Cognition 207 (2021) 104526

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Incidental encoding of visual information in temporal reference frames in working memory

Anna Heuer *, Martin Rolfs

Department of Psychology, Humboldt-Universität zu Berlin, Berlin 10117, Germany

ARTICLE INFO

Keywords:
Visual working memory
Temporal cognition
Spatial cognition

ABSTRACT

Visual events are structured in space and time, yet models of visual working memory (VWM) have largely relied on tasks emphasizing spatial aspects. Here, we show that temporal properties of visual events are incidentally encoded along with spatial properties. In five experiments, participants performed change-detection tasks, in which items had unique spatial and temporal coordinates at encoding. Crucially, neither space nor time was task-relevant. The key manipulation concerned the retrieval context: The test array was identical to the memory array either in its entire spatiotemporal structure, or only its spatial or temporal structure. Removing spatial or temporal information at retrieval resulted in costs, indicating that memory relied on both spatial and temporal context in which items were initially perceived. Encoding of spatiotemporal structure occurred incidentally, not strategically, as it was robust even when the retrieval context was perfectly predictable. However, spatial and temporal inter-item spacings influenced the weighting of spatial and temporal information: It favoured the domain in which items were more widely spaced, facilitating their individuation and, likely, access to representations. Across individuals, the weighting of spatial and temporal information varied substantially, but it remained consistent across sessions, suggesting stable preferences for coding in the spatial or temporal domain. No comparable incidental encoding occurred for other task-irrelevant feature dimensions (size or colour). We propose that temporal structure serves as fundamental a function in VWM as spatial structure, scaffolding events that unfold over time.

1. Introduction

The visual events we encounter unfold in space and time and mostly occur in a spatiotemporal context. It seems natural to assume that space and time are of particular importance for visual memory – spatiotemporal properties of events are not only initially perceived but we may also retain and utilize them to store that information in memory once the event itself has passed. Consider seeing a bike crashing into a car pulling out of a parking lot. The event itself may only last a couple of seconds, yet both the order of events and the locations of the objects at any given time are crucial for subsequent eye witness reports.

The special role of space in visual working memory (VWM) is indeed widely acknowledged, for instance in models positing a location-based architecture, in which non-spatial features are bound via their shared position in space (Pertzov & Husain, 2014; Rajsic & Wilson, 2014; Schneegans & Bays, 2017; Treisman & Zhang, 2006). Objects are not only bound to their own location, even spontaneously (Foster, Bsales, Jaffe, & Awh, 2017), but also to their spatial context: Memory is typically better when inter-object relations remain intact and the spatial context in which the information was encoded is also present during recall (Hollingworth, 2006, 2007; Jiang, Olson, & Chun, 2000; Olson & Marshuetz, 2005; Sun & Gordon, 2009, 2010; Timm & Papenmeier, 2019).

Much less is known about the role of temporal aspects for the short-term retention of visual events. In stark contrast to other areas of memory research such as episodic memory (e.g., Bellmund, Deuker, & Doeller, 2019; Howard & Kahana, 2002), the prevalence of space at the expense of time is deeply ingrained in VWM research. The vast majority of studies employs variants of change detection tasks that rely on spatial codes, which almost universally involve a simultaneous presentation of to-be-memorised items that lacks any distinctive temporal structure (e.g., Luck, 2008; Luck & Vogel, 2013; Souza & Oberauer, 2016). Nevertheless, it has been suggested that time plays a similar role as space in the organization of VWM (Manohar, Pertzov, & Husain, 2017; Schneegans & Bays, 2018): It may provide a context that allows binding of features and access to contents. Empirical evidence for this appealing

* Corresponding author at: Humboldt-Universität zu Berlin, Department of Psychology, Rudower Chaussee 18, Berlin 12489, Germany.
E-mail address: anna.heuer@hu-berlin.de (A. Heuer).

https://doi.org/10.1016/j.cognition.2020.104526
Received 27 May 2020; Received in revised form 21 November 2020; Accepted 23 November 2020
Available online 3 December 2020

idea is sparse but paints a promising picture. For one, studies using a sequential presentation of visual memoranda have demonstrated that temporal aspects such as order or duration can be successfully maintained (e.g., Delogu, Nijboer, & Postma, 2012; Manohar et al., 2017; Rondina, Curtiss, Meltzer, Barense, & Ryan, 2016; Ryan & Villate, 2009), and that changes to this temporal structure interfere with memory (Rondina et al., 2016). Spatially or temporally proximal objects are also more likely to be confused with a target than objects that are distant in either dimension, indicating a spatiotemporal coordinate system (Sapkota, Pardhan, & van der Linde, 2016). However, in these studies, the encoding of spatiotemporal properties was explicitly encouraged by using spatial or temporal information as retrieval cues or by directly testing memory for changes in spatial or temporal information. Whereas these studies show that spatiotemporal context *can* be used to access VWM, it remains unclear if time *is* used naturally in representing past visual events.

Here, we asked whether temporal information is incidentally encoded along with spatial information to provide spatiotemporal reference frames in VWM. In a variant of a change detection task, participants memorised colours (or item sizes; Exp. 5) presented sequentially with different inter-stimulus-intervals and at different locations. This task provides rich spatiotemporal information, as it might be encountered in natural environments: Each item on a given trial could be uniquely identified by its spatial or temporal coordinates and its relations in these coordinate systems to other items. However, neither spatial nor temporal information was required to perform the task, as there were never any swaps between item positions. Indeed, participants' task was always the same: to report if there was a change in one of the colours. The logic behind our approach was simple: If the task-irrelevant spatiotemporal information is incidentally encoded, memory should be impaired if that information is not available at retrieval. We systematically manipulated the retrieval context by removing either spatial or temporal information from the test array and compared performance under these conditions to a baseline condition, in which the test array contained the same spatiotemporal information that was present at encoding.

2. Experiments 1 and 2: incidental encoding of spatiotemporal properties

In a first step, we varied retrieval context conditions across trials in a randomly interleaved fashion (Experiment 1) and, in a second step, compared the interleaved variation with a variation across blocks of trials (Experiment 2) to determine if the encoding of spatiotemporal properties always occurs incidentally or if it can be strategically adjusted to predictable retrieval contexts.

2.1. Methods

2.1.1. Participants

Twenty-four volunteers participated in each experiment (Experiment 1: 21 female, mean age 22 years; Experiment 2: 16 female, mean age 25 years) for course credit or monetary compensation. The sample size for Experiment 1 was based on comparable VWM studies and then kept constant for the following experiments. All participants had normal or corrected-to-normal visual acuity and colour vision and were naive to the purpose of the experiment. The experimental protocol was approved by the ethics committees of the Faculty of Psychology at Philipps-Universität Marburg (Experiment 1) and of the Department of Psychology at Humboldt-Universität zu Berlin (Experiment 2), and conducted in accordance with the Declaration of Helsinki (2008). All participants provided informed written consent.

2.1.2. Apparatus and stimuli

Experiment 1 was conducted at Philipps-Universität Marburg. Participants sat in a dimly-lit room, facing a monitor (Samsung Syncmaster 2233, 22″, 1680 × 1050 pixels) at a distance of 104 cm. Stimulus presentation and response collection were controlled by a Windows PC using *E*-Prime 2.0 software. Experiments 2–5 were conducted at Humboldt-Universität zu Berlin, where stimuli were presented on a ViewPixx/3D monitor (24″, 1920 × 1080 pixels) at a viewing distance of 53 cm. Matlab (Mathworks, Natick, MA) and the Psychophysics Toolbox 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) were used to control stimulus presentation and response collection. Participants responded with a button press on a keyboard using their left and right index fingers. The assignment of buttons to responses (no change vs. change) was balanced across participants, randomly assigned for each person, and kept constant throughout experimental sessions.

