Chunking by Social Relationship in Working Memory

Ilenia Paparella and Liuba Papeo

Institut des Sciences Cognitives— Marc Jeannerod, UMR5229, Centre National de la Recherche Scientifique (CNRS) & Université Claude Bernard Lyon1, France

Author Note

Ilenia Paparella https://orcid.org/0000-0002-7683-2503

Liuba Papeo https://orcid.org/0000-0003-3056-8679

Correspondence concerning this article should be addressed to Ilenia Paparella, now at GIGA-Research, Cyclotron Research Center-In Vivo Imaging Unit, 8 allée du Six Août, Batiment B30, University of Liège, 4000 Liège (Belgium).

Email: ipaparella@uliege.be

Abstract

Working memory (WM) uses knowledge and relations to organize and store multiple items in fewer structured units, or chunks. We investigated: *a)* whether a crowd that exceeds the WM capacity is retained better if individuals can be grouped in *social chunks*; and *b)* what counts as a social chunk: two individuals involved in a meaningful interaction or just spatially close and face-to-face. In a delayed change-detection task, participants were more accurate in reporting changes in arrays involving facing (*vs.* non-facing) dyads whether they depicted meaningful interactions or not (Experiment 1, 2 and 4). This advantage survived a secondary task that increased WM load, only when facing dyads formed meaningful interactions (Experiment 3). Thus, WM uses representation of interaction to chunk crowds in social groups. The mere face-to-face positioning is sufficient to trigger social chunking, although without a semantic anchor this process is fainter and more susceptible to interference.

Keywords: working memory, chunking, perceptual grouping, social working memory, scene perception

Introduction (1080)

Living in a social world requires humans to process information about conspecifics and the relationships between them. In scenarios that feature multiple faces or bodies, such as an urban scene, vision exploits markers of interpersonal involvement to detect and recognize social groups –i.e., people who engage in social relationship. One of such markers is the relative positioning of bodies in space: nearby bodies in a face-to-face configuration are more likely to be interpreted as interacting than bodies in other spatial configurations (Zhou et al., 2019); they are more likely to be attended to in a crowd (Papeo et al., 2019; Vestner et al., 2019, 2020), and to break into visual awareness under low-visibility conditions (Papeo et al., 2017). Visual efficiency has been explained by grouping, that is, the processing of multiple bodies as a single perceptual/attentional unit, promoted by visuo-spatial cues of interaction such as spatial proximity and face-to-face positioning (Papeo, 2020).

Here, we asked whether the advantage of grouping people by virtue of socially relevant spatial relations extends beyond visual perception. A system that may benefit from the representation of relationship between social agents is working memory (WM). WM supports the temporary storage of a limited amount of information for further cognitive operations (Ardila, 2003; Baddeley, 2000, 2003; Baddeley & Logie, 1999). The limits of WM capacity, corresponding to about four items (Anderson, et. al., 2015; Gao et al., 2015; Shen et al., 2014; Wood, 2008) can be exceeded through chunking, the process of binding and storing multiple items into a single unit (Cowan, 2000; Mathy & Feldman, 2012; Miller, 1956). Chunking in WM exploits a variety of cues, from perceptual similarity and low-level perceptual features to semantic relatedness, and statistical regularities (Brady, et al., 2009; Brady & Tenenbaum, 2013; Hollingworth, 2007; Kaiser, et al., 2014; Luck & Vogel, 1997; O'Donnell, et al., 2018).

Recent findings suggest that social relationship may be an effective principle of chunking in WM. It has been shown that infants as young as 16 months rely on knowledge about social relations to chunk sets of dolls in social units (Stahl & Feigenson, 2014). In particular, after seeing dolls interacting in pairs, infants were capable of remembering two pairs, i.e., four dolls, which exceeded the three-item limit of their WM. In another study on adults, it has been reported that, presented in arrays of four bodies, body movements performed as part of a meaningful dyadic interaction were more likely to be recognized in a short-delayed recognition task, relative to movements performed by isolated agents (Ding et al., 2017). In the authors' interpretation, movements that gave rise to interaction were chunked and stored as a single unit, thus increasing WM efficiency.

The above effects have been interpreted as the result of embedding individuals into the representation of a meaningful social interaction. But, can socially relevant spatial cues (e.g., spatial proximity and face-to-face positioning) alone, in the absence of familiar, meaningful interaction, trigger chunking of bodies in WM?

In visual perception and attention, effects of grouping have been found for bodies postures oriented toward one another without necessarily representing a meaningful, coherent interaction (Papeo et al., 2019, 2017). This circumstance raises the possibility that face-to-face positioning –i.e., the mutual perceptual accessibility of two bodies– is sufficient on its own to trigger the representation of interaction that binds two bodies together. In other words, it is possible that individuals represent the face-to-face positioning of bodies as an intrinsically meaningful relation that would yield a WM advantage, irrespective of whether the two facing bodies realize a familiar, coherent interaction.

We addressed this hypothesis in four experiments on female and male human adults, using a delayed change-detection task. The task was adapted from Kaiser, et al. (2015), who used it to show a benefit to the WM capacity, for multiple objects in spatial relations that respected real-world regularities (e.g., a lamp above a table, rather than a lamp below a table). In our version of the task, participants saw static arrays of four or six bodies, which approached or exceeded the WM capacity, arranged in two or three face-to-face dyads (facing arrays) or back-to-back dyads (non-facing arrays). Facing pairs could give rise to a coherent, familiar interaction (meaningful set –MF), or not (meaningless set –ML).

In Experiments 1-2 we asked: *are facing arrays remembered better than non-facing arrays? And, if so, is this advantage afforded by only meaningful interactions, or could it be* *found for any face-to-face body dyad?* We reasoned that, since the WM capacity in adults is of four items (Anderson et al., 2015; Cowan, 2000; Gao et al., 2015; Shen et al., 2014; Wood, 2008), the advantage of chunking by perceived social relationship could emerge already with arrays of four bodies and, certainly so, with arrays of six bodies, which exceed the WM capacity. We tested so with arrays composed of MF facing (*vs.* non-facing) dyads (Experiment 1) or of ML facing (vs. non-facing) dyads (Experiment 2). To assess whether relational cues favored chunking, performance on MF- and ML-facing arrays was compared with performance with the corresponding non-facing arrays involving the very same bodies.

Since Experiments 1-2 revealed a WM advantage for facing, over non-facing arrays, in Experiments 3-4, we asked whether WM represents MF and ML facing arrays in the same way. More precisely, we asked whether a semantically specified relation could provide an anchor point that would make the representation of MF-facing dyads stronger and therefore less subjected to interference in WM. To test so, in Experiments 3, all participants performed the delayed change-detection task on both MF and ML arrays, while performing a concurrent shadowing task (continuous word repetition), to increase WM load. On our reasoning, this condition could expose differences in the processing of MF-facing and ML-facing (*vs.* non-facing) arrays in WM, related to differences in the underlying relation (meaningful/familiar or not). In Experiment 4, we repeated the design of Experiment 3 without verbal shadowing. We aimed to confirm that, if in Experiment 3 a difference was found between MF-facing and ML-facing (*vs.* non-facing) arrays, it was actually due to the introduction of the secondary task.

In summary, with this study, we sought to evaluate the effect of body positioning (facing *vs.* non-facing) and the effect of representing a familiar, semantically specified interaction, in promoting chunking by social relationship in WM. Given that task performance depended on the possibility to structure a crowd in social (multiple-person) units, the results of this study shed light on what counts as a social unit in WM.

Experiment 1 (28)

Experiment 1 tested the participants' performance in detecting a change in a crowded array, where bodies formed facing or non-facing dyads. All facing dyads depicted coherent, meaningful interactions.

Participants (132)

Twenty healthy adults (18 females; mean age 22.8 ± 3.9 standard deviation, *SD*) participated in Experiment 1 as paid volunteers. All had normal or corrected-to-normal vision, reported no history of neurological or psychiatric disease and no consumption of psychoactive substances or medications. Participants gave informed consent prior to participation in the study. Experiments 1 was exploratory with respect to the sample size. With a sample size of 20, sensitivity analysis (G*Power 3.1; Erdfelder, et al., 2009) estimated a medium to large minimum detectable effect (i.e., the smallest true effect, which would be statistically significant with *alpha* = 0.05, and power = 0.80) of $\eta_p^2 = 0.10$ for the effect of positioning (facing *vs.* non-facing arrays). The local ethics committee (Comité de Protection des Personnes, CPP SUD-EST II, IRB: 00009118) approved this study.

Stimuli (1186)

Meaningful-interaction (MF) dyads. We created gray-scale images of a human body in 48 different poses in lateral view, using Daz3D (Daz Productions, Salt Lake City, UT) and the Image Processing Toolbox in MATLAB (The MathWorks, Natick, MA). Body poses were compatible with one of six types of social interaction: communicating, talking, dancing, fighting, quarreling and waving goodbye. Single bodies were paired and positioned face-to-face, so to depict one of the six aforementioned interactions, yielding a total of 24 interacting dyads. The meaningfulness of each interaction was evaluated in a rating study involving an independent sample of participants (see below).

Meaningless-interaction (ML) dyads: Forty-eight images of a human body in 48 new poses in lateral view were created as above, and then randomly paired in 24 facing dyads, so that the pairing gave rise to non-familiar, meaningless interactions.

