PNAS*, 113(17), 4682–4687. <https://doi.org/10.1073/pnas.1524993113>
- Burling, J. M., & Lu, H. (2018). Categorizing coordination from the perception of joint actions. *Attention, Perception, & Psychophysics*, 80(1), 7–13. <https://doi.org/10.3758/s13414-017-1450-2>
- Bütepage, J., & Kragic, D. (2017). Human-robot collaboration: From psychology to social robotics. *arXiv*, <https://doi.org/10.48550/arXiv.1705.10146>
- Capozzi, F., & Ristic, J. (2018). How attention gates social interactions. *Annals of the New York Academy of Sciences*, 1426(1), 179–198. <https://doi.org/10.1111/nyas.13854>
- Centelles, L., Assaiante, C., Nazarian, B., Anton, J.-L., & Schmitz, C. (2011). Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: A neuroimaging study. *PLoS ONE*, 6(1), Article e15749. <https://doi.org/10.1371/journal.pone.0015749>
- Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: The limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology* (2006), 62(7), 1420–1429. <https://doi.org/10.1080/17470210802453977>
- Cheng, Y., Liu, W., Yuan, X., & Jiang, Y. (2021). The eyes have it: Perception of social interaction unfolds through pupil dilation. *Neuroscience Bulletin*, 37(11), 1595–1598. <https://doi.org/10.1007/s12264-021-00739-z>
- Cheng, Y., Yuan, X., & Jiang, Y. (2021). The cognitive characteristics of and the brain mechanisms underlying social interaction processing from a third-person perspective. *Advances in Psychological Science*, 29(3), 472–480. <https://doi.org/10.3724/SP.J.1042.2021.00472>
- Collegio, A. J., Nah, J. C., Scotti, P. S., & Shomstein, S. (2019). Attention scales according to inferred real-world object size. *Nature Human Behaviour*, 3(1), 40–47. <https://doi.org/10.1038/s41562-018-0485-2>
- Cousineau, D., & O'Brien, F. (2014). Error bars in within-subject designs: A comment on Baguley (2012). *Behavior Research Methods*, 46(4), 1149–1151. <https://doi.org/10.3758/s13428-013-0441-z>
- Ding, X., Gao, Z., & Shen, M. (2017). Two equals one: Two human actions during social interaction are grouped as one unit in working memory. *Psychological Science*, 28(9), 1311–1320. <https://doi.org/10.1177/0956797617707318>
- Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. *Science (American Association for the Advancement of Science)*, 317(5843), 1344–1347. <https://doi.org/10.1126/science.1145463>
- Fedorov, L. A., Chang, D.-S., Giese, M. A., Bülthoff, H. H., & De la Rosa, S. (2018). Adaptation aftereffects reveal representations for encoding of contingent social actions. *Proceedings of the National Academy of Sciences—PNAS*, 115(29), 7515–7520. <https://doi.org/10.1073/pnas.1801364115>
- Feng, C., Eickhoff, S. B., Li, T., Wang, L., Becker, B., Camilleri, J. A., Hétu, S., & Luo, Y. (2021). Common brain networks underlying human social interactions: Evidence from large-scale neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, 126, 289–303. <https://doi.org/10.1016/j.neubiorev.2021.03.025>
- Gao, Z., Bentin, S., & Shen, M. (2015). Rehearsing biological motion in working memory: An EEG study. *Journal of Cognitive Neuroscience*, 27(1), 198–209. https://doi.org/10.1162/jocn_a_00687
- Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research (Oxford)*, 41(10–11), 1475–1482. [https://doi.org/10.1016/S0042-6989\(00\)00317-5](https://doi.org/10.1016/S0042-6989(00)00317-5)
- Gu, Q., Li, W., Lu, X., Chen, H., Shen, M., & Gao, Z. (2019). Agent identity drives adaptive encoding of biological motion into working memory. *Journal of Vision*, 19(14), Article 6. <https://doi.org/10.1167/19.14.6>
- Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*, 25(6), 475–492. <https://doi.org/10.1016/j.tics.2021.01.006>
- Hari, R., Henriksson, L., Malinen, S., & Parkkonen, L. (2015). Centrality of social interaction in human brain function. *Neuron*, 88(1), 181–193. <https://doi.org/10.1016/j.neuron.2015.09.022>
- Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, 1(10), 743–747. <https://doi.org/10.1038/s41562-017-0208-0>
- Henschel, A., Hortensius, R., & Cross, E. S. (2020). Social cognition in the age of human–robot interaction. *Trends in Neurosciences*, 43(6), 373–384. <https://doi.org/10.1016/j.tins.2020.03.013>
- Huang, L., & Awh, E. (2018). Chunking in working memory via content-free labels. *Scientific Reports*, 8(1), Article 23. <https://doi.org/10.1038/s41598-017-18157-5>
- Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences—PNAS*, 114(43), E9145–E9152. <https://doi.org/10.1073/pnas.1714471114>
- Isik, L., Mynick, A., Pantazis, D., & Kanwisher, N. (2020). The speed of human social interaction perception. *NeuroImage*, 215, Article 116844. <https://doi.org/10.1016/j.neuroimage.2020.116844>
- Jeffreys, H., Sir. (1961). *Theory of probability* (3rd ed.). Clarendon Press.