Stimuli were presented on a grey background. On each trial, four different memory item colours were randomly selected from a set of seven colours (CIE coordinates x/y): blue (0.093/0.347), green (0.052/0.716), orange (0.478/0.441), pink (0.310/0.295), red (0.400/0.361), violet (0.231/0.288), and yellow (0.341/0.497). Colours in the test array were either identical to the memory item colours (no-change trials) or one of the colours changed to one of the remaining colours that were not memory item colours on that trial (change trials). Items were arranged on two imaginary circles around the fixation dot, at eccentricities of 4.68° and 5.23° of visual angle. There were six possible spatial configurations with one item in each quadrant and two at each eccentricity. All items were circle-shaped and 1.16° in diameter. The fixation dot subtended 0.17° and the enlarged fixation dot signalling response time 0.23°.

2.1.3. Procedure and design

The trial procedure is illustrated in Fig. 1a. Each trial started with the memory array: Participants memorised the colours of four items presented sequentially at different locations and with different inter-stimulus-intervals (ISIs), each for 100 ms. Time intervals between items were varied in six different temporal structures, which were permutations of short (100 ms), medium (250 ms) and long (400 ms) ISIs. Spatial locations were varied in six different spatial structures (see Section 2.1.2). Spatial and temporal structures were fully crossed. Thus, each item on a given trial had unique coordinates and relations to the other items in both the spatial and the temporal domain. In other words, each item could be identified based on absolute, relative or ordinal spatial or temporal information. After a 1000 ms retention interval, the test array was presented. In the spatiotemporal condition, this was identical to the memory array in its spatiotemporal structure, that is, items were presented at the same locations and sequentially with the same ISIs, each for 100 ms. In the spatial condition, temporal information was removed from the test array: Items were presented in the same spatial locations, but simultaneously for 400 ms. In the temporal condition, spatial information was removed from the test array: Items were presented sequentially with the same ISIs and each for 100 ms, but centrally around fixation. After presentation of the test array, the fixation dot was enlarged, signalling the onset of response time (10 s maximum). Participants were asked to indicate if the colours had changed from memory to test array. Importantly, there were never any swaps between spatial or temporal locations, so the task did not require participants to bind colour to spatial or temporal position. When there was a change in colour, it was to a new colour. The next trial started after a 1000 ms inter-trial-interval. Participants were instructed to maintain fixation during the trials.

Experiment 1 consisted of 864 trials, equally distributed among retrieval context conditions. Retrieval contexts were varied randomly. The six spatial structures were crossed with the six temporal structures, yielding 36 spatiotemporal structures, which were randomly selected on each trial but presented equally often for each retrieval context. Half of all trials were change trials, and a change was equally likely to occur at any of the item positions. Every 48 trials, participants had the opportunity of a short rest.

Experiment 2 was conducted in two sessions, each consisting of 864 trials: In one session, retrieval context conditions were randomly

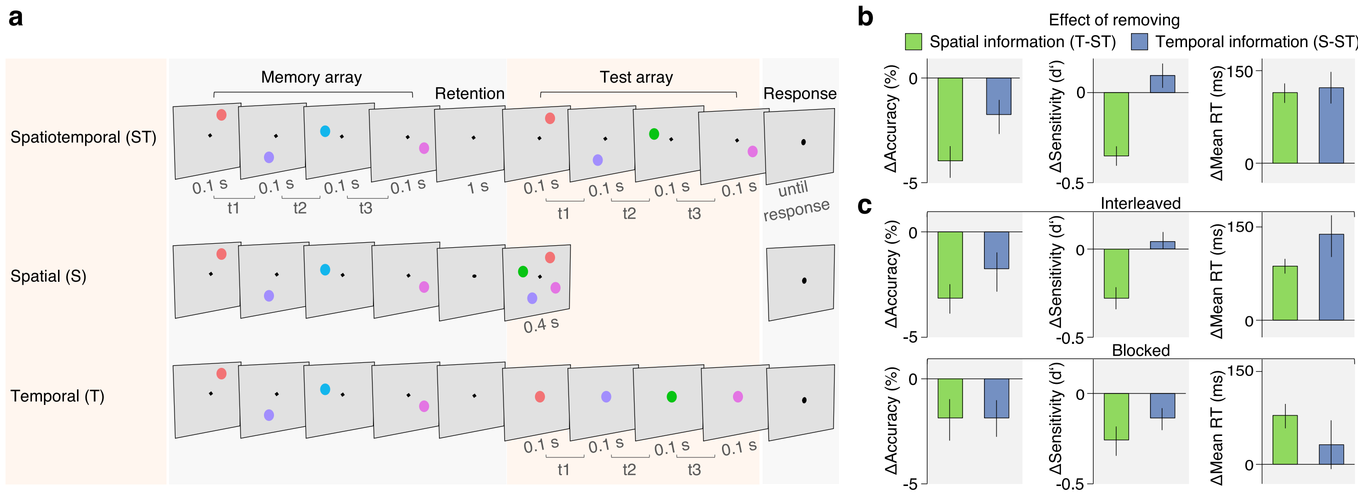


Fig. 1. Task and results of Experiments 1 and 2. (a) Trial procedure for the main retrieval context conditions. Participants memorised four colours, presented sequentially at different locations and with different inter-stimulus-intervals (t1, t2, t3), across a retention interval in order to indicate whether one of the colours had changed at test. Here, the blue item changed to green. We manipulated retrieval context (i.e., the test array): In the spatiotemporal condition (ST), the test array was identical to the memory array in its spatiotemporal structure; in the spatial condition (S), temporal information was removed at retrieval; in the temporal condition (T), spatial information was removed. Spatial or temporal positions of colours never changed within a trial. (b) Results of Experiment 1: The effect of removing temporal information (performance in the spatial condition minus performance in the spatiotemporal condition; S-ST) or spatial information (performance in the temporal condition minus performance in the spatiotemporal condition; T-ST) on accuracy in percent, sensitivity (d'), and reaction times. (c) Results of Experiment 2: The effect of removing temporal information or spatial information in a setting with a randomly interleaved or blocked manipulation of retrieval contexts. All error bars show standard errors of the means. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

interleaved. In the other, they were blocked, changing after six blocks of 48 trials. The order of sessions and the order of retrieval contexts in the blocked session were balanced across participants.

2.1.4. Data analysis

Reaction time outliers (±2.5 SD from individual mean RT) were excluded from the analyses (2.6% of all trials in Experiment 1, 2.2% in Experiment 2). To quantify memory performance, we analysed accuracy (percent correct), the sensitivity to detect a change (*d'*) and mean reaction times. Sensitivity was calculated as d' = z(hit rate) – z(false alarm rate). Rates of 0 or 1 were replaced with 0.5/n or (n − 0.5)/n, respectively, where n is the number of signal or noise trials (Stanislaw & Todorov, 1999). Only correct trials were included in the analysis of reaction times. Costs of removing spatial or temporal information were computed relative to the spatiotemporal condition and tested against zero (one-tailed *t*-tests). When appropriate, we compared the magnitude of costs associated with the removal of spatial or temporal information (two-tailed t-tests).

2.2. Results and discussion

The spatiotemporal condition (ST) served as a baseline: We subtracted measurements obtained in this condition (accuracy: 78.14 ± 1.6%; d': 1.75 ± 0.13; RT: 674 ± 39 ms; Mean ± SEM) from the spatial condition (S) to directly examine the effect of removing temporal information (S-ST), and from the temporal condition (T) to examine the effect of removing spatial information (T-ST; see Supplementary Fig. 1 for overall performance levels in all conditions and experiments). We report reaction times in addition to measures of accuracy (percent correct and *d'*), but these should be interpreted with caution, as sequential stimulus presentation at test allows for more response preparation time than simultaneous presentation. Consequently, none of our conclusions are based solely on reaction time data. A control experiment confirmed that measures of accuracy were not influenced by the additional time available during test array presentation with sequential as compared to simultaneous presentation (for details see Supplementary methods and Supplementary results).