Rating study. The meaningfulness of the above dyads was evaluated with a rating study involving 19 native-French speakers (11 females, mean age 27.5 ± 6.8 *SD*) external to the main study. All participants saw all the MF and ML facing dyads in random order. Dyads appeared at the center of a computer screen subtending a visual angle of ~10°. For each dyad, participants were asked to rate the meaningfulness of the scene by clicking on a Likert scale from 0 to 10 (0 = meaningless; 10 = very meaningful) displayed under the dyad. After a blank, participants were instructed to provide a verbal description of the stimulus using one or a few words. There was no time limit to respond. The study was conducted online using Google Forms.

For each dyad, we computed: *a)* a score of meaningfulness corresponding to the mean rate across participants; and *b)* a score of semantic consistency, representing the percentage of participants who agreed on the expected meaning of a stimulus (descriptions with similar meaning were considered semantically consistent; for example, “parler”, *to talk,* “argumentée un point”, *to make a point,* were taken as descriptions compatible with the general meaning “talking”).

The results of this study confirmed our *a priori* categorization of the dyads as MF or ML. In particular, dyads that we had categorized as MF had high meaningfulness scores (mean = 7.71 ± 1.14) and high semantic consistency (mean = 84.8 ± 17.06). Dyads that we had categorized as ML had low meaningfulness scores (mean = 4.9 ± 0.92) and low consistency (mean = 32.5 ± 15.95). The two sets differed significantly for both meaningfulness, $t(23) = 9.54$, $p < .001$, Cohen's $d = 2.71$, and consistency, $t(23) = 9.97$, $p < .001$, Cohen's $d = 3.16$.

Next, we created two sets of stimuli for the main experiments. For the MF set, we selected three exemplars for each of the four categories that had obtained the highest values of meaningfulness and semantic consistency: “Talking”, “Dancing”, “Fighting” and “Quarreling” (mean meaningfulness = 7.67 ± 1.35; mean consistency = 94.75 ± 10.39). For each of the selected dyads, we created a mirror version, yielding a total of 24 meaningful dyads. Twenty-four non-facing dyads were created by swapping the position of the two figures in each facing dyad (i.e., the figure on the left side was moved to the right side and *vice versa*). The distance between two bodies in a dyad was matched across facing and non-facing stimuli. To this end, we considered: (i) the distance between the centers of the two minimal bounding boxes around each body (facing *vs.* non-facing dyads: $t(23) = 1.04$, $p = .308$, Cohen's $d = .13$); (ii) the distance between the closest points of the two bodies (facing *vs.* non-facing dyads: $t(23) = 1.16$, $p = .257$, Cohen's $d = .01$); and (iii) the center of mass (facing *vs.* non-facing dyads: $t(23) = .13$, $p = .893$, Cohen's $d < .01$). Thus, facing and non-facing dyads based on the MF set involved the very same bodies at matched distances, and only differed for the relative spatial positioning.

For the ML set, 12 dyads were randomly selected amongst the dyads with the lowest values of meaningfulness and consistency. Those dyads were flipped on the horizontal axis, yielding a total of 24 meaningless dyads. The positioning of the two bodies within each dyad was swapped to obtain 24 non-facing dyads. Across ML facing and non-facing stimuli, we matched distances between: (i) the centers of the two bounding boxes around each body, $t(23) = .14$, $p = .885$, Cohen's $d < .01$; (ii) the closest extremities of bodies, $t(23) = .25$, $p = .802$, Cohen's $d < .01$; and (iii) the centers of mass, $t(23) = .87$, $p = .391$, Cohen's $d < .02$.

The MF and ML set of dyads were also matched in terms of center of mass (facing dyads: $t(23) = .41$, $p = .683$, Cohen's $d = .12$; non-facing dyads: $t(23) = 1.07$, $p = .295$, Cohen's $d = .12$), distance between the closest points of the two bodies (facing dyads: $t(23) = 1.38$, $p = .178$, Cohen's $d = .28$; non-facing dyads: $t(23) = 1.47$, $p = .152$, Cohen's $d = .29$), and distance between the centers of the two bounding boxes containing the bodies (facing dyads: $t(23) = .93$, $p = .360$, Cohen's $d = .20$; non-facing dyads: $t(23) = .77$, $p = .447$, Cohen's $d = .08$). In Experiment 1, we used the MF set. The ML set will be considered in Experiment 2.

Arrays. We created arrays featuring two (set2; 50% of arrays) or three (set3) facing dyads (Figure 1B). In set2-arrays, dyads were placed on the right and left side, equally distant from a cross in central fixation; in set3, dyads were placed in correspondence of the three angles of a (invisible) triangle centered on the fixation. Individual dyads subtended ~3° of visual angle and their center was far ~2° from the central fixation cross. The distance between two bodies in a dyad was about one third of the distance between two different dyads, making spatial proximity the first spatial cue to chunk the crowd in dyads –for both facing and non-facing arrays. In each array, dyads were all facing (50%) or all non-facing.

In a facing array, facing dyads belonged to different semantic categories. For each facing array, we created a non-facing array involving the very same dyads with bodies presented back-to-back. Each array (sample) was paired with another array (probe) that could be identical (same trials; 50%) or differ from the sample for one dyad of a category not shown in the sample array (different trials). For example, in a different trial, if the sample-array showed exemplars of “Fighting” and “Talking”, the probe-array would show the same “Fighting” dyad and a new dyad chosen from the remaining two categories (“Quarreling” or “Dancing”). For each participant, a new set of 432 arrays was created (160 sample-arrays and corresponding same/different probe-arrays for the main experiment; 24 and 32 sample-arrays and corresponding same/different probe-arrays for familiarization and training, respectively), equally distributed across the eight experimental conditions: set2 and set3 of facing and non-facing same and different arrays (see Figure 1B-C for examples of facing and non-facing arrays of MF and ML dyads).

Procedure (297)

Participants were seated on a height-adjustable chair in front of a screen for stimulus presentation, with their eyes aligned to, and 60 cm away from the center of the screen. Each trial began with a central fixation cross presented for 1400 ms, followed by the sample-array shown for 2000 ms. After an interval of 1000 ms, the probe-array appeared for 2000 ms (Figure 1A). For each trial, participants were asked to report whether the probe was the same or different relative to the sample. Using the numeric keypad of a keyboard in front of them, they had to press “1” for *same* and “2” for *different*, with the right index and middle finger, respectively. They were encouraged to respond as fast and accurately as possible, from the onset of the probe array. Participants performed five blocks of 32 trials, four for each experimental condition, yielding 20 trials for each of the eight conditions (set2 or set3, facing or non-facing arrays in same or different trials) and a total of 160 trials. Stimuli of the eight conditions were randomly interleaved in a block. Every two blocks, participants were invited to take a break. The experiment was preceded by a familiarization block of 24 trials, during which participants were free to ask questions and address the experimenter, and a training block of 32 trials, identical to the proper experiment. Response accuracy and reaction times (RTs) were recorded for each trial. Images were displayed on a 17-in CRT monitor (1024x768 pixels resolution, 85-Hz refresh rate). Stimulus presentation and response collection were controlled using the Psychophysics Toolbox extensions (Brainard, 1997) through MATLAB. The entire experiment lasted ~30 minutes. In the debriefing at the end of the experiment, participants were asked about the strategies they may have used to complete the task.

Results (906)

For each participant, for each condition, performance was analyzed in terms of proportion of correct responses (*hereafter*, accuracy) and in terms of signal detection theory (SDT). In the latter approach, we computed both the *A'* values (Zhang & Mueller, 2005) as a measure of sensitivity, that is, the participants' ability to distinguish different from same arrays:

$$A' = \begin{cases} \frac{3}{4} + \frac{H-F}{4} - F(1-H) & \text{if F} \leq \ 0.5 \leq \ \text{H} \\ \frac{3}{4} + \frac{H-F}{4} - \frac{F}{4H} & \text{if F} \leq \ \text{H} \leq \ 0.5 \\ \frac{3}{4} + \frac{H-F}{4} - \frac{1-H}{4(1-F)} & \text{if } 0.5 \leq \ \text{F} \leq \ \text{H} \end{cases}$$

and the criterion *c* (Zhang & Mueller, 2005) to measure the response bias, that is, the participants' tendency to respond "same" or "different":

$$c = -\frac{Z(H) + Z(F)}{2}$$

SDT measures were used to clarify the mechanism behind differences in accuracy rates: a change in perceptual sensitivity between facing and non-facing conditions, and/or a response criterion, more or less conservative with respect to reporting a change in an array. In this and in the following experiments, participants' accuracy proved more sensitive than RTs to the effects of experimental manipulations, as it could be predicted for this task (Kaiser et al., 2015). Therefore, here we focus on accuracy measures and provide full report of RT results as Supplementary Information.

In Experiment 1, no participant performed below or above 2 *SD* from the group accuracy mean; therefore, they were all included in the forthcoming analyses. Participants' data were analyzed with repeated-measures ANOVAs. For each effect, we computed the Bayesian factor, which is reported as Supplementary Information (Tables s1-S4).

Accuracy. Mean accuracy rates were analyzed in a repeated-measures ANOVA with factors Spatial position (facing/non-facing), Set size (set2/set3) and Trial type (same/different). As illustrated in Figure 2A, accuracy values revealed an advantage for facing over non-facing arrays, which especially emerged in different-trials. This pattern was confirmed by statistical analysis. The ANOVA revealed a main effect of Spatial position, $F(1,19) = 10.5$, $p = .004$, $\eta_p^2 = .354$, showing that participants were more accurate with facing than non-facing arrays. Spatial position significantly interacted with Trial type, $F(1,19) = 15.1$, $p < .001$, $\eta_p^2 = .443$. That is, the advantage for facing over non-facing arrays emerged in different-trials, $t(19) = 5.03$, $p < .001$, Cohen's $d = .35$, but not in same-trials, $t(19) = .17$, $p = .864$, Cohen's $d = .04$. This pattern applied to conditions with both set2, $t(19) = 29.29$, $p < .001$, Cohen's $d = 2.09$; and set3, $t(19) = 2.72$, $p = .013$, Cohen's $d = .30$ (Trial type x Set size x Spatial position, $F(1,19) = 2.14$, $p = .159$).

