Psychon Bull Rev
DOI 10.3758/s13423-016-1204-5

BRIEF REPORT

Decoding working memory content from attentional biases

Emma Wu Dowd[1,2] · **John M. Pearson**[1] · **Tobias Egner**[1]

Abstract What we are currently thinking influences where we attend. The finding that active maintenance of visual items in working memory (WM) biases attention toward memory-matching objects—even when WM content is irrelevant for attentional goals—suggests a tight link between WM and attention. To test whether this link is reliable enough to infer specific WM content from measures of attentional bias, we applied multivariate pattern classification techniques to response times from an unrelated visual search task during a WM delay. Single-trial WM content was successfully decoded from incidental attentional bias within an individual, highlighting the specificity and reliability of the WM-attention link. Furthermore, classifiers trained on a group of individuals predicted WM content in another, completely independent individual—implying a shared cognitive mechanism of memory-driven attentional bias. The existence of such classifiers demonstrates that memory-based attentional bias is both a robust and generalizable probe of WM.

Keywords Working memory · Visual attention · Multivariate pattern classification · Attentional bias

Electronic supplementary material The online version of this article (doi:10.3758/s13423-016-1204-5) contains supplementary material, which is available to authorized users.

✉ Emma Wu Dowd
dowd.45@osu.edu

[1] Department of Psychology & Neuroscience, Center for Cognitive Neuroscience, and Duke Institute for Brain Sciences, Duke University, Durham, USA

[2] Department of Psychology, The Ohio State University, 1827 Neil Ave, Columbus, OH 43210, USA

Our thoughts guide our behavior, as exemplified by the process of visual search—finding targets among distractors—in which a mental representation of the target is thought to be held in working memory (WM; i.e., the active maintenance and manipulation of internal information), facilitating the detection of target-matching features in the visual environment (e.g., Desimone & Duncan, 1995). Moreover, this link between WM and visual attention (i.e., the selective processing of a subset of information in the visual field) is not limited to intentional uses of WM, as it has been shown repeatedly that maintaining *any* information in WM, even information unrelated to a search target, is sufficient to bias visual attention to memory-matching items in the environment (for review, see Soto, Hodsoll, Rotshtein, & Humphreys, 2008).

Specifically, studies of memory-based attentional guidance typically use dual-task paradigms, in which participants remember an item (e.g., a colored shape) while performing an intervening but unrelated visual search during the delay period between the WM cue and subsequent probe (e.g., Soto, Heinke, Humphreys, & Blanco, 2005). Critically, the memory item can reappear in the search display, either coinciding with the target location (i.e., "valid") or with a distractor location (i.e., "invalid"); or fail to reappear in the display at all (i.e., "neutral"). The canonical finding is that search is speeded when the memory item matches the target location and slowed when the memory item matches a distractor location (relative to neutral; see Soto et al., 2008). This attentional bias by WM is clearly unintentional, as it occurs even under conditions where it is reliably detrimental to search performance (e.g., Dowd, Kiyonaga, Beck, & Egner, 2015; Kiyonaga, Egner, & Soto, 2012).

One explanation for these incidental biasing effects is that holding an item in WM recruits the sensory representation of that item, which in turn facilitates processing of memory-matching items in the sensory environment

Published online: 02 February 2017

(Desimone & Duncan, 1995). This view implies a tight link between what we think about and where we attend, which has profound implications for our understanding of cognition and behavior. In fact, if this link were truly reliable, then one should be able to infer what a person is holding in mind based on their unintentional attentional bias. Robust evidence for memory-based attentional bias across various paradigms and populations have been found as mean differences in search response times (RTs) when aggregated over many trials and subjects, leaving open the question about whether the effect is present within single subjects or generalizes well to new subjects. Mean group effects do not necessarily translate into successful classification of categorical data at the single-trial level (see Franz & von Luxburg, 2015), as aggregate effects can be driven by a subset of trials or subjects, and/or the signal-to-noise ratio of a given effect measurement might be too small for reliable single-trial inference. Previous work has also demonstrated high variability in WM-based attentional effects across individuals, even within the same task (e.g., Dowd, Kiyonaga, Egner, & Mitroff, 2015). Thus, in this study, we asked whether the effects of WM maintenance on attentional orienting are robust enough to be used diagnostically for inferring a person's mental content. Specifically, we applied multivariate pattern classification techniques to the pattern of RTs observed in an unrelated search task during the WM delay period to "decode" the specific item an individual is holding in WM on a particular trial. Moreover, this multivariate approach shows that predictions about WM-based attentional bias are generalizable to completely new subjects, indicating a shared cognitive mechanism for how WM impacts attention.