In Experiment 1, the removal of spatial information at retrieval was associated with large costs across all measures (accuracy: $t(23) = -5.29$, $d = -1.08$, 95% CI [−1.58, −0.57]; d': $t(23) = -6.32$, $d = -1.29$, 95% CI [−1.83, −0.74]; RT: $t(23) = 7.30$, $d = 1.49$, 95% CI [0.90, 2.07]; all $p < .001$; Fig. 1b, green bars). Removing temporal information (blue bars) resulted in an equivalent cost in reaction time ($t(23) = 4.67$, $p < .001$, $d = 0.95$, 95% CI [0.46, 1.43]) and a smaller cost in accuracy ($t(23) = -2.11$, $p = .023$, $d = -0.43$, 95% CI [−0.85, −0.01]; S-ST vs. T-ST: $t(23) = 2.68$, $p = .013$, $d = 0.55$, 95% CI [0.11, 0.97]), but no loss of sensitivity ($t(23) = 1.40$, $p = .912$). The discrepancy between accuracy and sensitivity suggests a shift in the criterion to report a change. Indeed, in all experiments, we observed an overall bias towards responding no-change. Criterion differences between retrieval contexts, however, were not related to memory impairments (see Supplementary Fig. 2 for criterion values for all conditions and experiments).

These findings indicate that both spatial and temporal properties aided memory, even though they were not informative as to the colour change, let alone required to perform the task. But in this randomly interleaved setting, either type of information could be present at test and thereby prove useful in accessing the memorised information on any given trial. Thus, participants might have adopted the strategy to always encode the spatiotemporal context to some degree. In Experiment 2, therefore, we additionally varied retrieval context conditions across blocks of trials. If encoding of spatiotemporal properties can be strategically adjusted to a predictable retrieval context, colour information should only be bound to the dimension that is known to be an effective retrieval cue in a given block of trials and no, or at least smaller, costs should be observed.

Performance in the spatiotemporal baseline condition was at a similar level in the interleaved (accuracy: 76.54 ± 1.3%; d': 1.75 ± 0.11; RT: 542 ± 35 ms) and blocked (accuracy: 75.88 ± 1.8%; d': 1.74 ± 0.15; RT: 518 ± 30 ms) conditions of Experiment 2. The interleaved condition replicated the results of Experiment 1 (Fig. 1c): The removal of spatial information (S-ST) impaired memory across all measures (accuracy: $t(23) = -4.34$, $d = -0.89$, 95% CI [−1.35, −0.40]; d': $t(23) = -4.58$, $d = -0.93$, 95% CI [−1.41, −0.45]; RT: $t(23) = 8.06$, $d = 1.65$, 95% CI [1.02, 2.26]; all $p < .001$) and the removal of temporal information (T-ST) incurred costs in accuracy ($t(23) = -1.83$, $p = .04$, $d = -0.37$, 95% CI [−0.78, −0.05]) and reaction time ($t(23) = 4.22$, $p < .001$, $d = 0.86$, 95% CI [0.38, 1.33]), but not in sensitivity ($t(23) = 0.77$, $p = .78$). Costs did not differ between the spatial and temporal dimensions (accuracy: $t(23) = 1.23$, $p = .23$); RT: $t_{(23)} = 1.74$, $p = .10$). In the blocked setting, a highly similar pattern of results was obtained. In fact, for accuracy and reaction time, there was no difference between settings (accuracy: $F(1,23) = 0.25$, $p = .62$; RT: $F(1,23) = 2.31$, $p = .14$) nor an interaction of setting and the type of information removed (accuracy: $F(1,23) = 1.66$, $p = .21$; RT: $F(1,23) = 4.04$, $p = .06$). Costs were associated with removing both spatial (accuracy: $t(23) = -1.95$, $p = .032$, $d = -0.40$, 95% CI [−0.81, 0.02]; RT: $t(23) = 3.84$, $p < .001$, $d = 0.78$, 95% CI [0.32, 1.24]) and temporal information (accuracy: $t(23) = -2.16$, $p = .021$, $d = -0.44$, 95% CI [−0.86, −0.02]). Only in terms of sensitivity did the effects of removing spatial or temporal information differ between settings ($F(1,23) = 5.25$, $p = .03$, partial $\eta^2 = 0.19$): In the blocked setting, unlike in the interleaved setting, we observed a cost of removing either spatial ($t(23) = -3.10$, $p = .003$, $d = -0.63$, 95% CI [−1.07, −0.19]) or temporal information ($t(23) = -1.99$, $p = .029$, $d = -0.41$, 95% CI [−0.82, 0.02]). A reason for the rather counterintuitive observation that participants seemed to rely more on temporal information when it was known to be an ineffective retrieval cue is not obvious, but may lie in differential practice effects across conditions: Neglecting the beginning of the experiment, the interleaved condition strongly resembled the pattern observed in the blocked condition, which was fairly stable across the experiment (Supplementary Fig. 3). In any case, these findings clearly show that encoding in a spatiotemporal frame of reference is not strategically adjusted to a predictable retrieval context: Costs of removing spatial or temporal information were evident (and unscathed) even if participants had advance knowledge about which dimension would not be helpful to access memorised information.

3. Experiment 3: flexibility and stability of the weighting of spatial and temporal information

The results from the first two experiments have shown that both spatial and temporal structures are incidentally stored along with task-relevant information, but they also appear to indicate that more weight is assigned to spatial information: The removal of spatial information resulted in more consistent and larger costs. However, the particular spatial and temporal parameters that we used might have encouraged reliance on spatial information. Whereas items were spatially spaced far apart, with one item in each quadrant of the display, the time intervals between items were very brief and their differences just noticeable. A reference frame should ideally allow access to specific representations while avoiding confusion between them, and wider spacings in both the spatial and temporal dimension have repeatedly been shown to allow for better item identification and individuation across a variety of perceptual and memory tasks (e.g., Bahcall & Kowler, 1999; Emrich & Ferber, 2012; Franconeri, Jonathan, & Scimeca, 2010; Intriligator & Cavanagh, 2001; Whitney & Levi, 2011; Yeshurun & Marom, 2008). The wide spacings in space and close spacings in time might thus have resulted in a prioritisation of a spatial frame of reference over a temporal one.

In Experiment 3, we systematically varied the spatial and temporal inter-item spacings, expecting that the weighting of spatial and temporal information would shift towards the more widely spaced dimension.

3.1. Methods

Unless stated otherwise, the methods of Experiment 3 were identical to those of Experiment 2.

3.1.1. Participants

Twenty-four volunteers (20 female, mean age 24 years) participated in the experiment.

3.1.2. Apparatus and stimuli

We varied the spatial spacing by moving the central point around which the configurations were arranged from the fixation dot (wide) 5° to the left or right hemifield (medium) or to either of the four quadrants (close), and by reducing eccentricities with respect to that central point by a factor of 0.75 (medium) or 0.5 (close).

3.1.3. Procedure and design

We used three spacing levels (close, medium and wide) in each dimension and fully crossed them, while keeping the 36 spatiotemporal structures from Experiments 1 and 2 intact. The wide spatial spacing was the same as in the previous experiments, arranged around fixation. For the medium and close spacings, the centre of the configurations was shifted and eccentricities were reduced (see Section 3.1.2). In the temporal domain, the close spacing was the same as in Experiments 1 and 2. Interval durations and differences between durations were increased for the medium (100, 300 and 600 ms) and wide (100, 400 and 800 ms) spacing levels. It is essentially impossible to perfectly match spacing levels across dimensions, so we do not claim that, for instance, the medium spatial distance corresponds to the medium temporal distance.