Also significant were the main effects of Set size, $F(1,19) = 244.5$, $p < .001$, $\eta_p^2 = .927$, and Trial type, $F(1,19) = 64.8$, $p < .001$, $\eta_p^2 = .773$. The interaction between the two factors was significant, $F(1,19) = 53.1$, $p < .001$, $\eta_p^2 = .736$, revealing a larger difference between set2 and set3, in different-trials, $t(19) = 11.57$, $p < .001$, Cohen's $d = 1.78$, than in same-trials, $t(19) = 2.093$, $p = .05$, Cohen's $d = 0.59$. The Set size x Spatial position interaction was not significant, $F(1,19) = 1.834$, $p = .192$, $\eta_p^2 = .088$.

Sensitivity. *A′* values were entered in a repeated-measures ANOVA with factors Spatial position and Set size. Results showed a main effect of Set size $F(1,19) = 140.4$, $p < .001$, $\eta_p^2 = .880$, revealing that participants' discrimination between same and different arrays was higher in set2-trials compared to set3-trials. All other effects were not significant (Spatial position, $F(1,19) = 2.32$, $p = .143$, $\eta_p^2 = .109$; Spatial position x Set size, $F(1,19) = .42$, $p = 521$, $\eta_p^2 = .021$).

Response bias. We calculated both participants' general response bias (participants' tendency to respond *same* or *different* throughout the experiment) and participants' response bias as a function of the experimental conditions: facing or non-facing. Participants showed a general bias to respond "same", $t(19) = 9.77$, $p < .001$, Cohen's $d = 2{,}18$. Importantly, however, the Spatial position x Set size ANOVA showed that the bias was stronger in non-facing trials than in facing trials (main effect of Spatial position, $F(1,19) = 6.25$, $p = .021$, $\eta_p^2 = .247$). There was also a main effect of Set size, $F(1,19) = 34.09$, $p < .001$, $\eta_p^2 = .642$, reflecting a stronger bias to respond "same" in set3 trials, compared to set2 ones. The Set size by Spatial position interaction was not significant, $F(1,19) = .28$, $p = .599$, $\eta_p^2 = .014$.

Summary of results. Experiment 1 showed a WM advantage of facing over non-facing dyads in a task that required participants to hold the representation of visual stimuli in WM, for delayed recognition in the probe array. STD analyses showed that the advantage, found in different-trials only, reflected not so much a difference in sensitivity as a difference in the criterion, between facing and non-facing trials. In particular, in processing non-facing arrays, participants were more inclined to respond "same", that is, less inclined to report –or less certain about– the change in the array (Stanislaw & Todorov, 1999). The overall bias to respond "same" can also account for the near ceiling performance in same trials, observed in the accuracy analysis.

The performance advantage with facing trials is compatible with the hypothesis that being face-to-face improves the WM representation of the arrays, by promoting chunking of the crowd in dyads. Alternatively, it could reflect the difference between processing of meaningful events in facing arrays *vs.* processing of meaningless scenes in non-facing arrays, where body positioning broke the meaning of interaction-events. Experiment 2 speaks to that question.

Experiment 2 (79)

In Experiment 1, facing dyads depicted meaningful interactions, while non-facing dyads depicted meaningless events, as positioning the bodies back-to-back disrupted the representation of the interaction. Thus, results of Experiment 1 could reflect the advantage of processing facing (*vs.* non-facing) dyads, or the advantage of processing meaningful (*vs.* meaningless) scenes. In Experiment 2, we sought to disentangle the effect of spatial positioning from the effect of meaning, by presenting dyads of facing and non-facing bodies with no obvious semantic content.

Participants (73)

Twenty healthy adults (14 females; mean age 24.2 ± 4.4) participated as paid volunteers. All had normal or corrected-to-normal vision, reported no history of neurological or psychiatric conditions and no assumption of psychoactive substances or medications. They gave informed consent prior to participation. The sample size was the same as in Experiment 1, as Experiment 2 sought to test whether the effects in Experiment 1 could be replicated with the new stimulus set.

Stimuli and procedures (160)

Stimuli of Experiment 2 were formed using the same procedure of Experiment 1 (Stimulus section of Experiment 1), except that they involved dyads from the ML set. As in Experiment 1, a unique set of 432 arrays was created for each participant (160 arrays of two/three facing or non-facing dyads and as many identical or different arrays), in addition to 64 unique arrays (32 sample and 32 probes) for training and 48 arrays (24 samples and 24 probes) for the instructions and familiarization phase (the same 48 arrays were used for all participants). To create the different-probe arrays, one dyad in the 50% of sample arrays was replaced by a new dyad randomly selected from the remaining dyads of the ML set. Experimental setting, task, and procedures were identical to Experiment 1, except that here participants were instructed to respond after the probe-array disappeared from the screen (after 2000 ms from probe onset) to encourage the exploration of the arrays.

Results (750)

Mean accuracy values of all participants were within 2 *SD* from the group mean; therefore, they were all included in the following analyses.

Accuracy. Mean proportions of correct responses were analyzed in a 2 Spatial position x 2 Trial type x 2 Set size repeated-measures ANOVA. As shown in Figure 2B, we found no main effect of Spatial Position, $F(1,19) = 3.53$, $p = .075$, but a significant interaction between Spatial Position and Trial type, $F(1,19) = 11.4$, $p = .003$, $\eta_p^2 = .375$. Congruent with Experiment 1, the interaction revealed an advantage for facing over non-facing dyads in different-trials, $t(19) = 3.06$, $p = .006$, Cohen's $d = 0.43$, but not in same-trials, $t(19) = 0.98$, $p = .336$, Cohen's $d = 0.12$. A trend for the Set size by Spatial position interaction, $F(1,19) = 3.74$, $p = .068$, $\eta_p^2 = .164$, showed that the advantage of facing *vs.* non-facing arrays was stronger with set3-arrays, $t(19) = 2.33$, $p = .030$, Cohen's $d = 0.41$, than with set2-arrays, $t(19) = 0.23$, $p = .814$, Cohen's $d = 0.05$. The three way interaction between Trial type, Set size and Spatial position, however, did not reach the significance, $F(1,19) = 1.97$, $p = .175$, $\eta_p^2 = .094$. We also found the main effect of Trial type, $F(1,19) = 7.38$, $p = .014$, $\eta_p^2 = .279$, the main effect of Set size, $F(1,19) = 54.25$, $p < .001$, $\eta_p^2 = .74$, and an interaction between the two, $F(1,19) = 8.29$, $p = .009$, $\eta_p^2 = .303$, showing that the difference in performance with set2- and set3-arrays (better with set2 than set3), was larger for different trials.

Sensitivity. A 2 Spatial position x 2 Set size repeated-measures ANOVA on *A'* values showed a main effect of Spatial position, $F(1,19) = 5.14$, $p = .03$, $\eta_p^2 = .213$. That is, participants were better at discriminating different, from same arrays in facing trials, compared to non-facing trials. There was also a main effect of Set size, $F(1,19) = 42.7$, $p < .001$, $\eta_p^2 = .692$, and a significant interaction between the two factors, $F(1,19) = 5.51$, $p = .02$, $\eta_p^2 = .224$, showing that the difference in sensitivity between facing and non-facing trials emerged with set3 arrays, $t(19) = 3.22$, $p = .004$, Cohen's $d = 0.44$, but not with set2 arrays, $t(19) = 0.06$, $p = .948$, Cohen's $d = 0.05$.

Response bias. Like in Experiment 1, participants showed a general bias to respond "same", $t(19) = 2.52$, $p = .02$, Cohen's $d = .56$, which can account for the near ceiling performance with same-trials. The Spatial position x 2 Set size repeated-measures ANOVA showed that the response bias towards "same" was stronger with non-facing arrays (effect of Spatial position, $F(1,19) = 8.36$, $p = .009$, $\eta_p^2 = .305$). All other effects were not significant (Set size, $F(1,19) = 2.86$, $p = .107$, $\eta_p^2 = .130$, Set size x Spatial position, $F(1,19) = .61$, $p = .443$, $\eta_p^2 = .031$).

Summary of results. Results of Experiment 2 based on accuracy rates were congruent with those of Experiment 1 in showing a performance advantage in trials with facing arrays, when participants had to report a change in the probe (i.e., in different-trials). In Experiment 1, the advantage was found for arrays of facing dyads depicting meaningful interactions. Here, we replicated the effect with arrays of facing dyads that did not give rise to any obvious meaningful interaction. Further analyses with Experiment (1 or 2) as a between-subjects factor confirmed that performance did not differ between the two experiments (Supplementary Information). SDT results further clarified that, in Experiment 2, the advantage of facing arrays reflected both a greater sensitivity to the same-different distinction in facing trials and a stronger bias to respond "same" in non-facing trials. Altogether, these effects are compatible with the hypothesis that the face-to-face positioning of bodies improves the processing of crowded arrays in a WM task. This advantage was not affected by the type of relation represented in the dyads (meaningful or meaningless), suggesting that being face-to-face, even in the absence of a meaningful, familiar interaction, defines a relation that binds two bodies together in WM. Moreover, in both experiments, many participants reported to have spontaneously labeled (or attempted to) facing dyads and rehearsed those labels in the interval elapsing between sample and probe, further suggesting similar processing of MF and ML facing arrays.