Method

Participants

One hundred Amazon Mechanical Turk workers (mean age = 35.2 years, *SD* = 10.2; 50 male) participated in exchange for monetary compensation ($3.50–$4.00). All participants provided informed consent in accordance with the Duke University Institutional Review Board. Prior to collecting data, we set a data inclusion threshold of 85% accuracy on both memory and search tasks; this performance threshold was set in order to maximize the number of instances (correct memory responses) and number of features (correct search RTs) fed into the different classifiers. Thus, data from an additional 19 participants were excluded for below-threshold performance: eight for memory accuracy, three for search accuracy, and eight for both memory and search accuracies.

Stimuli and procedure

The dual-task paradigm consisted of a delayed match-to-sample WM test with an intervening visual search task comprised of 12 successive visual searches (see Fig. 1). We included 12 separate searches within each delay period in order to form a multidimensional space of classifier inputs, described further in this section. Each trial began by presenting a central fixation cross for 500 ms, followed by the WM cue item under the text "Remember this color!" for 1,000 ms. Each WM cue was a single colored circle (113 × 113 pixels) in one of four colors (RGB values: red = 227,2,24; blue = 48,62,152; green = 95,180,46; yellow = 251,189,18). The WM cue was followed by a blank delay for 2,000 ms, then by a series of 12 visual searches. Each visual search presented a central fixation cross for 500 ms, followed by a search array for 300 ms, and then a blank screen until a response was recorded (or up to 1,200 ms). Each search array consisted of two colored circles presented along the upper and lower halves of the vertical midline, and each circle surrounded a white line (10 × 75 pixels). One line (i.e., search distractor) was vertical, while the other line (i.e., search target) was tilted 45° to the left or right. The visual search stimuli were thus distinct from, but embedded within, the class of stimuli that could match WM (see Soto et al., 2005). Participants were instructed to indicate the orientation of the line via key press, as quickly and accurately as possible. Target locations and orientations occurred equally often in a randomized order. After the 12th search, a four-alternative forced-choice memory probe array was presented under the text "Which color did you remember?" for up to 3,000 ms; participants were instructed to report which color matched the initial WM cue color. Color-response mappings of the memory probe array changed across trials. Feedback was presented following all memory responses (i.e., correct or incorrect) and only after erroneous (i.e., time-outs or wrong key presses) search responses. Participants completed four practice trials, followed by 80 experimental trials (20 of each WM color, randomly intermixed) across 10 blocks.

Importantly, the series of 12 visual searches during the WM delay contained one of each possible combination of two-color (${}_4C_2 = 6$) and target-distractor (2) arrangements (see Fig. 2), resulting in 12 unique search arrays that were presented in randomized order. This scheme ensured that for each trial, no matter the WM color, there was always a decoupling between the specific validity relationships and that memory color. In other words, the full set of 12 possible search arrays was completely agnostic to the WM color. Thus, although each WM color cue was followed by the exact same 12 search combinations (but randomized in order), the corresponding validity condition of each search changed, depending on the specific WM color (but always resulting in three valid, three invalid, and six neutral searches; see Fig. 1). This changing