The resulting nine combinations of spatial and temporal spacings were presented equally often and distributed equally across retrieval contexts, but randomly chosen on each trial. Retrieval contexts were varied across blocks of trials. The order of retrieval contexts was kept constant across sessions for each participant but balanced across participants. There were three sessions of 648 trials each, yielding a total of 1944 trials.

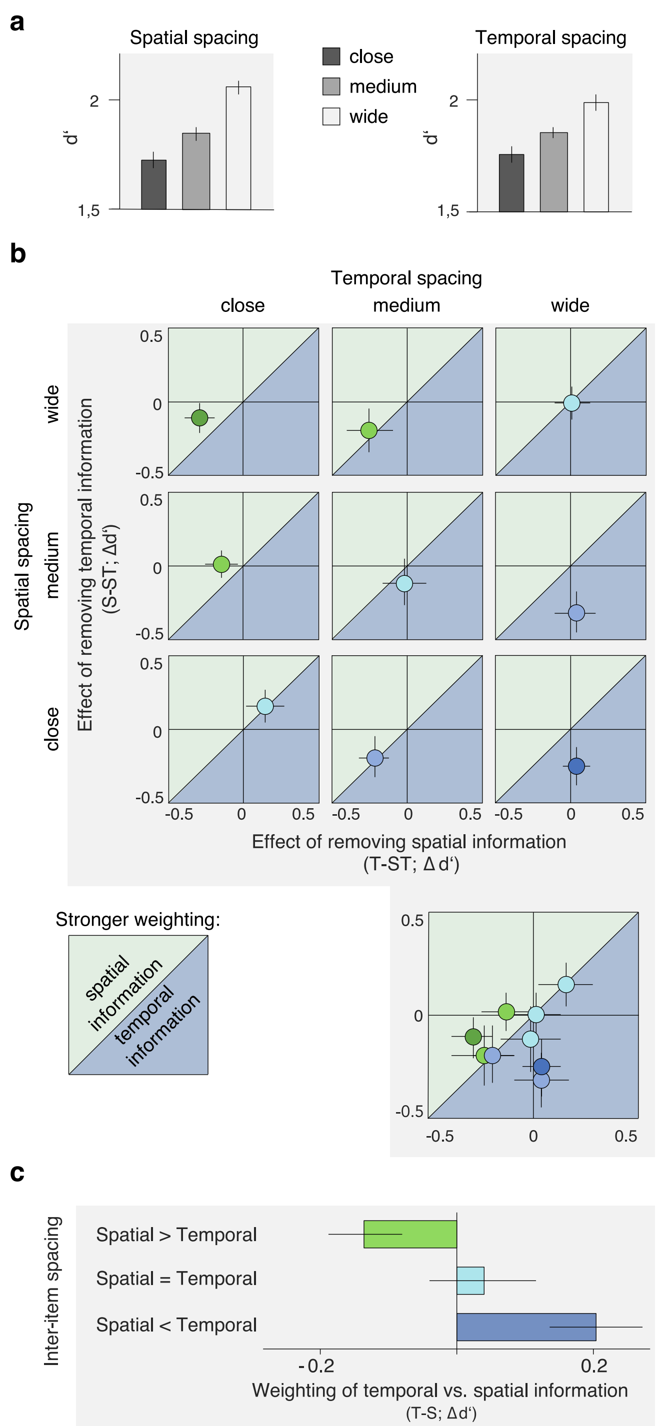
3.1.4. Data analysis

Reaction time outliers (2.1% of all trials) were excluded from the analyses.

3.2. Results and discussion

For the sake of brevity, we only report sensitivity (d') here, but analogous results were observed for accuracy and reaction times (see Supplementary Results). The key prediction that inter-item spacings affect the relative weighting of spatial and temporal information relied on the assumption that memory would benefit from a more widely spaced item representation in either reference frame. Indeed, the sensitivity to detect a colour change increased with both a wider spatial spacing (Fig. 2a; $F(2,46) = 24.65$, $p < .001$, partial $\eta^2 = 0.52$) as well as with a wider temporal spacing ($F(2,46) = 12.88$, $p < .001$, partial $\eta^2 = 0.36$). More importantly, different combinations of spacing levels modulated the weighting of spatial and temporal information (performance in the temporal condition minus performance in the spatial condition) as predicted ($F(8,184) = 3.09$, $p = .003$, partial $\eta^2 = 0.118$; Fig. 2b). This is most evident when looking at those combinations with the largest differences between spatial and temporal spacings. When items were far apart in space but close in time (Fig. 2b, top left panel), the removal of spatial information at retrieval incurred a larger cost than the removal of temporal information ($t(23) = 2.17$, $p = .04$, $d = 0.44$, 95% CI [0.02, 0.86]), replicating the results from the previous experiments. In the opposite situation, by contrast, when items were close in space and widely separated in time (bottom right panel), temporal information was weighted more strongly ($t(23) = 2.71$, $p = .013$, $d = 0.55$, 95% CI [0.12, 0.98]). In fact, while the removal of temporal information resulted in a loss in sensitivity in this condition ($t(23) = 1.88$, $p = .037$, $d = -0.38$, 95% CI [−0.79, 0.04]), removing spatial information had no effect ($t(23) = 0.40$, $p = .65$).

In Fig. 2b, it is apparent that we observed no costs of removing information from either dimension in the three conditions, in which spatial and temporal spacings were approximately matched (turquoise values). This might point to certain boundary conditions, under which spatial and temporal context no longer aid memory (e.g., severe crowding or masking at very close spacings or absence of relational



(caption on next column)

Fig. 2. Results of Experiment 3. (a) Sensitivity (d') for the different levels (close, medium, wide) of spatial and temporal spacings. (b) The effect of removing spatial (T-ST) and temporal information (S-ST) on sensitivity (d') for different combinations of spatial and temporal spacings. Values falling within the green-shaded area indicate a stronger weighting of spatial information, meaning that the removal of spatial information incurred larger costs or smaller benefits than the removal of temporal information. Conversely, the blue-shaded area indicates a stronger weighting of temporal information. Colours represent the relative spacing of spatial and temporal structures: spatial > temporal (green), spatial = temporal (turquoise) and spatial < temporal (blue). The bottom right panel shows all values gathered in one plot. (c) The weighting of temporal and spatial information (T-S; Δd') as a function of the relative inter-item spacing, that is, averaged across spacing conditions in which the spatial spacing was larger (green values in 2B), approximately equal to (turquoise values) or smaller (blue values) than the temporal spacing. Negative values indicate a stronger weighting of spatial information, positive values a stronger weighting of temporal information. All error bars show standard errors of the means. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

coding at wide spacings). However, this is not a critical matter for the present purpose of demonstrating that time can be as important a reference frame – or even more so – as space, because the weighting of spatial and temporal information shifts as a function of inter-item spacing.

To highlight this shift in the weighting of spatial and temporal information, we averaged the difference in sensitivity in the temporal and spatial conditions (T-S) across spacing conditions in which the spatial spacing was larger than, equal to, or smaller than the temporal spacing (green, turquoise, and blue values, respectively, in Fig. 2). With this difference measure, negative values indicate a stronger weighting of spatial information, and positive values a stronger weighting of temporal information. As can be seen in Fig. 2c, there was a stronger weighting of spatial information with wider spatial spacings ($t(23) = 2.54$, $p = .009$, $d = -0.518$, 95% CI [−0.94, −0.09]) and a stronger weighting of temporal information with wider temporal spacings ($t(23) = 2.97$, $p = .003$ $d = 0.61$, 95% CI [0.16, 1.04]). No preference was observed for equal spacing levels ($t(23) = 0.50$, $p = .62$), suggesting that the levels were approximately matched across dimensions.