So, was the representation of MF and ML facing arrays really the same in WM? Experiments 3-4 speak to this question.

Experiment 3 (85)

In Experiment 3, we introduced verbal shadowing through the continuous repetition of a verbal message, in order to increase the WM load during processing of both MF and ML sets. The goal was to expose possible differences between the MF and ML set. In particular, we tested whether the secondary task would impact the WM processing (i.e., chunking) of stimuli, and whether it would do it differently for stimuli depicting a familiar, meaningful relation (MF set) *versus* stimuli with a weaker semantic relation (ML set).

Participants (73)

Twenty-seven healthy adults (24 females; mean age 23.61 ±4.82 *SD*) participated as paid volunteers. This sample size was established taking into account the size of the effect of Spatial Position found in Experiment 1 (η_p^2 = .354, alpha = 0.05, beta = 0.95). All participants had normal or corrected-to-normal vision, reported no history of neurological or psychiatric conditions and no consumption of psychoactive substances or medications. All gave informed consent prior to participation.

Stimuli and procedure (233)

Stimuli, apparatus and procedures were identical to Experiments 1-2, except for the following aspects. First, all participants saw all the stimuli of both Experiments 1 and 2, which doubled the number of trials (320 in total) and conditions (16 conditions: same- and different-trials with set2 and set3 arrays of MF- and ML- facing and non-facing dyads). Stimuli based on the MF set and those based on the ML set were presented in independent runs, with the order alternating between participants. Second, concurrently with the delayed change-detection task, participants performed a shadowing task. To implement this task, the trial began with two target-digits presented for 500 ms in red ink, at the center of the screen (1°×1° visual angle). Participants were instructed to read the two target-digits and repeat them aloud throughout the trial. Meanwhile, the trial unfolded identical to Experiments 1-2, with sample- and probe-arrays for the change-detection task. Participants were instructed to wait until the probe disappeared to respond. After the participant responded to the change-detection task, a red digit appeared on the screen and participants had to decide whether this was one of the two target-digits. If so, they had to press the key "F" with the left index. If no response was provided within 2000 ms, it counted as a "no" response and the next trial began. The experiment lasted ~85 min (~70 minutes for task + ~15 of breaks).

Results (1047)

All participants performed within 2 *SD* from the group accuracy mean and were included in the analysis. Only trials in which participants provided a correct response in the secondary task were considered for further analysis (mean rejected trials 5.92 ± 5.94 *SD*).

Accuracy. Mean accuracy values were analyzed in a 2 Set (MF/ML) x 2 Spatial position (facing/non-facing) x 2 Set size (set2/set3) x 2 Trial type (same/different) repeated-measures ANOVA. Results showed no effect of Spatial position, $F(1,26) = 1.27$, $p= .269$, but a significant interaction between Spatial position and Trial type, $F(1,26) = 9.65$, $p = .004$, $\eta_p^2 = .270$. This interaction revealed a stronger advantage of facing over non-facing arrays in different-trials, $t(26) = 2.36$, $p = .026$, Cohen's $d = .29$, than in same-trials, $t(26) = 1.97$, $p = .056$, Cohen's $d = .19$ (Figure 3A). Results also showed an interaction between Set and Spatial position, $F(1,26) = 12.68$, $p = .001$, $\eta_p^2 = .327$. All effects were qualified by the interaction between Set, Set size and Spatial position, $F(1,26) = 20.44$, $p = .001$, $\eta_p^2 = .440$. This interaction showed that the advantage for facing over to non-facing dyad-arrays in set3 trials, was found only when facing dyads belonged to the MF set, $t(26) = 3.52$, $p = .001$, Cohen's $d = .57$. The opposite effect (an advantage for non-facing over facing set3-arrays) was observed when facing dyads were from the ML set, $t(26) = 2.27$, $p = .031$, Cohen's $d = .42$. Finally, we found an effect of Trial type, $F(1,26) = 73.72$, $p < .001$, $\eta_p^2 = .739$, reflecting higher accuracy with same- than different-trials, an effect of Set size, $F(1,26) = 422.01$, $p < .001$, $\eta_p^2 = .941$, reflecting higher accuracy with set2- than set3-trials, and a significant Set x Trial Type interaction, $F(1,26) = 45.2$, $p < .001$, $\eta_p^2 = .635$. All other effects were not significant (Set x Set size, $F(1,26) < 1$, $p = .941$, $\eta_p^2 < .001$; Set size x Spatial position, $F(1,26) < 1$, $p = .81$, $\eta_p^2 = .002$; Set x Trial type x Set size, $F(1,26) = 1.073$, $p = .309$, $\eta_p^2 = .039$; Set x Trial type x Spatial Position, $F(1,26) = 2.598$, $p = .119$, $\eta_p^2 = .090$; Trial type x Set size x Spatial position, $F(1,26) < 1$, $p = .684$, $\eta_p^2 = .006$; Set, $F(1,26) = .75$, $p = .391$, $\eta_p^2 = .028$; Set x Trial type, $F(1,26) = .38$, $p = .537$, $\eta_p^2 = .014$; Set x Spatial position x Set size x Trial type, $F(1,26) = 3.78$, $p = .063$, $\eta_p^2 = .127$).

Since participants processed MF and ML stimuli in separate blocks, we tested whether the order of conditions affected the performance. We repeated the above analysis adding the Order (MF first or ML first) as a between-subjects factor in the ANOVA. Results showed no effect of Order or interaction of this factor with any other factor in the model (Supplementary Information).

Sensitivity. *A'* values were entered in a 2 Set x 2 Spatial position x 2 Set size repeated-measures ANOVA. Results showed an interaction between Spatial position and Set, $F(1,19) = 11.5$, $p = .002$, $\eta_p^2 = .307$, reveling a greater sensitivity to the same-different distinction in facing arrays, than in non-facing arrays, with the MF set only, $t(26) = 1.99$, $p = .05$, Cohen's $d = .29$. The opposite trend was observed with the ML set, $t(26) = 2.09$, $p = .04$, Cohen's $d = .30$. The interaction between Spatial position, Set and Set size was also significant, $F(1,19) = 19.2$, $p < .001$, $\eta_p^2 = .425$, showing that the above effects only emerged in set3-trials (MF-facing *vs.* non-facing arrays, $t(26) = 2.71$, $p = .01$, Cohen's $d = .39$; ML-facing *vs.* non-facing arrays, $t(26) = 2.44$, $p = .02$, Cohen's $d = .52$). The main effect of Set size was significant, $F(1,19) = 186.3$, $p < .001$, $\eta_p^2 = .877$. All other effects were not significant (Spatial position, $F(1,19) = .01$, $p = .914$, $\eta_p^2 < .001$; Set, $F(1,19) = .06$, $p = .789$, $\eta_p^2 = .002$; Set size x Spatial position, $F(1,19) = .21$, $p = .645$, $\eta_p^2 = .008$; Set size x Set, $F(1,19) = .02$, $p = .869$, $\eta_p^2 = .001$).

Response bias. Participants showed a general bias to respond “same” (MF: $t(26) = 7.89$, $p < .001$, Cohen's $d = 1.51$; ML: $t(26) = 9.78$, $p < .001$, Cohen's $d = 1.88$). A 2 Set x 2 Spatial position x 2 Set size repeated-measures ANOVA showed that the bias was stronger for non-facing arrays than for facing arrays (effect of Spatial position: $F(1,19) = 8.13$, $p = .008$, $\eta_p^2 = .238$), and for set3-trials than for set2-trials (effect of the Set size: $F(1,19) = 27.57$, $p < .001$, $\eta_p^2 = .514$). All other effects were not significant (Set, $F(1,19) = 1.13$, $p = .297$, $\eta_p^2 = .041$; Set size x Spatial position, $F(1,19) = .29$, $p = .589$, $\eta_p^2 = .011$; Set x Set size , $F(1,19) = 1.32$, $p = .260$, $\eta_p^2 = .048$; Set x Spatial position , $F(1,19) = .205$, $p = .654$, $\eta_p^2 = .007$; Set x Set size x Spatial position , $F(1,19) = .51$, $p = .479$, $\eta_p^2 = .019$).

Summary of results. Like in Experiments 1-2, the WM advantage of facing, over non-facing arrays emerged in the most demanding of the experimental conditions, i.e., the condition in which participants had to detect a change in an array of six bodies. With concurrent shadowing, however, the advantage was only found for facing arrays featuring familiar, meaningful interactions (MF set). SDT analyses clarified that, with the MF set, the advantage of facing (*vs.* non-facing) arrays reflected better discrimination of different arrays from same arrays. This effect was erased, and indeed reversed, in the processing of the ML set.

These results showed that the semantic relation defined by MF-facing dyads provided an effective principle for chunking in WM, regardless of the opportunity for verbal labeling. The increase in WM load due to the secondary task –or, maybe, the specific interference of that task with labeling and rehearsal– instead abolished the advantage of ML-facing arrays. This suggests that the advantage of ML-facing dyads in Experiment 2 reflected a spontaneous, impromptu attribution of meaningful relations, represented in WM in the form of verbal labels.