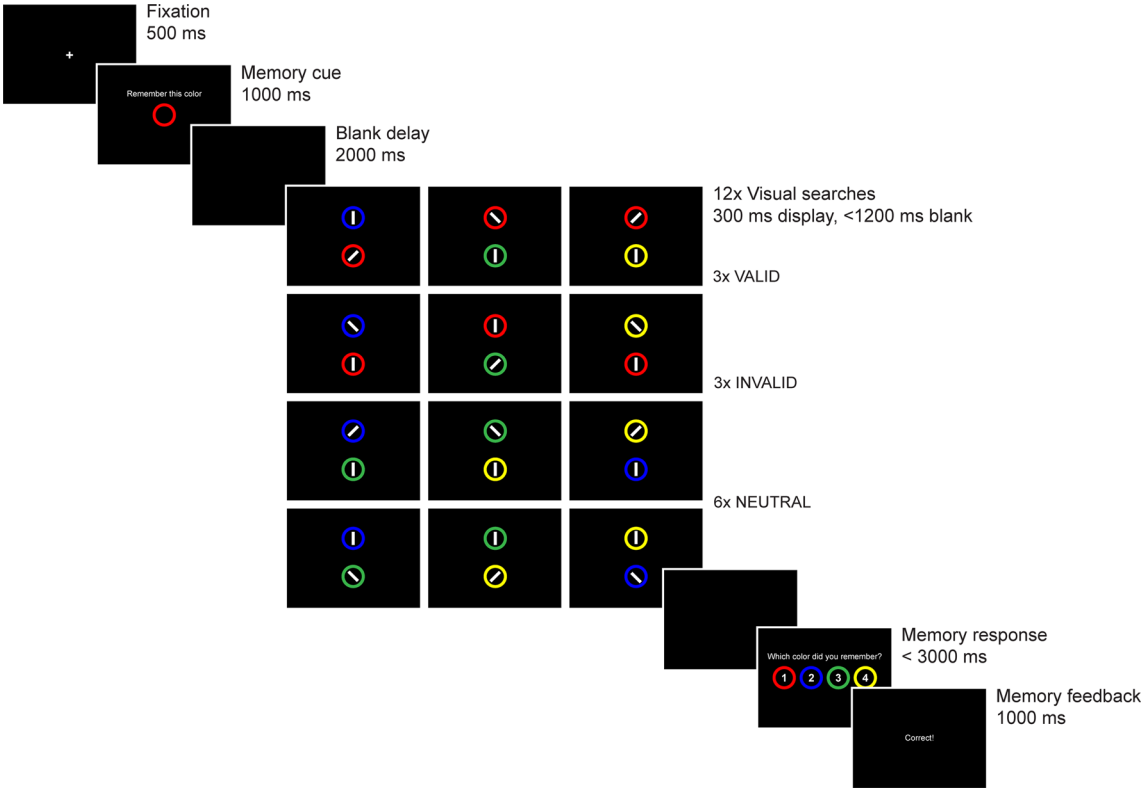

Fig. 1 Example trial sequence. Participants were to remember a cue color (red, blue, green, or yellow) while performing a series of 12 intervening visual searches, in which they reported the direction of the tilted line. Participants were then given a four-alternative forced-choice memory probe array and asked to report which color matched the initial memory cue. In this example, the cue color is *red*; in the series of 12 searches, three are valid (*top row*; target is red), three are invalid (*second row*; distractor is red), and six are neutral (*bottom two rows*; neither target nor distractor is red) (Color figure online)

pattern of validity condition—and subsequent attentional biasing effects—is thus the targeted behavioral index by which to decode the contents of WM.

Classification analysis

Our key analysis focused on using multivariate pattern classification to decode the contents of WM from behavior. The multidimensional feature space consisted of search RTs, organized by the 12 search combinations within each trial; although the order of the 12 search combinations was randomized within each trial, these RTs were rearranged such that each feature column of the input matrix represented a single search combination across all trials. Each trial thus contributed a vector of 12 search RTs, and was labeled by the corresponding WM color (see Fig. 2).

Preprocessing of the classifier inputs included (1) excluding trials with incorrect WM responses (4.1%) from the classifier, and replacing incorrect search responses (3.7% of all visual searches) with noninformative mean values, which allowed us to discount individual search errors without excluding the entire trial vector; (2) detrending RTs across blocks and removing any mean differences by memory color; (3) rescaling each feature column to [0 1], so that each feature was encoded with a similar dynamic range (i.e., relative to a given subject's RT means and variances); and (4) including the squares and square-roots of each feature column (see Supplement S4), resulting in a 36-dimensional feature space.