- Ji, H., Yin, J., Huang, Y., & Ding, X. (2020). Selective attention operates on the group level for interactive biological motion. *Journal of Experimental Psychology: Human Perception and Performance*, 46(12), 1434–1442. <https://doi.org/10.1037/xhp0000866>
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2), 201–211. <https://doi.org/10.3758/BF03212378>
- Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., & Alami, R. (2017). Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence*, 247, 45–69. <https://doi.org/10.1016/j.artint.2016.07.002>
- Liu, R., Yuan, X., Chen, K., Jiang, Y., & Zhou, W. (2018). Perception of social interaction compresses subjective duration in an oxytocin-dependent manner. *eLife*, 7, Article e32100. <https://doi.org/10.7554/eLife.32100>
- Lu, X., Dai, A., Guo, Y., Shen, M., & Gao, Z. (2022). Is the social chunking of agent actions in working memory resource-demanding? *Cognition*, 229, Article 105249. <https://doi.org/10.1016/j.cognition.2022.105249>
- Manera, V., Becchio, C., Schouten, B., Bara, B. G., & Verfaillie, K. (2011). Communicative interactions improve visual detection of biological motion. *PLoS ONE*, 6(1), Article e14594. <https://doi.org/10.1371/journal.pone.0014594>
- McMahon, E., Bonner, M. F., & Isik, L. (2023). Hierarchical organization of social action features along the lateral visual pathway. *Current Biology*, 33(23), 5035–5047. <https://doi.org/10.1016/j.cub.2023.10.015>
- Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, 9(9), 1186–1192. <https://doi.org/10.1038/nn1759>
- Nessler, J. A., & Gilliland, S. J. (2009). Interpersonal synchronization during side by side treadmill walking is influenced by leg length differential and altered sensory feedback. *Human Movement Science*, 28(6), 772–785. <https://doi.org/10.1016/j.humov.2009.04.007>
- O'Donnell, R. E., Clement, A., & Brockmole, J. R. (2018). Semantic and functional relationships among objects increase the capacity of visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(7), 1151–1158. <https://doi.org/10.1037/xlm0000508>

- Okruszek, L. (2018). It is not just in faces! Processing of emotion and intention from biological motion in psychiatric disorders. *Frontiers in Human Neuroscience*, 12, Article 48. <https://doi.org/10.3389/fnhum.2018.00048>
- Okruszek, L., & Chrustowicz, M. (2020). Social perception and interaction database—A novel tool to study social cognitive processes with point-light displays. *Frontiers in Psychiatry*, 11, Article 123. <https://doi.org/10.3389/fpsy.2020.00123>
- Okruszek, L., Haman, M., Kalinowski, K., Talarowska, M., Becchio, C., & Manera, V. (2015). Impaired recognition of communicative interactions from biological motion in schizophrenia. *PLoS ONE*, 10(2), Article e0116793. <https://doi.org/10.1371/journal.pone.0116793>
- Okruszek, L., Wordecha, M., Jarkiewicz, M., Kossowski, B., Lee, J., & Marchewka, A. (2018). Brain correlates of recognition of communicative interactions from biological motion in schizophrenia. *Psychological Medicine*, 48(11), 1862–1871. <https://doi.org/10.1017/S0033291717003385>
- Paparella, I., & Papeo, L. (2022). Chunking by social relationship in working memory. *Visual Cognition*, 30(5), 354–370. <https://doi.org/10.1080/13506285.2022.2064950>
- Papeo, L. (2020). Twos in human visual perception. *Cortex*, 132, 473–478. <https://doi.org/10.1016/j.cortex.2020.06.005>
- Papeo, L., Goupil, N., & Soto-Faraco, S. (2019). Visual search for people among people. *Psychological Science*, 30(10), 1483–1496. <https://doi.org/10.1177/0956797619867295>
- Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The two-body inversion effect. *Psychological Science*, 28(3), 369–379. <https://doi.org/10.1177/0956797616685769>
- Quadflieg, S., Gentile, F., & Rossion, B. (2015). The neural basis of perceiving person interactions. *Cortex*, 70, 5–20. <https://doi.org/10.1016/j.cortex.2014.12.020>
- Quadflieg, S., & Koldewyn, K. (2017). The neuroscience of people watching: How the human brain makes sense of other people's encounters. *Annals of the New York Academy of Sciences*, 1396(1), 166–182. <https://doi.org/10.1111/nyas.13331>
- Quadflieg, S., & Penton-Voak, I. S. (2017). The emerging science of people-watching: Forming impressions from third-party encounters. *Current Directions in Psychological Science: A Journal of the American Psychological Society*, 26(4), 383–389. <https://doi.org/10.1177/0963721417694353>
- Quadflieg, S., & Westmoreland, K. (2019). Making sense of other people's encounters: Towards an integrative model of relational impression formation. *Journal of Nonverbal Behavior*, 43(2), 233–256. <https://doi.org/10.1007/s10919-019-00295-1>
- Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, 22(2), 322–339. <https://doi.org/10.1037/met0000061>
- Schweinberger, S. R., & Dobel, C. (2021). Why twos in human visual perception? A possible role of prediction from dynamic synchronization in interaction. *Cortex*, 135, 355–357. <https://doi.org/10.1016/j.cortex.2020.09.015>
- Sliwa, J., & Freiwald, W. A. (2017). A dedicated network for social interaction processing in the primate brain. *Science (American Association for the Advancement of Science)*, 356(6339), 745–749. <https://doi.org/10.1126/science.aam6383>
- Snow, J. C., & Culham, J. C. (2021). The treachery of images: How realism influences brain and behavior. *Trends in Cognitive Sciences*, 25(6), 506–519. <https://doi.org/10.1016/j.tics.2021.02.008>
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–149. <https://doi.org/10.3758/BF03207704>
- Su, J., Van Boxtel, J. J. A., & Lu, H. (2016). Social interactions receive priority to conscious perception. *PLoS ONE*, 11(8), Article e0160468. <https://doi.org/10.1371/journal.pone.0160468>
- Thurman, S. M., & Lu, H. (2014). Perception of social interactions for spatially scrambled biological motion. *PLoS ONE*, 9(11), Article e112539. <https://doi.org/10.1371/journal.pone.0112539>
- Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a “life detector”? *Current Biology*, 16(8), 821–824. <https://doi.org/10.1016/j.cub.2006.03.022>
- Vestner, T., Flavell, J. C., Cook, R., & Tipper, S. P. (2021). Remembered together: Social interaction facilitates retrieval while reducing individuation of features within bound representations. *Quarterly Journal of Experimental Psychology*, 75(9), 1593–1602. <https://doi.org/10.1177/17470218211056499>
- Vestner, T., Gray, K. L. H., & Cook, R. (2020). Why are social interactions found quickly in visual search tasks? *Cognition*, 200, Article 104270. <https://doi.org/10.1016/j.cognition.2020.104270>
- Vestner, T., Gray, K. L. H., & Cook, R. (2021). Visual search for facing and non-facing people: The effect of actor inversion. *Cognition*, 208, Article 104550. <https://doi.org/10.1016/j.cognition.2020.104550>
- Vestner, T., Over, H., Gray, K. L. H., & Cook, R. (2021). Objects that direct visuospatial attention produce the search advantage for facing dyads. *Journal of Experimental Psychology: General*, 151(1), 161–171. <https://doi.org/10.1037/xge0001067>
- Vestner, T., Over, H., Gray, K. L. H., Tipper, S. P., & Cook, R. (2021). Searching for people: Non-facing distractor pairs hinder the visual search of social scenes more than facing distractor pairs. *Cognition*, 214, Article 104737. <https://doi.org/10.1016/j.cognition.2021.104737>
- Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S.-A. (2019). Bound together: Social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *Journal of Experimental Psychology: General*, 148(7), 1251–1268. <https://doi.org/10.1037/xge0000545>
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 92–114. <https://doi.org/10.1037/0096-1523.27.1.92>
- von der Lühse, T., Manera, V., Barisic, I., Becchio, C., Vogeley, K., & Schilbach, L. (2016). Interpersonal predictive coding, not action perception, is impaired in autism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1693), Article 20150373. <https://doi.org/10.1098/rstb.2015.0373>
- Walbrin, J., Downing, P., & Koldewyn, K. (2018). Neural responses to visually observed social interactions. *Neuropsychologia*, 112, 31–39. <https://doi.org/10.1016/j.neuropsychologia.2018.02.023>
- Walbrin, J., & Koldewyn, K. (2019). Dyadic interaction processing in the posterior temporal cortex. *NeuroImage*, 198, 296–302. <https://doi.org/10.1016/j.neuroimage.2019.05.027>
- Wurm, M. F., & Caramazza, A. (2019). Lateral occipitotemporal cortex encodes perceptual components of social actions rather than abstract representations of sociality. *NeuroImage (Orlando, Fla.)*, 202(Journal Article), Article 116153. <https://doi.org/10.1016/j.neuroimage.2019.116153>
- Wurm, M. F., Caramazza, A., & Lingnau, A. (2017). Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *The Journal of Neuroscience*, 37(3), 562–575. <https://doi.org/10.1523/JNEUROSCI.1717-16.2016>
- Yovel, G., & O'Toole, A. J. (2016). Recognizing people in motion. *Trends in Cognitive Sciences*, 20(5), 383–395. <https://doi.org/10.1016/j.tics.2016.02.005>
- Zhou, C., Han, M., Liang, Q., Hu, Y.-F., & Kuai, S.-G. (2019). A social interaction field model accurately identifies static and dynamic social groupings. *Nature Human Behaviour*, 3(8), 847–855. <https://doi.org/10.1038/s41562-019-0618-2>

Received December 6, 2022

Revision received November 10, 2023

Accepted November 29, 2023 ■