Experiment 4 (142)

In Experiment 3, we reported a difference in the performance with the MF *vs.* the ML set: the advantage for facing over non-facing dyads survived the introduction of a secondary task in the case of the MF set, while it disappeared (and was even reversed) for the ML set. We attributed this change to the secondary task impacting the performance on stimuli with the weaker semantic relation (i.e., ML set). To single out and confirm the effect of the secondary task, in Experiment 4, we replicated the design of Experiment 3 without verbal shadowing. If the performance difference between MF and ML stimuli in Experiment 3 reflected the selective effect of the secondary task on processing stimuli with the weaker relation (i.e., ML set), the abolition of the secondary task should restore the advantage of facing (vs. non-facing) arrays for ML stimuli.

Participants (38)

Twenty-eight healthy adults (20 females, mean age = 22.15 ± 2.48) participated as paid volunteers. All participants had normal or corrected-to-normal vision and gave informed consent before participating. The sample size is the same as in Experiment 3.

Stimuli and procedure (38)

Stimuli, apparatus and procedures were identical to Experiments 3, except for the presence of the secondary task, which reduced the total duration of the experiment to about 75 min (~60 minutes of tasks + ~15 minutes of breaks).

Results (933)

Data from one participant were discarded, as the accuracy rate was >2 *SD* lower than the group mean.

Accuracy. A 2 Set (MF, ML) x 2 Spatial position (facing, non-facing) x 2 Set size (set2, set3) x 2 Trial type (same, different) repeated-measures ANOVA on accuracy rates from the remaining 27 participants confirmed the results of Experiments 1-2. In particular, there was a significant interaction between Spatial Position and Trial type, $F(1,26) = 8.84$, $p = .006$, $\eta_p^2 = .253$, reflecting higher accuracy with facing than with non-facing arrays in different-trials, $t(26) = 2.4$, $p= .023$, Cohen's $d = .30$, and the opposite trend in same-trials, $t(26)= 2.7$, $p= .011$, Cohen's $d = .39$. The effect of Spatial Position did not interact with any other factor; importantly, the effect of Spatial Position did not interact with the Set (MF/ML), meaning that the processing of facing (*vs.* non-facing) arrays was not affected by semantic relations in the dyads (Figure 3B).

Results also showed: an effect of Trial type, $F(1,26) = 174.60$, $p < .001$, $\eta_p^2 = .870$, reflecting higher accuracy in same- *vs.* different-trials; an effect of Set size, $F(1,26) = 171.9$, $p < .001$, $\eta_p^2 = .868$, reflecting better performance with set2- *vs.* set3-trials; and a significant interaction between Set and Set size, $F(1,26) = 7.73$, $p = .010$, $\eta_p^2 = .229$, showing that the performance difference between set2-trials and set3-trials was stronger for the ML (*vs.* the MF) set. Finally, there was a significant interaction between Trial type and Set size, $F(1,26) = 84.9$, $p < .001$, $\eta_p^2 = .765$, whereby the difference between set2 and set3-trials was stronger in different-trials, relatively to same-trials. No other effect or interaction approached the significance (Spatial Position: $F(1,26) = 1.43$, $p = .241$, $\eta_p^2 = .052$; Set: $F(1,26) = 2.874$, $p = .101$, $\eta_p^2 = .099$; Set x Trial type: $F(1,26) = 1.63$, $p = .212$, $\eta_p^2 = .059$; Set x Spatial position, $F(1,26) < 1$, p =. 899, $\eta_p^2 < .001$; Set size x Spatial position: $F(1,26) < 1$, $p = .443$, $\eta_p^2 = .028$; Set x Trial type x Set size, $F(1,26) = 3.36$, $p = .078$, $\eta_p^2 = .114$; Set x Trial type x Spatial position, $F(1,26) < 1$ p =. 945, $\eta_p^2 < .001$; Set x Set size x Spatial position: $F(1,26) < 1$, $p = .744$, $\eta_p^2 = .004$; Trial type x Set size x Spatial position: $F(1,26) < 1$, $p = .976$, $\eta_p^2 < .001$; Set x Trial type x Set size x Spatial position: $F(1,26) < 1$, $p = .358$, $\eta_p^2 = .032$).

In a secondary analysis, we tested the effect of the order in which participants performed the task and found no change in the key effect of Spatial Position for the MF and the ML set, depending on the order of blocks (MF first or ML first; Supplementary Information).

Sensitivity. *A'* values, analyzed in a 2 Set (MF/ML) x 2 Spatial position (facing/non-facing) x 2 Set size (set2/set3) repeated-measures ANOVA, showed a main effect of Set size, $F(1,19) = 122.1$, $p < .001$, $\eta_p^2 = .824$, reflecting increased sensitivity in set2-trials than in set3-trials, and an interaction between Set and Set size, $F(1,19) = 6.24$, $p = .019$, $\eta_p^2 = .193$, indicating that the difference in sensitivity between set2- and set3-trials was more pronounced with the ML set, $t(26) = 12.14$, $p < .001$, Cohen's $d = 1.86$, then with the MF set, $t(26) = 6.65$, $p < .001$, Cohen's $d = 1.19$. All other effect were not significant (Spatial position, $F(1,19) < .001$, $p = .998$, $\eta_p^2 < .001$; Set, $F(1,19) = 2.90$, $p = .100$, $\eta_p^2 = .100$; Set size x Spatial position, $F(1,19) = 1.68$, $p = .205$, $\eta_p^2 = .060$; Spatial position x Set, $F(1,19) = .03$, $p = .855$, $\eta_p^2 = .001$; Set size x Spatial position x Set, $F(1,19) = .43$, $p = .513$, $\eta_p^2 = .016$).

Response bias. Consistent with all previous experiments, participants showed a general bias to respond “same” (MF: $t(26) = 10.77$, $p < .001$, Cohen's $d = 2.43$; ML: $t(26) = 13.92$, $p < .001$, Cohen's $d = 2.63$). A 2 Set x 2 Spatial position x 2 Set size repeated-measures ANOVA showed that the bias was stronger in non-facing trials than in facing trials (effect of Spatial position: $F(1,19) = 15.12$, $p < .001$, $\eta_p^2 = .367$), and with set3-trials than with set2-trials (effect of Set size: $F(1,19) = 33.59$, $p < .001$, $\eta_p^2 = .563$). All other effects were not significant (Set, $F(1,19) = 1.60$, $p = .216$, $\eta_p^2 = .058$; Set size x Spatial position, $F(1,19) = .45$, $p = .506$, $\eta_p^2 = .017$; Set size x Set, $F(1,19) = .04$, $p = .827$, $\eta_p^2 = .001$; Spatial position x Set, $F(1,19) = .009$, $p = .924$, $\eta_p^2 < .001$; Set size x Spatial position x Set, $F(1,19) = .80$, $p = .377$, $\eta_p^2 = .030$).

Summary of results. Accuracy rates in Experiment 4 showed an advantage for facing over non-facing dyads in different trials, comparable across the MF and the ML set. We found the opposite trend in same trials, which we refrain from interpreting, as no such effect was found in all previous experiments. The advantage of facing arrays reflected a stronger bias to respond “same” for non-facing arrays. In sum, Experiment 4 converged with Experiments 1-2 in showing that the face-to-face positioning of bodies on its own offered an advantage for WM processing. Moreover, Experiment 4 demonstrates that the performance difference between the MF and ML set in Experiment 3 was the consequence of the secondary task, which was the only difference between the last two experiments.

Discussion (1825)

Keeping in mind an array of more than four items for a short period of time, depends on the possibility to chunk multiple items in units in WM. We investigated whether representing a relation between social agents (i.e., human bodies) benefited the processing of crowded scenarios in WM, by offering a structure to organize single individuals in social chunks. In addressing this question, we asked what kind of relation could bind two bodies together in WM. We considered the representation of a familiar, meaningful face-to-face interaction (i.e., fighting, talking, quarreling, or dancing; Experiment 1) and the mere face-to-face positioning of two bodies (Experiment 2). In both cases, we found that arrays of facing dyads were retained and recognized better (higher accuracy rates) than arrays of non-facing dyads, despite differences between stimuli (MF set in Experiment 1, ML set in Experiment 2), participants, and task instructions. The conditions that proved most sensitive to the effect of positioning were those in which a change occurred in arrays that exceeded the WM capacity of four items (Gao et al., 2015; Shen et al., 2014; Wood, 2008), making chunking mandatory to succeed in the task.

SDT analyses clarified that, when present, the advantage of facing over non-facing dyads, found in accuracy rates, was driven by participants' stronger bias to respond "same" to non-facing arrays (Experiments 1, 2 and 4) and, less consistently (in Experiments 2-3), by a greater perceptual sensitivity to the distinction between same and different trials in facing arrays. In other words, participants showed higher accuracy rates with facing arrays because they were less certain, or more cautious, about reporting a change in non-facing (*vs.* facing) arrays and, sometimes (in Experiments 2-3), detected a change in facing arrays, more often than in non-facing arrays.

Participants in Experiments 1-2 reported using verbal labeling to encode and remember facing dyads. Verbal labeling is a common, relatively undemanding strategy to hold information by phonological maintenance or rehearsal. In visual WM tasks, it can provide an additional source of storage, where a verbal code is used to recall visual information. This strategy can be prevented with shadowing by continuous word repetition (Baddeley, et al., 1998; Robbins et al., 1996). Experiment 3 showed that verbal shadowing left unhindered the advantage for meaningful-facing (*vs.* non-facing) dyads but abolished –even inverted– the advantage of meaningless-facing (*vs.* non-facing) dyads. These findings suggests that, in the case of meaningless stimuli, when allowed (Experiment 2), verbal labeling enhanced a faint relationship prompted by spatial positioning, thus facilitating binding of face-to-face bodies.