We analyzed three separate classifiers: distance analysis, logistic regression, and linear support vector machines. In the first distance analysis, trials were classified according to the smallest Euclidean distance between a test vector and the mean training vector for each label (i.e., WM color). That is, trials were classified by their similarity to the population mean response patterns. Distance analysis thus represents a simple, naïve model for classifying specific WM content. The second analysis employed multinomial logistic regression to calculate probabilistic predictions for each label, using the "glmnet" package in R (Hastie & Junyang, 2014). Finally, we trained linear support vector machines (SVM), which are commonly used in multivariate decoding analyses of neural data (Haynes, 2015). For multiclass SVM classification, we used the "e1071" package in R to train six binary one-against-one classifiers; for each test instance, each binary classifier output a winning label, and the predicted WM color for that test instance was the label with the most wins among all binary pairwise classifiers (Dimitriadou, Hornik, Leisch, Meyer, & Weingessel, 2006; see Supplement S3).

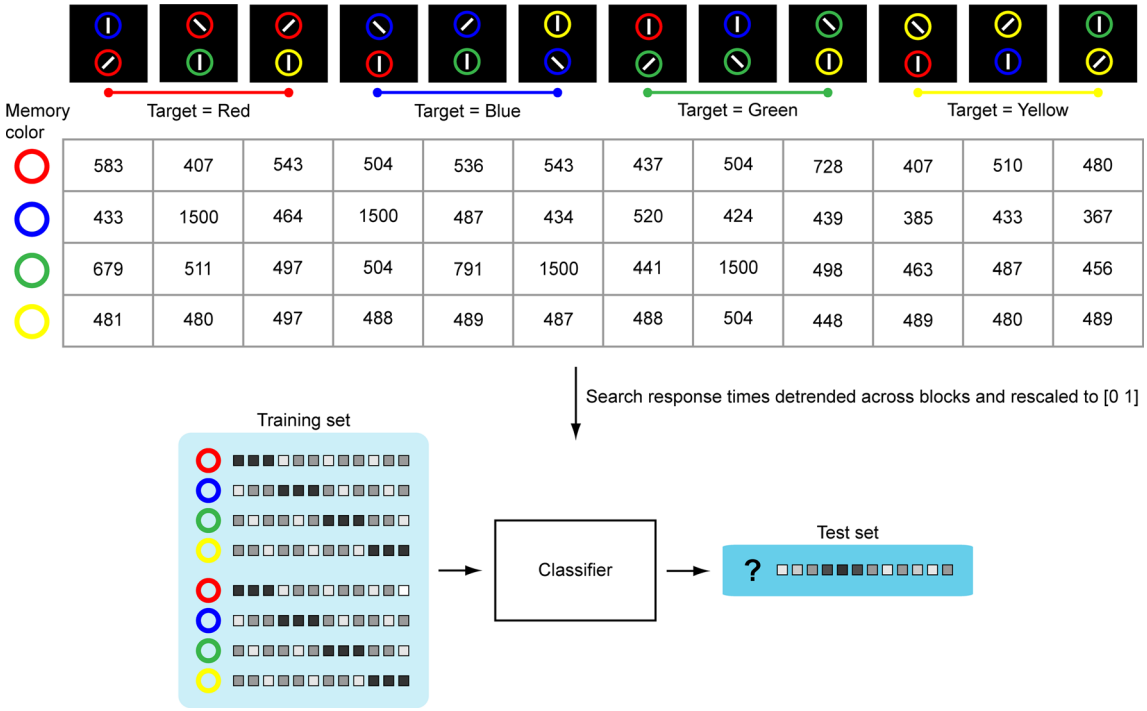

Fig. 2 The multidimensional feature space was organized according to the 12 possible two-color (6) and target-distractor (2) arrangements (*top*). Classifier inputs were search response times for each feature. Thus, each trial contributed a vector of 12 search response times, labeled by the corresponding memory color; four example, input vectors are shown in the grid. Response times were detrended across blocks and memory colors and rescaled within each feature column to [0 1]. For within-subject classification analyses, we adopted a leave-one-trial-out cross-validation scheme; within each participant's data, the classifier was trained on the vectors corresponding to all but one trial (*training set*), then tested on its ability to decipher the memory label on that one trial (*test set*) (Color figure online)