These findings demonstrate that space does not always trump time in VWM, but that there are, in fact, situations in which temporal context is more important than spatial context. More generally, they indicate that storage in spatial and temporal reference frames depends on the usefulness of either domain for differentiating between items and accessing specific memory representations.

Across all experiments, we observed that individuals greatly differed in their use of spatial and temporal information, not only in the extent to which they made use of the task-irrelevant spatiotemporal structure but also in the relative weighting of spatial and temporal information. To examine if such preferences for representing information in either domain might be stable over time, we analysed the data separately for each of the three identical sessions of Experiment 3, which were separated by at least one day (six days on average). Individual weightings of spatial and temporal information (T-S; averaged across spacing levels; Fig. 3) remained fairly consistent from session 1 to session 2 (d': $r = 0.293$, $p = .082$; accuracy: $r = 0.495$, $p < .01$) and even more so from session 2 to session 3 (d': $r = 0.578$, $p < .01$; accuracy: $r = 0.556$, $p < .01$), when participants had gained more experience with the task.

4. Are space and time special?

So far, our findings do not necessarily imply that space and time play a special role in the architecture of VWM that distinguishes them from other feature dimensions. First, the impairments of memory performance could also be interpreted in terms of discriminability: Each item could be identified based on its temporal or spatial locations, so taking either information away likely reduced discriminability. The same

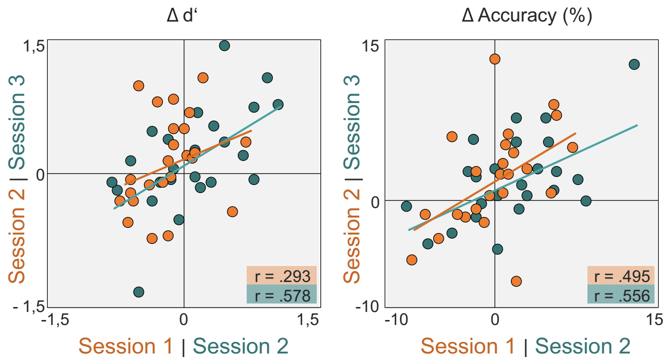


Fig. 3. Stability of individual weightings (T-S) of spatial and temporal information across sessions of Experiment 3 (in orange from session 1 to session 2; in dark turquoise from session 2 to session 3). Negative values indicate a stronger weighting of spatial information, positive values a stronger weighting of temporal information. The left panel shows sensitivity (d'), the right panel accuracy in percent. Each dot in each colour represents one of the 24 participants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

might hold for any other feature that similarly allows to distinguish between items. Second, costs in the temporal and spatial conditions might have been observed because, unlike in the spatiotemporal condition, there was an inconsistency between encoding and retrieval (*encoding specificity principle*; Thomson & Tulving, 1973). Conceivably, any other inconsistency would produce similar effects. To rule out these two alternative accounts, we introduced additional task-irrelevant feature dimensions.

4.1. Experiment 4: variations in space, time and size

In Experiment 4, we additionally included variations along a third, easily discriminable feature dimension: size. At encoding, memory items varied not only in their temporal and spatial locations, but also in their size (in the following abbreviated as D for diameter, to avoid confusion with the spatial dimension, abbreviated as S). The test array was then either identical to the memory array with respect to spatial, temporal and size information (STD), or we removed information from one dimension by presenting items in the same location (TD), at the same time (SD) or in the same size (ST).

4.1.1. Methods

4.1.1.1. Participants. Twenty-eight volunteers participated in the experiment. Four participants were excluded, because their performance did not exceed chance level. The analyses were performed on the remaining twenty-four participants (17 female, mean age 25 years).

4.1.1.2. Apparatus and stimuli. Four different item sizes were chosen on each trial from two size sets varied across sessions: In the set with small size differences, steps between items were 1/6 of the area of the original item size (0.94°, 1.05°, 1.16°, 1.25°, 1.33° in diameter) and in the set with large size differences, steps were 1/3 of the area of the original item size (0.67°, 0.94°, 1.16°, 1.34°, 1.49° in diameter).

4.1.1.3. Procedure and design. Memory items differed not only in spatial and temporal location, but also in size. At test, one type of information was removed, yielding four retrieval context conditions: spatiotemporal+size (STD), spatiotemporal (ST), spatial + size (SD) and temporal + size (TD). Size differences between items were small or large, varied across sessions. Retrieval context conditions were varied across blocks of trials. The order of sessions and of retrieval contexts within sessions was balanced across participants. The experiment consisted of 1728 trials (216 per retrieval context condition; 864 per session).

4.1.1.4. Data analysis. Reaction time outliers (2.1% of all trials) were excluded from the analyses.

4.1.2. Results and discussion

In order to test if there were similar spacing effects as we observed for time and space in Experiment 3, size differences were either small or large. However, there was no effect of the magnitude of size differences (accuracy: $F(1,23) = 0.04$, $p = .847$) or an interaction with retrieval context ($F(3,69) = 0.33$, $p = .801$), so we collapsed the data across size conditions. The condition including all types of information at retrieval (STD) served as a baseline (accuracy: 75.82 ± 1.6%; d': 1.65 ± 0.13; RT: 569 ± 45 ms). If the memory decrements we observed for removing temporal or spatial information were due to reduced discriminability or a general encoding specificity effect, similar costs should be observed when size information is not available at test. But while the removal of temporal or spatial information led to equivalent costs in accuracy (Fig. 4a; SD-STD: $t(23) = -1.84$, $p = .039$, $d = -0.38$, 95% CI [−0.79, 0.04]; TD-STD: $t(23) = -2.91$, $p = .004$, $d = -0.59$, 95% CI [−1.02, −0.15]; SD-STD vs. TD-STD: $t(23) = 1.04$, $p = .308$), and the removal of spatial information additionally to costs in sensitivity ($t(23) = -3.92$, $p < .001$, $d = -0.80$, 95% CI [−1.25, −0.33]) and reaction time ($t(23) = 2.51$, $p = .01$, $d = 0.51$, 95% CI [0.08, 0.93]), no costs were associated with removing size information (accuracy: $t(23) = 1.87$, $p = .963$; d': $t(23) = 0.99$, $p = .835$; RT: $t(23) = -0.29$, $p = .388$). If anything, we observed effects in the opposite direction: Performance tended to be better when size information was not present at retrieval. Notably, this also affects the magnitude of the reported costs of removing spatial or temporal information, because the condition including size information (STD) served as baseline in this experiment. Costs with respect to the spatiotemporal condition (ST), which was used as baseline in the previous experiments, were accordingly larger (see Supplementary Fig. 1).

4.2. Experiment 5: variations in space, time, size and colour

The findings from Experiment 4 indicate that the costs of removing spatial or temporal information were not due to reduced item discriminability or an inconsistency between encoding and retrieval. Yet they are not sufficient to allow us to conclude that time and space have a special status. For one, variations in three task-irrelevant feature dimensions may simply have been too much, exceeding capacity limitations or discouraging their encoding to avoid an unnecessarily high memory load. Moreover, size differences may not be as easily discriminable and helpful as other features. In Experiment 5, we addressed these issues by additionally manipulating encoding context and switching memorised and task-irrelevant feature dimensions: Participants had to remember colours while items could also vary in size, as in Experiment 4; or they memorised item sizes and colours were task-irrelevant (see Fig. 4b). At encoding, items varied in the to-be-memorised feature as well as in two task-irrelevant dimensions: space and time, space and size/colour, or time and size/colour. Retrieval contexts were either identical to encoding contexts, or they were reduced to a spatial or temporal context by removing the second task-irrelevant feature dimension (time, space, size or colour).