Could lower-level differences between stimuli in the MF and ML set account for the effect in Experiment 3? Several facts concur to rule out this possibility. First, MF and ML stimuli were matched, as much as possible, for low-level features that could affect grouping, such as distance and center of mass. Second, the MF and ML sets were not compared directly: each facing set was compared with the corresponding non-facing set involving the very same bodies, and we relied on interactions for assessing the effects of Set. Third, the comparison between Experiment 1 (MF set) and 2 (ML set) showed no statistical difference between the two patterns of results (Supplementary Information). The results of Experiment 4, in which we repeated the design of Experiment 3 without the secondary task, ultimately supported the conclusion that the performance difference between meaningful and meaningless stimuli in Experiment 3 reflected the impact of the secondary task on the representation of stimuli with the weaker relation in WM –i.e., the ML set.

Our results suggest that different subsystems supported the maintenance of facing and non-facing dyads in WM: the visuo-spatial sketchpad holding visual representations for non-facing dyads and –in addition to, or instead of the visuo-spatial sketchpad– the phonological loop holding information for facing dyads in a verbal code (Baddeley, 2000; Baddeley & Hitch, 1994). While it is possible that chunking and storage of both meaningful-facing bodies in Experiment 1 and meaningless-facing bodies in Experiment 2 took advantage of verbal labeling, only the former could still be bound together without labeling (Experiment 3).

The resistance of the advantage for meaningful-facing (over non-facing) dyads to the concurrent secondary task highlights the contribution of semantic knowledge of social interaction, which would provide a structure to organize new information in WM. Different semantic content for the different MF-facing dyads in an array might also make the dyads more distinguishable and therefore easier to be individuated and discriminated in the crowded array. WM can exploit various cues in the visual input, which are related to prior knowledge, and can be used to form groups of associated items that are thus stored together (Brady et al., 2009, 2011, 2016; Chen & Cowan, 2009; Cowan, 2000; Ericsson & Kintsch, 1995). Current models of WM suggest that this mechanism recruits the episodic buffer, a component of WM for temporary storage of episodes, with access to and from long-term memory (Baddeley, 2000; Ericsson & Kintsch, 1995; Rossi-Arnaud, et al., 2006).

A similar mechanism might have operated in the processing of meaningless-facing dyads. Our contention is that being face-to-face triggered a general, underspecified representation of social interaction; however, without an anchor to a specific semantic entry, and without the additional support for chunking provided by labeling (Experiment 3), the underspecified representation often failed to support discrimination between two instances of interaction (two meaningless-facing dyads), making the change from sample to probe harder to detect, or more uncertain, for meaningless-facing dyads. While, in this interpretation, the effect of the secondary task is taken to reflect interference with verbal labeling, it could instead be the consequence of a general increase in the WM load. If the latter is true, we should expect to observe the abolition of the advantage for ML-facing over non-facing dyads, using any other (non-verbal) secondary task. Another method to demonstrate the role of verbal labelling in the current WM task would be to drastically reduce the stimulus duration, so to prevent the verbal-labeling strategy (Vogel et al., 2001).

The current results contribute to demonstrate that, among familiar associations and semantic relations (Brady et al., 2009; Chase & Simon, 1973; Curby et al., 2009; Feigenson & Halberda, 2008; Kaiser et al., 2015; Kibbe & Feigenson, 2013), individuals can use social relationship (i.e., the knowledge about the typical structure of dyadic interactions) for chunking in WM. In previous studies, chunking by social relationship was emphasized by showing meaningful social interactions (physical/communicative exchanges) between social agents acting on, or towards each other (Ding et al., 2017). Here, we set conditions to tell apart the effects of spatial cues (i.e., spatial proximity and positioning) and semantic relations (i.e., category of interaction). In this way, we showed that just being face-to-face, without a familiar, meaningful interaction, can establish a relationship, as faint as it might be, that triggers chunking, to the benefit of WM capacity.

Our results also shed light on the relationship between visual perception and WM. Research on scene perception has shown that visuo-spatial cues of interaction, such as proximity and face-to-face positioning, independently from the meaningfulness of the stimuli, trigger perceptual grouping of multiple bodies (Adibpour et al., 2021), which would account for increased efficiency in stimulus detection and recognition (Papeo, 2020; Papeo et al., 2019, 2017; Strachan, et al., 2019; Vestner et al., 2019). The current results show that grouping triggered by face-to-face positioning, extends to WM. In sum, being face-to-face defines a relationship that is exploited for efficiency in visual perception and for chunking in WM.

A number of questions remain open. One concerns the representation of the single bodies that form an interacting (or seemingly interacting) dyad. In visual search for bodies through a crowd, participants rapidly access two facing bodies as a group (*vs.* non-facing bodies), but with a cost in the access to individual bodies of that group (Papeo et al., 2019). A similar cost might be found in WM processing. Previous research involving familiar objects has shown that compressing information in WM increases capacity but reduces the number of features that are encoded for each individual component (Alvarez, 2011; Alvarez & Cavanagh, 2004; Brady & Alvarez, 2011). This cost however might vary depending on the object class. For socially relevant stimuli such as faces, WM exhibits not only greater capacity relative to non-face objects (Curby & Gauthier, 2007), but also improved resolution (Scolari et al., 2008). Ding et al. (Ding et al., 2017) tested WM for arrays of interacting or non-interacting body dyads by asking participants to report whether a single body was present in the previous array or not. A performance advantage was found for bodies seen in interacting dyads. Those results encourage the hypothesis that the representation of single bodies (or single actions) in WM may be enhanced, rather than impoverished, in the context of a meaningful interaction (see also Abassi & Papeo, 2020; Bellot, et al., 2020; Neri, et al., 2006).

Another open question concerns the features that are more likely to be encoded in the WM representation of an interacting dyad. Research on visual perception of the *gist of events* has shown that individuals are extremely rapid and efficient at extracting information about agent-patient roles (Hafri, et al., 2013, 2018) and action coherence (Glanemann et al., 2016), from the physical structure of the visual input. Are these the features of an interaction that are most likely to be encoded in WM and pass into the long-term memory representation of a social event? Future research shall also investigate what are the other visuo-spatial features that, alone or in interaction with the face-to-face positioning, can trigger representation of social interaction in WM. For example, it is possible that the representation of social interaction becomes weaker as the distance between two bodies increases, and the advantage of facing arrays is abolished beyond a certain distance threshold. And, what happens to the WM representation of social interactions that lack prototypical features such as face-to-face positioning and spatial proximity? Finally, it remains to be established whether the effects reported here are to ascribe to the general WM system, or rather capture the functioning of the so-called social working memory, that is, a set of operations –and, possibly, neural structures– specialized in maintaining and manipulating social information (Druzgal & D'esposito, 2003; LoPresti et al., 2008; Meyer & Lieberman, 2012; Meyer, et al., 2012; Thornton & Conway, 2013). In addressing these questions, future studies will contribute to understand how people encode and remember one of the most important aspects of their visual world and social life: social interaction.

In conclusion, we showed that WM uses information on social relationship to chunk bodies in groups (dyads), thus increasing its capacity. Being face-to-face alone can drive this mechanism: It solicits (tentative) semantic encoding of the stimuli, as suggested by spontaneous labeling, providing an effecting principle for chunking in WM. Thus, two people mutually accessible to one another form an intrinsically meaningful representation that human cognition readily processes as a social unit, before the interaction is fully realized or understood.

Competing interests

The authors declare that they have no competing interest

Consent for publication

Not applicable

Ethics approval and consent to participate

The study was approved by the local ethics committee (Comité de Protection des Personnes, CPP SUD-EST II, IRB: 00009118). Participants gave informed consent prior to participation in the study.

Funding

This work was supported by a European Research Council Starting Grant awarded to L.P. (Project: THEMPO, Grant Agreement 758473) and by the Amici di Claudio Demattè Scholarship (12° edition) awarded to I.P.

Availability of data and materials

All datasets used for supporting the conclusions of this article and some test stimuli are available from the public data repository at the website https://osf.io/2abue/

Authors' contributions

IP and LP both conceived and designed this research and drafted the manuscript. LP coordinated this research. IP carried out experiments and data analysis. The authors read and approved the final manuscript.

References

Abassi, E., & Papeo, L. (2020). The representation of two-body shapes in the human visual cortex. *Journal of Neuroscience*, *40*(4), 852–863.

Adibpour, P., Hochmann, J.-R., & Papeo, L. (2020). Spatial relations trigger visual binding of people. *BioRxiv*.

Alvarez, G. A. (2011). *Representing Multiple Objects as an Ensemble Enhances Visual Cognition Representing multiple objects as an ensemble enhances visual cognition*. https://doi.org/10.1016/j.tics.2011.01.003

Alvarez, G. A., & Cavanagh, P. (2004). The Capacity of Visual Short- Term Memory Is Set Both by Visual Information Load and by Number of Objects. *PSYCHOLOGICAL SCIENCE*, *17*(5), 1019–1029. https://doi.org/10.1208/s12249-015-0428-4

Anderson, D. E., Vogel, E. K., & Awh, E. (2015). "Selection and storage of perceptual groups is constrained by a discrete resource in working memory": Retraction of Anderson et al. (2013). *Journal of Experimental Psychology: Human Perception and Performance*, *41*(5), 1189–1189. https://doi.org/10.1037/xhp0000136

Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders*, *36*(3), 233–240.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*(4), 485.