All classifiers were evaluated with a cross-validation approach, wherein a subset of trials was used to train the classifier, and the WM color was predicted for an independent set of test trials. For within-subject classification, we adopted a leave-one-trial-out cross-validation scheme, wherein the classifier was trained on vectors corresponding to all but one trial, then tested on its ability to decipher the WM label on that one trial (see Fig. 2). The training set was iteratively shuffled so that each trial served once as the test set, and classifier accuracy for a given participant reflected the average classifier performance across all iterations. In other words, a within-subject classifier might be trained on 79 trials and tested on one trial, iteratively shuffled 80 times. For between-subject classification, we adopted a similar leave-one-subject-out cross-validation scheme, in which the classifier was trained on the data corresponding to all but one participant, whose single-trial data then served as the test set. All between-subject analyses were also iteratively shuffled until each participant had served once in the testing set, resulting in a single average classification accuracy across all 100 iterations. In other words, a between-subject classifier would be trained on 99 subjects (up to 7,920 trials) and tested on one subject (up to 80 trials), iteratively shuffled 100 times.

Statistical significance of classification was assessed with permutation testing. For within-subject permutations, the memory labels for each subject's data were randomly shuffled and subjected to leave-one-trial-out classification 100 times. To produce a null distribution of group-level accuracies, the subject-level permutations were combined with a bootstrapping procedure, in which one of the 100 values was randomly chosen (with replacement) for each subject, then averaged across all participants, 1,000 times (see Stelzer, Chen, & Turner, 2013). The significance of the empirical group-level accuracy was assessed relative to this bootstrapped null distribution. For between-subject permutations, the memory labels within each subject's data were randomly shuffled, and the entire set of data was subjected to a leave-10-subjects-out (i.e., 10-fold) classification; this was performed 1,000 times, producing a null distribution of 1,000 permuted accuracies.

Results

Mean performance data

Both memory accuracy (M = 95.9%, SD = 4.0%) and search accuracy (M = 96.3%, SD = 3.2%) were near ceiling. A repeated-measures ANOVA with a Greenhouse–Geisser correction revealed that RTs for correct visual searches differed

significantly across validity, $F(2, 198) = 250.4$, $p < .001$, $\eta^2 = .72$. Search times were overall faster on valid trials ($M = 580$ ms, $SD = 134$ ms), $t(99) = 10.87$, $p < .001$, and slower on invalid trials ($M = 607$ ms, $SD = 149$ ms), $t(99) = 14.88$, $p < .001$, compared to neutral trials ($M = 589$ ms, $SD = 136$ ms), replicating canonical validity effects of attentional bias toward memory-matching contents (Fig. 3; see Supplement S1). These validity effects were significant and robust across the entire series of 12 intervening searches (see Supplement S2), validating the inclusion of entire patterns of RTs as classifier features.

Classification results

Our main analyses focused on using multivariate pattern classification to decode the contents of WM from behavioral indices of incidental attentional bias in the unrelated search task. We first classified the contents of WM within each participant's data, using a leave-one-trial-out cross-validation scheme. Decoding accuracy for the three classifiers is shown in Fig. 4a, where a violin plot illustrates the distribution of each classifier's accuracy over the 100 participants. Mean within-subject decoding accuracy was compared to chance-level null distribution as calculated by permutation tests. All three methods classified the trial-specific WM contents within a single participant's set of data with significantly above-chance accuracy (all ps < .001; see Table 1, Fig. 5a): Overall within-subject decoding accuracy for the distance analysis was 28.9%; for logistic regression it was 30.6%; and for linear SVM it was 31.3%. Logistic regression and linear SVM classifiers performed significantly better than the distance analysis, as estimated by Wilcoxon signed-ranks tests with a Bonferroni-corrected significance threshold of .017 (Demšar, 2006).

We hypothesized that classifiers were primarily driven by the univariate mean validity effects observed in previous studies. To test this, we examined the correlation between classifier performance and a coarse measure of the magnitude of attentional bias, as calculated from search RTs: (invalid – valid) / neutral. This normalized difference score reflects the absolute RT difference as a proportion of neutral RTs, such that positive values indicate greater attentional bias (Dowd, Kiyonaga, Egner, et al., 2015). Individual differences in classification accuracy for all three approaches were significantly predicted by the magnitude of attentional bias, with R^2 values ranging from .30 to .44, ps < .001 (see Fig. 4b). Thus, greater average attentional bias by WM contents was associated with better decoding, suggesting that the trained classifiers indeed capitalized on patterns of search validity effects that emerged across each trial. In sum, single-trial WM content could be inferred at above-chance levels based on incidental attentional bias within subjects.