4.2.1. Methods

4.2.1.1. Participants. Twenty-nine volunteers participated in the experiment. Five participants were excluded, because their performance did not exceed chance level. The analyses were performed on the remaining twenty-four participants (15 female, mean age 23 years).

4.2.1.2. Apparatus and stimuli. In the size session, four different to-be-memorised sizes were randomly selected from a set of five sizes

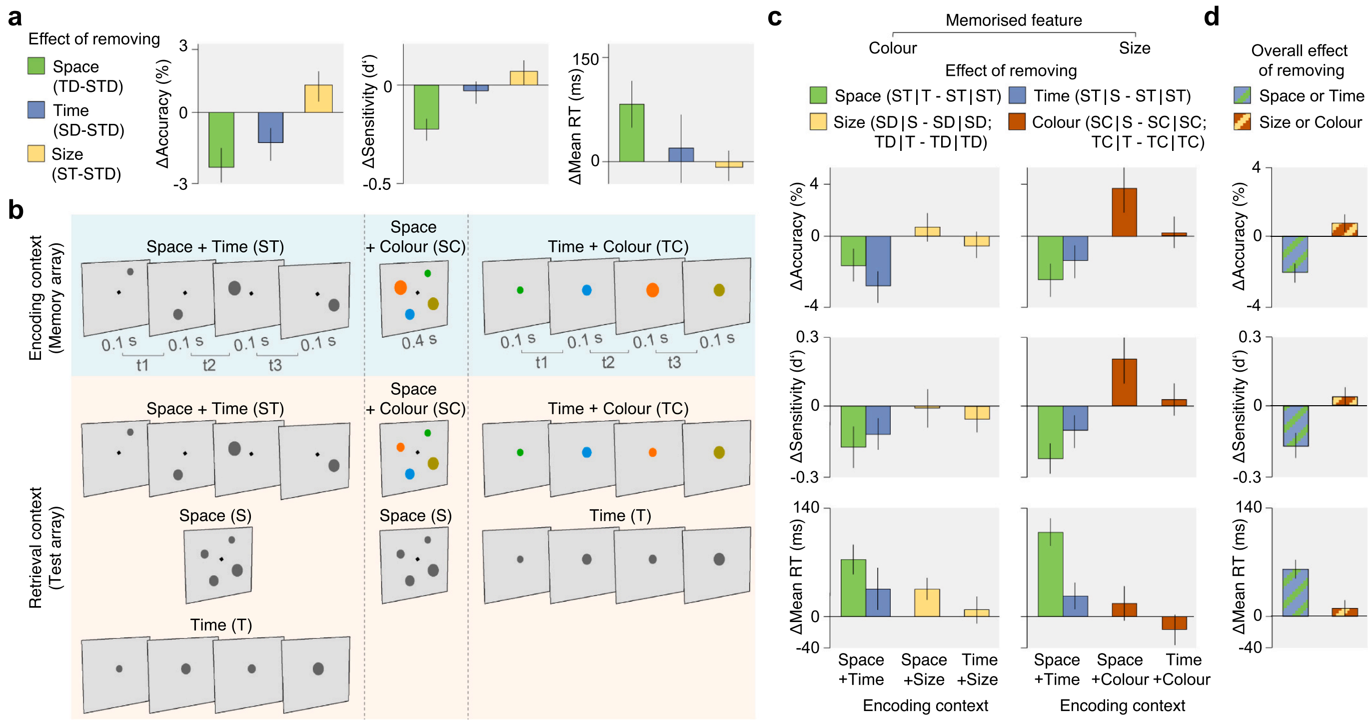


Fig. 4. Task and results of Experiments 4 and 5. (a) Results of Experiment 4: The effects of removing spatial (TD-STD), temporal (SD-STD) or size information (ST-STD) on accuracy in percent, sensitivity (d') and reaction times. (b) Trial procedure for the different encoding and retrieval context conditions of Experiment 5, illustrated for the version with size as to-be-memorised feature. At encoding, items varied in two task-irrelevant feature dimensions: space and time (ST), space and colour (SC) or time and colour (TC). Retrieval contexts were either identical to encoding context (encoding|retrieval; ST|ST, SC|SC, TC|TC), or they were reduced to spatial or temporal information by removing the other task-irrelevant dimension (time, space or colour). (c) Results of Experiment 5: Effects of removing space, time, size or colour on accuracy in percent, sensitivity (d') and reaction times, computed relative to the respective baseline (i.e., condition with identical encoding and retrieval context: ST|ST, SD|SD, SC|SC, TD|TD, TC|TC). (d) Overall effects of removing space or time and size or colour, averaged across encoding contexts and memorised features. All error bars show standard errors of the means.

(0.67°, 0.94°, 1.16°, 1.34°, 1.49° in diameter) on each trial. In change trials, one of the items changed to the remaining size not selected for a memory item on the respective trial. Task-irrelevant item colours or sizes were chosen in the same manner as when either were the to-be-memorised features.

4.2.1.3. Procedure and design. In two sessions on separate days, participants memorised either item colours or item sizes. At encoding, memory items additionally differed in two task-irrelevant features: space and time (ST), space and size/colour (SD/SC) or time and size/colour (TD/TC). Retrieval contexts were either identical to encoding contexts (encoding|retrieval: ST|ST, SD|SD, SC|SC, TD|TD, TC|TC) or they were reduced to spatial or temporal variations by removing the second task-irrelevant feature dimension (time, space, size or colour; ST|S, ST|T, SD|S, SC|S, TD|T, TC|T). Encoding and retrieval context conditions were varied across blocks of trials. The order of sessions, encoding contexts and retrieval contexts within encoding context conditions were balanced across participants. Colour and size sessions each consisted of 840 trials (120 per encoding/retrieval context condition), yielding 1680 trials in total. The time intervals between items were permutations of 100, 300 and 600 ms, as in the medium temporal spacing condition of Experiment 3.

4.2.1.4. Data analysis. Reaction time outliers (2.5% of all trials) were excluded from the analyses.

4.2.2. Results and discussion

The effects of removing specific feature dimensions at retrieval were computed relative to the respective baseline (ST|ST, SD|SD, SC|SC, TD|TD, TC|TC). With colour as memorised feature (left panels in Fig. 4c), the removal of both spatial and temporal information incurred costs in accuracy (ST|T − ST|ST: $t(23) = -1.80$, $p = .043$, $d = -0.37$, 95% CI [−0.78, 0.05]; ST|S − ST|ST: $t(23) = -3.06$, $p = .003$, $d = -0.62$, 95% CI [−1.06, −0.18]; ST|ST: 76.97 ± 1.52%) and sensitivity (ST|T − ST|ST: $t(23) = -1.99$, $p = .029$, $d = -0.41$, 95% CI [−0.82, 0.02]; ST|S − ST|ST: $t(23) = -1.79$, $p = .043$, $d = -0.37$, 95% CI [−0.78, 0.05]; ST|ST: 1.71 ± 0.10). The removal of spatial information was additionally associated with a cost in reaction time (ST|T − ST|ST: $t(23) = 3.89$, $p < .001$, $d = 0.79$, 95% CI [0.33, 1.25]; ST|S − ST|ST: $t(23) = 1.30$, $p = .104$; ST|ST: 594 ± 40 ms). The removal of size information, by contrast, only delayed reaction times with spatial contexts (SD|S − SD|SD: $t(23) = 2.48$, $p = .011$, $d = 0.51$, 95% CI [0.08, 0.93]; SD|SD: 463 ± 39 ms). There was no delay with temporal contexts (TD|T − TD|TD: $t(23) = 0.47$, $p = .323$; TD|TD: 672 ± 42 ms) and no costs in accuracy (SD|S − SD|SD: $t(23) = 0.62$, $p = .731$; SD|SD: 78.26 ± 1.10%; TD|T − TD|TD: $t(23) = -0.76$; $p = .227$; TD|TD: 70.06 ± 1.72%) or sensitivity (SD|S - SD|SD: $t(23) = -0.08$; $p = .468$; SD|SD: 1.96 ± 0.10; TD|T − TD|TD: $t(23) = -1.01$, $p = .161$; TS|TS: 1.19 ± 0.12).