Baddeley, A. D., & Logie, R. H. (1999). *Working memory: The multiple-component model.*

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158.

Bellot, E., Abassi, E., & Papeo, L. (2020). Moving toward versus away from another: how body motion direction changes the representation of bodies and actions in the visual cortex. *BioRxiv*.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392.

Brady, T. F., 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. (n.d.). *The role of meaning in visual working memory : Real- world objects , but not simple features , benefit from deeper processing*.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*(1), 85–109. https://doi.org/10.1037/a0030779

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Chase, W. G., & Simon, H. A. (1973). *The mind's eye in chess. Visual Information Processing. WG Chase*. New York, Academic Press.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity Behavioral and Brain Sciences. *Behavioral and Brain Sciences*, *4*, 87–185. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+magical+number+4+in+short-term+memory:+A+reconsideration+of+mental+storage+capacity#0

Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin & Review*, *14*(4), 620–628.

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 94.

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

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *22*(1), 170–178.

Druzgal, T. J., & D'esposito, M. (2003). Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *Journal of Cognitive Neuroscience*, *15*(6), 771–784.

Erdfelder, E., FAul, F., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences*, *105*(29), 9926–9930.

Gao, Z., Bentin, S., & Shen, M. (2015). Rehearsing biological motion in working memory: An EEG study. *Journal of Cognitive Neuroscience*, *27*(1), 198–209.

Glanemann, R., Zwitserlood, P., Bölte, J., & Dobel, C. (2016). Rapid apprehension of the coherence of action scenes. *Psychonomic Bulletin and Review*, *23*(5), 1566–1575. https://doi.org/10.3758/s13423-016-1004-y

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). *Getting the Gist of Events: Recognition of Two-Participant Actions from Brief Displays*. *71*(2), 233–236. https://doi.org/10.1038/mp.2011.182.doi

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, *175*(February), 36–52. https://doi.org/10.1016/j.cognition.2018.02.011

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 31–47. https://doi.org/10.1037/0096-1523.33.1.31

Kaiser, D., Stein, T., & Peelen, M. V. (2014). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proceedings of the National Academy of Sciences*, *111*(30), 11217–11222. https://doi.org/10.1073/pnas.1400559111

Kaiser, D., Stein, T., & Peelen, M. V. (2015). Real-world spatial regularities affect visual working memory for objects. *Psychonomic Bulletin and Review*, *22*(6), 1784–1790. https://doi.org/10.3758/s13423-015-0833-4

Kibbe, M., & Feigenson, L. (2013). Infants use statistical regularities to chunk items in visual working memory. *Journal of Vision*, *13*(9), 333.

LoPresti, M. L., Schon, K., Tricarico, M. D., Swisher, J. D., Celone, K. A., & Stern, C. E. (2008). Working memory for social cues recruits orbitofrontal cortex and amygdala: a functional magnetic resonance imaging study of delayed matching to sample for emotional expressions. *Journal of Neuroscience*, *28*(14), 3718–3728.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, *122*(3), 346–362.

Meyer, M. L., & Lieberman, M. D. (2012). Social working memory: neurocognitive networks and directions for future research. *Frontiers in Psychology*, *3*, 571.

Meyer, M. L., Spunt, R. P., Berkman, E. T., Taylor, S. E., & Lieberman, M. D. (2012). Evidence for social working memory from a parametric functional MRI study. *Proceedings of the National Academy of Sciences*, *109*(6), 1883–1888.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

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

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

Papeo, L. (2020). *Twos in human visual perception*.

Papeo, L., Goupil, N., & Soto-Faraco, S. (2019). Visual search for people among people. *Psychological Science*.

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

Robbins, T. W., Anderson, E. J., Barker, D. R., Bradley, A. C., Fearnyhough, C., Henson, R., Hudson, S. R., & Baddeley, A. D. (1996). Working memory in chess. *Memory & Cognition*, *24*(1), 83–93.

Rossi-Arnaud, C., Pieroni, L., & Baddeley, A. (2006). Symmetry and binding in visuo-spatial working memory. *Neuroscience*, *139*(1), 393–400.

Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review*, *15*(1), 215–222.

Shen, M., Gao, Z., Ding, X., Zhou, B., & Huang, X. (2014). Holding biological motion information in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1332.

Stahl, A. E., & Feigenson, L. (2014). Social knowledge facilitates chunking in infancy. *Child Development*, *85*(4), 1477–1490. https://doi.org/10.1111/cdev.12217

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149.

Strachan, J. W. A., Sebanz, N., & Knoblich, G. (2019). The role of emotion in the dyad inversion effect. *PloS One*, *14*(7), e0219185.

Thornton, M. A., & Conway, A. R. A. (2013). Working memory for social information: Chunking or domain-specific buffer? *Neuroimage*, *70*, 233–239.

Vestner, T., Gray, K. L. H., & Cook, R. (2020). Why are social interactions found quickly in visual search tasks? *Cognition*, *200*, 104270.

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*. 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.

Wood, J. N. (2008). Visual memory for agents and their actions. *Cognition*, *108*(2), 522–532.

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, *70*(1), 203–212.

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*. https://doi.org/10.1038/s41562-019-0618-2

Figure Captions

Figure 1. Example of trial and stimulus-arrays. A) Trial organization in Experiments 1-2. Participants saw two arrays (sample and probe) with either two (set2) or three (set3) facing or non-facing dyads. Participants had to report whether the probe was the same or different relative to the sample. Represented here is a same-trial with set2-facing array. B) Example arrays from the meaningful set (Experiment 1). C) Example arrays from the meaningless set (Experiment 2).

Figure 2. Results of analyses on accuracy rates, *A'* and *c* values in Experiments 1-2. A) Results of Experiment 1. B) Results of Experiment 2. Accuracy rate results are shown as a function of Spatial position (facing or non-facing), Set size (Set2 or Set3) and Trial type (different or same). *A'* and *c* results are shown as a function of Spatial position (facing or non-facing) and Set size (Set2 or Set3). Error bars represent ±1 within-subjects Standard Error of the Mean (SEM) (Cousineau, 2005). Asterisks indicate significance for pairwise comparisons (*$p < .05$, ** $p < .01$, ***$p < .001$).

Figure 3. Results of analyses on accuracy rates, *A'* and *c* values in Experiments 3-4. A) Results of Experiment 3. B) Results of Experiment 4. Accuracy rates are shown as a function of Spatial position (facing or non-facing), Set size (Set2 or Set3) and Trial type (different or same). *A'* and *c* results are shown as a function of Spatial position (facing or non-facing) and Set size (Set2 or Set3). Error bars represent ±1 within-subjects Standard Error of the Mean (SEM) (Cousineau, 2005). Asterisks indicate significance for pairwise comparisons (*$p < .05$, ** $p < .01$, ***$p < .001$).

Figure 1

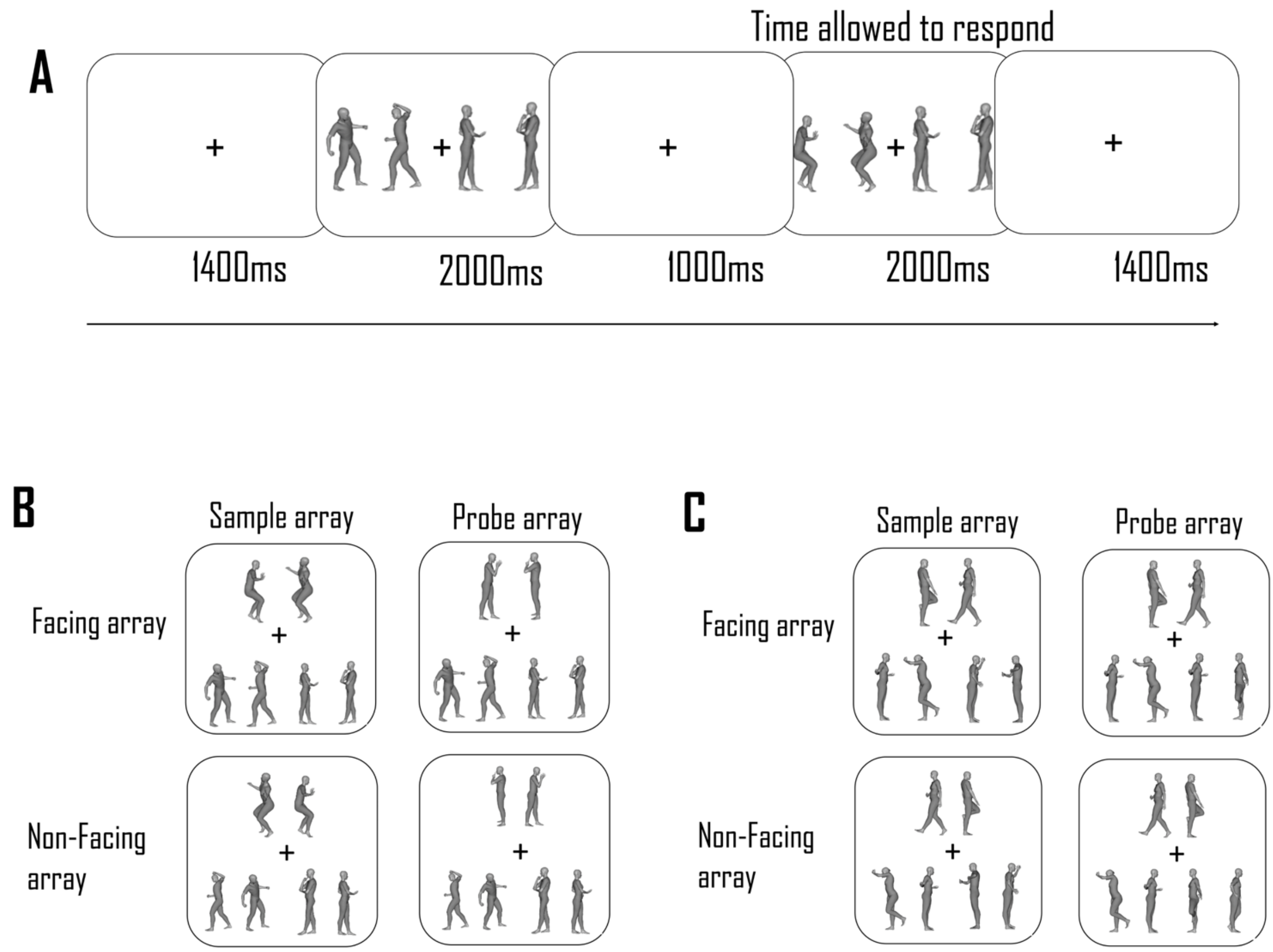

Time allowed to respond
A
1400ms
2000ms
1000ms
2000ms
1400ms
B
Sample array
Probe array
Facing array
Non-Facing array
C
Sample array
Probe array
Facing array
Non-Facing array