We next asked a more profound question of the robustness of incidental attentional bias by WM, namely, whether a classifier trained on data from a group of individuals could predict single-trial WM content in another, new individual. Permutation tests revealed that all three classifiers decoded the specific WM contents *across* the group of 100 participants significantly above chance (all ps < .001; see Table 1, Fig. 5b): Between-subject decoding accuracy for the distance analysis was 36.4%; for logistic regression it was 36.5%; and for linear SVM it was 36.1%. Wilcoxon signed-rank tests revealed no significant differences across between-subject classification methods. Thus, we found that the particular color a participant held in mind on a specific trial could be inferred based on classifiers trained on completely different participants. Our between-subject classifiers had generally higher decoding accuracy compared to within-subject approaches, due to higher volumes of input data (see Supplement S6).

Discussion

Recent behavioral and neural evidence supports the idea that WM recruits the same sensory representations that are activated by attending to external stimuli (see Sreenivasan, Curtis, & D'Esposito, 2014), such that holding an item in WM automatically facilitates the processing of memory-matching items in the sensory environment (Desimone & Duncan, 1995). Here, we tested the hypothesis that if the link between WM and

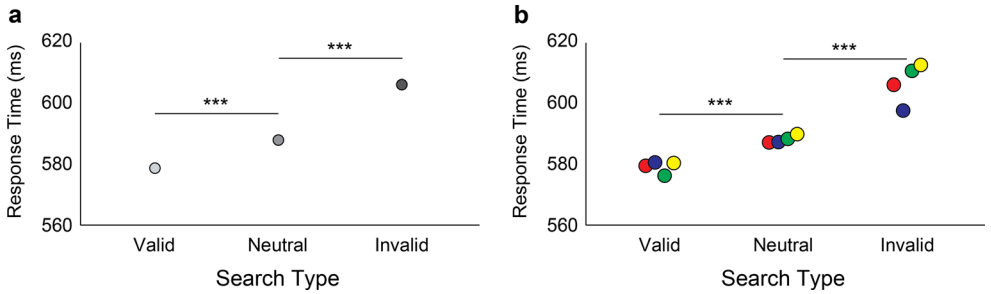

Fig. 3 Mean search response times by validity conditions (**a**) and by memory color (**b**) revealed significant validity effects, such that response times were overall faster on valid searches and slower on invalid searches, compared to neutral searches. The diameter of each *dot* represents 95% confidence intervals. *** $p < .001$. See Tables S1 and S2 (Color figure online)