While overall performance was somewhat worse (see also Fig. S1) – yet well above chance – when size was memorised and colour task-irrelevant, a similar pattern emerged. The removal of spatial information incurred costs across all measures (accuracy: ($t(23) = -2.64$, $p = .007$, $d = -0.54$, 95% CI [−0.96, −0.11]; ST|ST: 63.99% ± 1.34%; d': $t(23) = -3.48$, $p = .001$, $d = -0.71$, 95% CI [−1.15, −0.26]; ST|ST: 0.94 ± 0.10; RT: $t(23) = 6.16$, $p < .001$, $d = 1.26$, 95% CI [0.71, 1.79]; ST|ST: 642 ± 57 ms). The removal of temporal information reduced sensitivity ($t(23) = -2.43$, $p = .12$, $d = -0.50$, 95% CI [−0.92, −0.07]) and impairments were also observed in terms of accuracy and reaction time, but these failed to reach significance at the .05 level (accuracy: $t(23) = -1.46$; $p = .079$; RT: $t(23) = 1.54$, $p = .069$). No costs whatsoever were associated with the removal of colour information from spatial contexts (SC|S − SC|SC; accuracy: $t(23) = 2.04$, $p = .974$; SC|SC: 64.72 ± 1.54%; d': $t(23) = 1.92$, $p = .966$; SC|SC: 0.98 ± 0.11; RT: $t(23) = 0.77$, $p = .224$; SC|SC: 485 ± 47 ms) or temporal contexts (TC|T − TC|TC; accuracy: $t(23) = 0.20$, $p = .557$; TC|TC: 61.73 ± 1.51%; d': $t(23) = 0.40$, $p = .654$; TC|TC: 0.74 ± 0.10; RT: $t(23) = -0.84$, $p = .794$; TC|TC: 727 ± 67 ms).

Thus, as in Experiment 4, we observed consistent – yet rather small – impairments in memory performance when either spatial or temporal information was taken away at retrieval, but no costs (or instead even benefits) when a different task-irrelevant feature-dimension was removed. To illustrate this pattern, we averaged the effects of removing space or time and the effects of removing size or colour across memorised features and encoding contexts (Fig. 4d). Importantly, this pattern cannot be accounted for in terms of capacity limitations, because items were always defined by three different feature dimensions.

With this design, moreover, memory performance can be taken as a proxy of feature discriminability within dimensions, because we used the same item colours and sizes as in the previous experiments as either memorised or task-irrelevant feature dimension. Discriminability of the different colour categories is unquestionably high (average performance levels reached about 80% in all previous experiments; see also Supplementary Fig. 1), and memory performance for item sizes (as used in Experiments 4 and 5) confirmed that their discriminability was sufficiently high as well (i.e., to potentially provide a reference frame).

Together, these results may be taken to indicate that time and space, indeed, have a special status in VWM – if not qualitatively, then at least quantitatively, in that they are prioritized over other context dimensions given equal task-(ir)relevance.

5. General discussion

Space and time structure our visual experience, yet our understanding of how the temporal characteristics of visual events structure their short-term retention in memory lags far behind our understanding of the role of spatial aspects. Here, we have shown that both spatial and temporal properties are encoded along with to-be-memorised information, even when these properties are entirely task-irrelevant. Removing spatial or temporal information at retrieval impaired memory, while taking away different task-irrelevant feature dimensions at test (size or colour) did not produce similar effects. Encoding in a spatiotemporal reference frame occurred incidentally and was not strategically adjusted to a predictable retrieval context that relied on only one of the dimensions. The relative weighting of spatial and temporal information, however, was modulated by the inter-item spacing in either domain. A greater reliance on the reference frame in which items are more widely spaced likely allows for better item individuation and reduces interference or confusion between representations (e.g., Bahcall & Kowler, 1999; Emrich & Ferber, 2012; Franconeri et al., 2010; Intriligator & Cavanagh, 2001; Whitney & Levi, 2011; Yeshurun & Marom, 2008). Consequently, under certain conditions, temporal context matters more than spatial context. Individuals differed considerably in their weighting of spatial and temporal information, and these inter-individual differences were a stable preference that persisted across sessions.

For this series of experiments, we wanted to use rich spatiotemporal contexts, as they are typically provided in natural environments. To this end, we chose spatial and temporal parameters that ensured that each item could be uniquely identified based not only on its absolute position in space or time, but also its spatial and temporal relations to other items. This approach increases ecological validity, but its drawback is that, as of yet, we do not know specifically which aspects of the temporal structure constitute temporal reference frames in VWM – absolute, relational and/or ordinal properties. Based on previous findings for spatial information (e.g., Boduroglu & Shah, 2009; Jiang et al., 2000; Olson & Marshuetz, 2005), it seems reasonable to assume that it is not so much absolute position but relative coding – including both the distances to other items and item order – that is critical.

In a nutshell, our results show that visual events are represented in their spatiotemporal context. That context is important for memory is by no means a new idea (e.g., Baddeley, 2003; Eich, 1985; Howard & Kahana, 2002; Thomson & Tulving, 1973; Tulving, 1974), albeit it has mostly been investigated for verbal material. But this tradition of research on contextual dependencies in memory has only had little impact on studies on VWM and its conceptualizations. The most influential models of VWM capacity – variants of slot or resource models (e.g., Awh, Barton, & Vogel, 2007; Bays & Husain, 2008; Luck & Vogel, 2013; van den Berg, Shin, Chou, George, & Ma, 2012; Zhang & Luck, 2008) – more or less explicitly assume items to be represented independently, and do not account for inter-item and context effects (but see Swan & Wyble, 2014). This notion does not only seem unlikely to reflect mnemonic processes in natural environments, but also contrasts with a growing body of evidence demonstrating the existence of such effects even for simple visual arrays (e.g., Brady & Alvarez, 2011, 2015a, 2015b; Brady et al., 2019; Lew & Vul, 2015; Liesefeld, Liesefeld, & Müller, 2018; Schurgin & Brady, 2019).

One recent model does take context effects into account and resonates well with our findings: According to the interference model proposed by Oberauer and Lin (2016), memorised features (e.g., colours) are bound to their location on a context dimension in two-dimensional binding space, with limited precision in both the feature and the context dimension. Typically, this context dimension is space – the dominant reference system in VWM research – but the model would incorporate time equally well. Access to memory contents relies on an activation of the target location in the context dimension and the strengths of its bindings with features in binding space. This implies that a close spacing of items in a context dimension leads to greater confusion, because of overlapping activations in binding space.

The interference model clearly relates to our findings in many regards, but its focus is on contexts that are used as retrieval cues and are therefore task-relevant. Retrieval cue context refers to the dimension that is used to probe the memorised feature – in most cases, this is spatial location, but it can also be a nonspatial dimension such as colour (e.g., Kalogeropoulou, Jagadeesh, Ohl, & Rolfs, 2017; Schneegans & Bays, 2017; van Ede, Niklaus, & Nobre, 2017). In the present study, by contrast, we specifically designed the task so that it did not encourage the use of spatial or temporal location as retrieval cues: Participants were aware that any change in colour always involved a new colour but never a swap between items. This means that colour (or size; Exp. 5) was the only dimension required for successful performance and theoretically did not need to be bound to any context dimension.