Figure 2

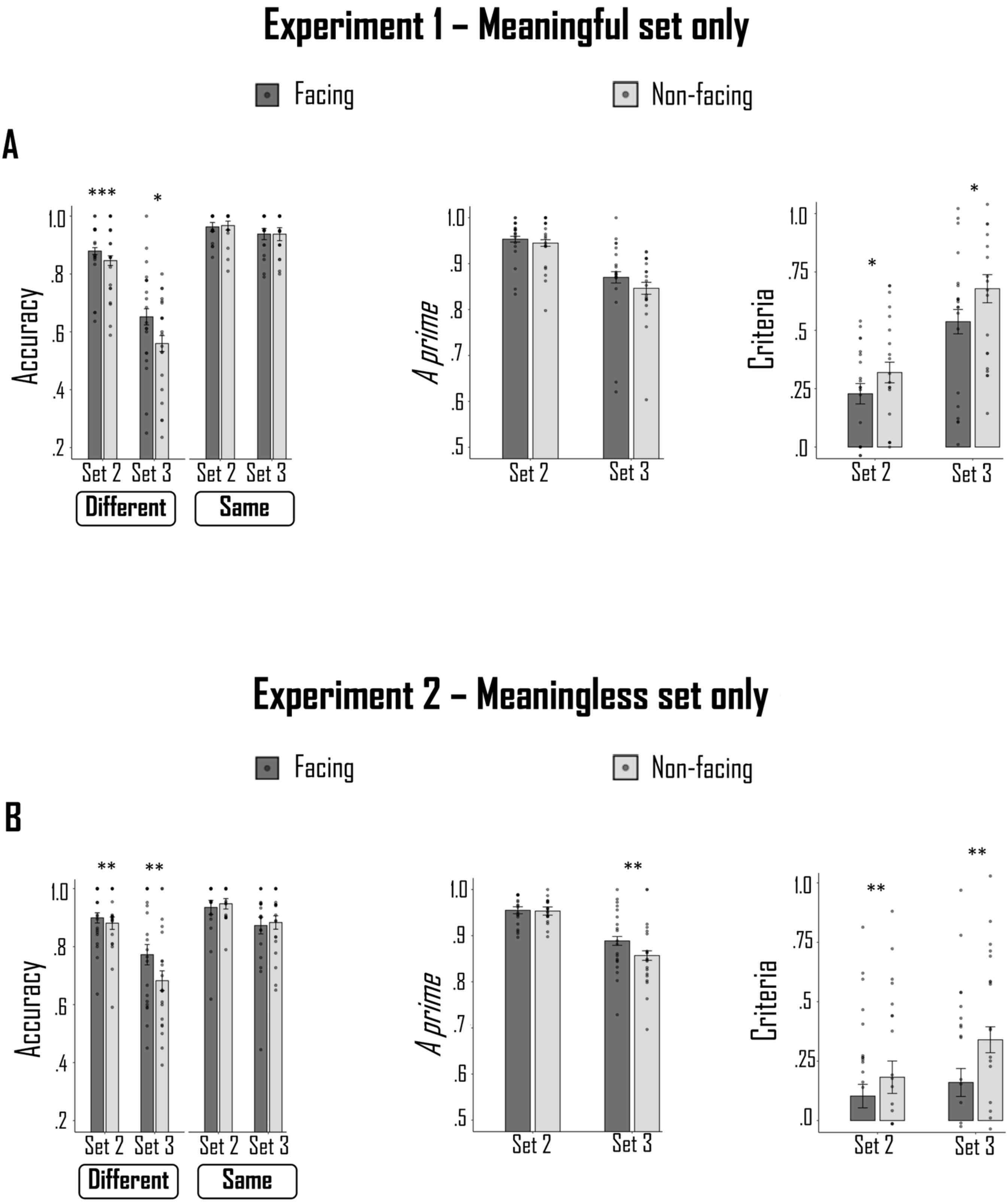
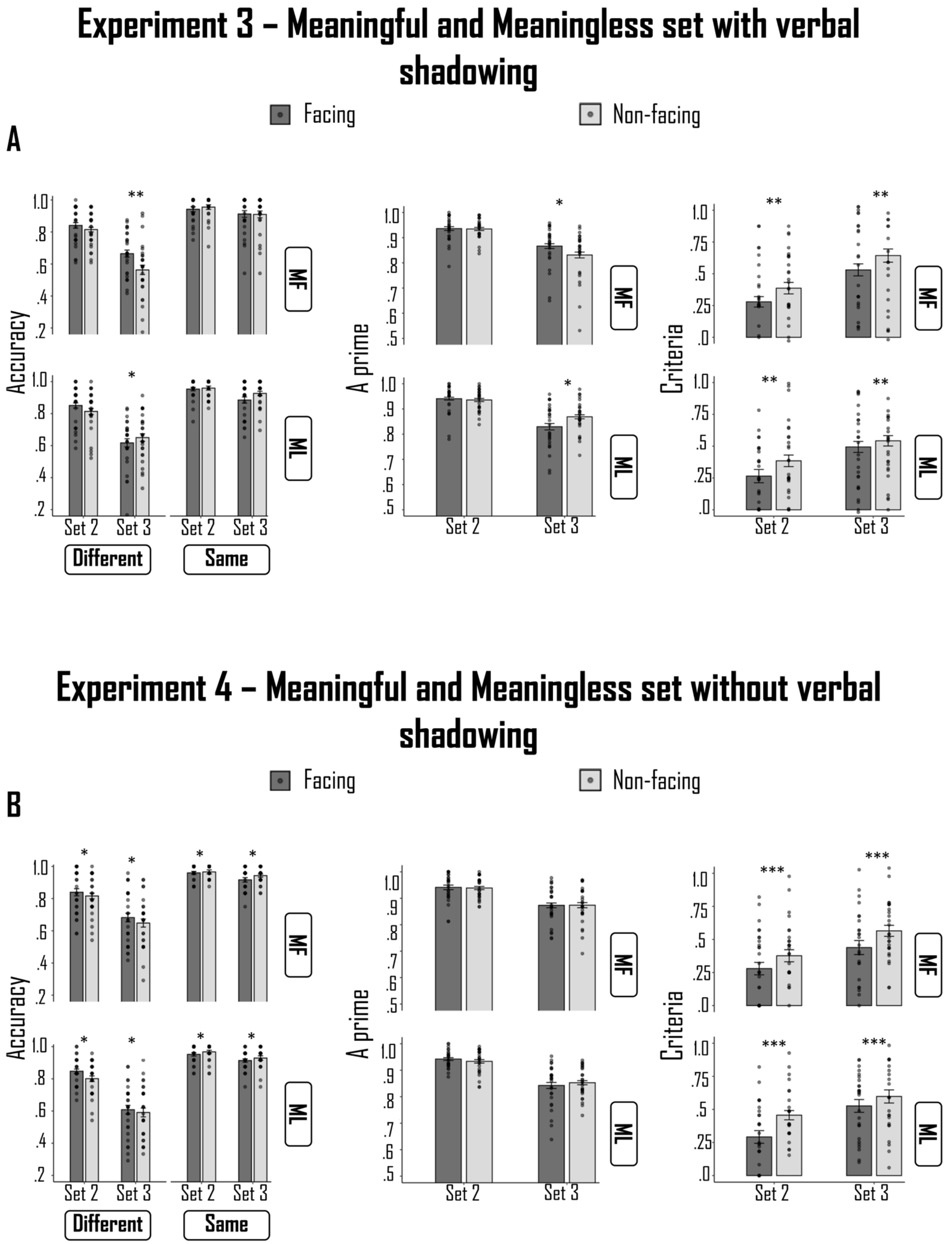

Experiment 1 – Meaningful set only
Facing
Non-facing
A
Accuracy
Set 2
Set 3
Different
Same
A prime
Criteria
Experiment 2 – Meaningless set only
B

Figure 3



Experiment 3 – Meaningful and Meaningless set with verbal shadowing
Facing
Non-facing
A
Accuracy
A prime
Criteria
MF
ML
Set 2
Set 3
Different
Same
Experiment 4 – Meaningful and Meaningless set without verbal shadowing
Facing
Non-facing
B
Accuracy
A prime
Criteria
MF
ML
Set 2
Set 3
Different
Same