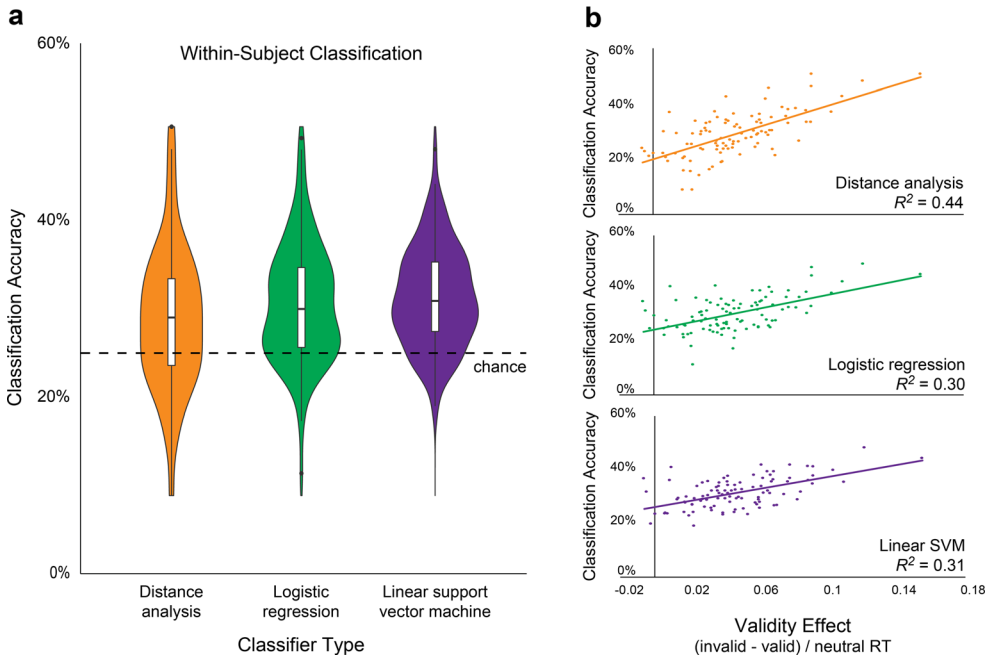

Fig. 4 Within-subject classification results. **a** Across three types of classification methods, within-subject decoding accuracy was significantly above chance (*dashed line* at 25%), as evaluated with permutation tests. For each classifier, a *violin plot* illustrates the distribution of that classifier's accuracy over the 100 participants. The *boxes* mark the middle two quartiles of classification performance, and the *whiskers* represent the full range of classification performance (outliers are marked with *dots*). **b** Individual differences in classification accuracy for all three approaches were significantly predicted by validity effect (i.e., magnitude of attentional bias), suggesting that trained classifiers capitalized on patterns of search validity effects that emerged across each trial

attentional orienting (i.e., in the form of memory-driven attentional bias) is so highly robust, then we should theoretically be able to infer what a person is holding in WM, based on their attentional bias. Importantly, in this study, attentional bias was derived from a search task *unrelated* to WM content, providing a crucial test of the link between WM and *incidental* attentional bias. We applied a series of classifiers to show that this effect is large and robust enough to provide out-of-sample prediction in *new* individuals. In doing so, we were able to eliminate several alternatives, including the possibility that significant effects at the population level were driven by small subsets of trials or only a few individuals.

We replicated canonical memory-based attentional biasing effects in which search RTs were speeded to memory-matching target locations and slowed by memory-matching distractor locations. However, mean univariate effects may be driven by a subset of trials or subjects, and/or the signal-to-noise ratio of a given effect measurement might be too small for reliable single-trial inference. Thus, these simple behavioral outputs (i.e., RTs) were fed into multivariate pattern classification analyses in order to predict the contents of WM on a single-trial level. Across three separate within-subject classification methods, patterns of incidental attentional bias within

Table 1 Overall classification accuracy and permutation distributions by classifier

Within subject	Classification	Permutation (1,000 repeats)	Permutation test
Distance analysis	.289 (.081)	.248 (.006)	$p < .001$
Logistic regression	.306 (.067)	.252 (.005)	$p < .001$
Support vector machine	.313 (.056)	.239 (.005)	$p < .001$
Between subject	**Classification**	**Permutation (1,000 repeats)**	**Permutation test**
Distance analysis	.364	.250 (.006)	$p < .001$
Logistic regression	.365	.255 (.003)	$p < .001$
Support vector machine	.361	.251 (.006)	$p < .001$

Note. Standard deviations are presented in parentheses. Significance values represent permutation tests of the mean classification accuracy compared to a null permutation distribution of 1,000 repeats