Admittedly, it is hard to imagine how single features could be stored as free-floating units without any differentiating context. It is thus likely, that *some* context is always stored along with to-be-memorised information. Our findings indicate that time and space are the preferred dimensions in this regard: Bindings to task-irrelevant spatial and temporal locations – but not size or colour – were still formed, even in trials in which either dimension was predictably an ineffective retrieval cue. Encoding in a spatiotemporal reference frame may constitute a minimum context, in which information is incidentally represented. In this mode, the weighting of spatial and temporal information depends on the overlap of representations in space and time. The amount of overlap may result from individual discrimination thresholds in the two domains, giving rise to pronounced and stable inter-individual differences. We assume, however, that the encoding of spatiotemporal properties is susceptible to specific task demands: The strength with which spatial and temporal properties are bound to memory contents as well as their relative weighting will drastically change as soon as other dimensions become task-relevant, or either space or time is rendered more important than the other. Accordingly, the maintenance of additional context dimensions (size, colour, orientation etc.) as reference frames should occur only if they are somehow relevant for a task at hand (Marshall & Bays, 2013; Sun & Gordon, 2010), manifesting as ensemble representations and inter-item dependencies along these dimensions (e.g., Brady & Alvarez, 2011, 2015a, 2015b; Brady et al., 2019; Lew & Vul, 2015; Liesefeld et al., 2018; Schurgin & Brady, 2019).

In sum, we suggest that time serves a similar function as space in VWM: It provides a frame of reference in which objects or features are bound (Manohar et al., 2017; Schneegans & Bays, 2018), thereby aiding maintenance and facilitating access to the memoranda. While the precise nature of spatial and temporal reference frames has yet to be determined, we believe that our findings open up a new avenue of research into the role of temporal aspects in the representational architecture of VWM, leading to a more naturalistic conceptualization of this system.

Declaration of Competing Interest

The authors declare no competing interest.

Acknowledgements

This research was supported by a DFG research grant to A.H. and M. R. (HE 8207/1-1 and RO 3579/11-1) and by the DFG's Heisenberg program (RO 3579/8-1 and RO 3579/12-1). We thank Jan-Nikolas Klanke, Lea Krätzig, Julius Krumbiegel, Tobias Richter, Olga Shurygina, Aaron Vetter and Hannah Wnendt for assistance with data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104526.

References

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*, 622–628.
Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839.
Bahcall, D. O., & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision Research, 39*, 71–86.
Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*, 851.
Bellmund, J. L. S., Deuker, L. D., & Doeller, C. F. (2019). Mapping sequence structure in the human lateral entorhinal cortex. *eLife, 8*, e45333.
Boduroglu, A., & Shah, P. (2009). Effects of spatial configurations on visual change detection: An account of bias changes. *Memory & Cognition, 37*, 1120–1131.
Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory. *Psychological Science, 22*, 384–392.
Brady, T. F., & Alvarez, G. A. (2015a). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision, 15*(6), 1–24.
Brady, T. F., & Alvarez, G. A. (2015b). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 41*, 921–929.
Brady, T. F., Störmer, V. S., Shafer-Skelton, A., Williams, J. R., Chapman, A. F., & Schill, H. M. (2019). Scaling up visual attention and visual working memory to the real world. *Psychology of Learning and Motivation, 70*, 29–69.
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.
Delogu, F., Nijboer, T. C. W., & Postma, A. (2012). Binding "when" and "where" impairs temporal, but not spatial recall in auditory and visual working memory. *Frontiers in Psychology, 3*, 62.
Eich, E. (1985). Context, memory and integrated item/context imagery. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 11*, 764–770.
Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision, 12*(12), 1–16.
Foster, J. J., Bsales, E. M., Jaffe, R. J., & Awh, E. (2017). Alpha-band activity reveals spontaneous representations of spatial position in visual working memory. *Current Biology, 27*, 3216–3223.
Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science, 21*, 920–925.
Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 32*, 58–69.
Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 31–47.
Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*, 269–299.
Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*, 171–216.
Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 26*, 683–702.
Kalogeropoulou, Z., Jagadeesh, A. V., Ohl, S., & Rolfs, M. (2017). Setting and changing feature priorities in visual short-term memory. *Psychonomic Bulletin & Review, 24*, 453–458.
Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3. *Perception, 36*.
Lew, T. F., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision, 15*(10), 1–14.
Liesefeld, H. R., Liesefeld, A. M., & Müller, H. J. (2018). Two good reasons to say "change!" – Ensemble representations as well as item representations impact standard measures of VWM capacity. *British Journal of Psychology, 110*, 328–356.
Luck, S. J. (2008). Visual short-term memory. In S. J. Luck, & A. Hollingworth (Eds.), *Visual Memory* (pp. 43–85). Oxford University Press.
Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*, 391–400.
Manohar, S. G., Pertzov, Y., & Husain, M. (2017). Short-term memory for spatial, sequential and duration information. *Current Opinion in Behavioral Sciences, 17*, 20–26.
Marshall, L., & Bays, P. M. (2013). Obligatory encoding of task-irrelevant features depletes working memory resources. *Journal of Vision, 13*(21), 1–13.
Oberauer, K., & Lin, H. (2016). An interference model of visual working memory. *Psychological Review, 124*, 21–59.
Olson, I. R., & Marshuetz, C. (2005). Remembering "what" brings along "where" in visual working memory. *Perception & Psychophysics, 67*, 185–194.
Pertzov, Y., & Husain, M. (2014). The privileged role of location in visual working memory. *Attention, Perception, & Psychophysics, 76*, 1914–1924.
Rajsic, J., & Wilson, D. E. (2014). Asymmetrical access to color and location in visual working memory. *Attention, Perception, & Psychophysics, 76*, 1902–1913.
Rondina, R., Curtiss, K., Meltzer, J. A., Barense, M. D., & Ryan, J. D. (2016). The organisation of spatial and temporal relations in memory. *Memory, 25*, 436–449.
Ryan, J. D., & Villate, C. (2009). Building visual representations: The binding of relative spatial relations across time. *Visual Cognition, 17*, 254–272.
Sapkota, R. P., Pardhan, S., & van der Linde, I. (2016). Spatiotemporal proximity effects in visual short-term memory examined by target–nontarget analysis. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 42*, 1304–1315.
Schneegans, S., & Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *Journal of Neuroscience, 37*, 3913–3925.
Schneegans, S., & Bays, P. M. (2018). New perspectives on binding in visual working memory. *British Journal of Psychology, 110*, 207–244.
Schurgin, M. W., & Brady, T. F. (2019). When "capacity" changes with set size: Ensemble representations support the detection of across-category changes in visual working memory. *Journal of Vision, 19*(3), 1–3.
Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics, 78*, 1839–1860.
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*, 137–149.
Sun, H. M., & Gordon, R. (2009). The effect of spatial and nonspatial contextual information on visual object memory. *Visual Cognition, 17*, 1259–1270.
Sun, H. M., & Gordon, R. D. (2010). The influence of location and visual features on visual object memory. *Memory & Cognition, 38*, 1049–1057.
Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophysics, 76*, 2136–2157.
Thomson, D. M., & Tulving, E. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.
Timm, J. D., & Papenmeier, F. (2019). Reorganization of spatial configurations in visual working memory. *Memory & Cognition*. https://doi.org/10.3758/s13421-019-00944-2.
Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition, 34*, 1704–1719.
Tulving, E. (1974). Cue-dependent forgetting. *American Scientist, 62*, 74–82.
van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences, 109*, 8780–8785.
van Ede, F., Niklaus, M., & Nobre, A. C. (2017). Temporal expectations guide dynamic prioritization in visual working memory through attenuated α oscillations. *Journal of Neuroscience, 37*, 437–445.
Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15*, 160–168.
Yeshurun, Y., & Marom, G. (2008). Transient spatial attention and the perceived duration of brief visual events. *Visual Cognition, 16*, 826–848.
Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 233–235.