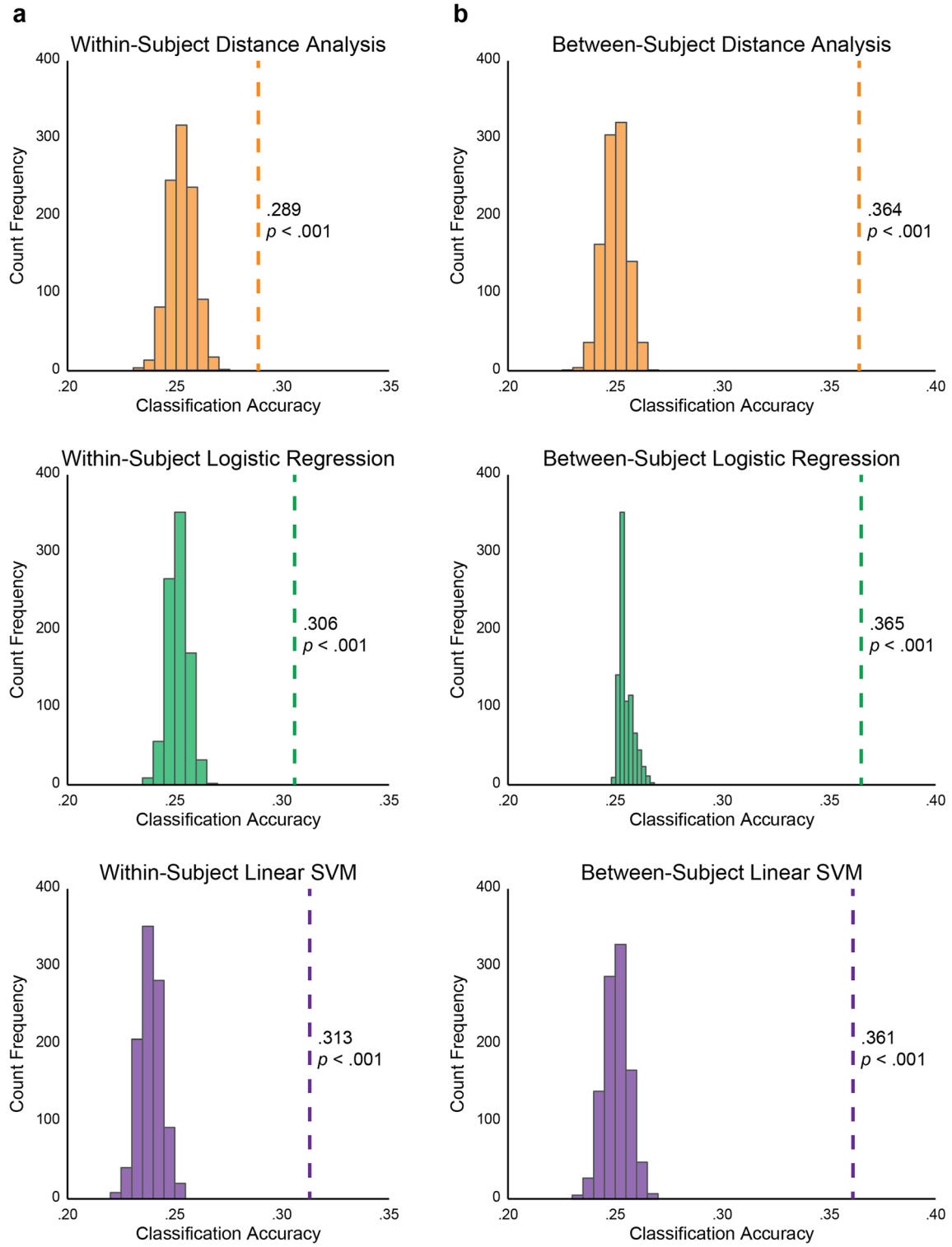

Fig. 5 Null distributions (1,000 permutations each; *colored bars*) compared to empirical classification performance (*dashed lines*) for within-subject (**a**) and between-subject (**b**) classifiers across distance analysis, logistic regression, and linear support vector machine (*SVM*). Permutation testing revealed that all classifiers performed significantly above chance, *p*s < .001 (Color figure online)

an individual's search task data could reliably decode which specific color the individual was maintaining in WM for a single, particular trial, highlighting the specificity and reliability of the link between WM and attention. Classifier evidence was furthermore significantly predicted by the magnitude of validity effects, indicating that the classifiers were indeed decoding WM contents from patterns of memory-based attentional bias.

Given the variation in standard univariate validity effects across individuals (see Dowd, Kiyonaga, Egner, et al., 2015), the generalization of decoding WM contents from attention bias was unknown. Thus, this study explicitly tested whether a classifier trained on data from a group of individuals could predict single-trial WM content in another, new individual. Again, across three separate between-subject classification methods, the specific contents of WM—for a completely independent participant—could be reliably inferred from patterns of incidental attentional bias within a separate group's search task data. While such between-subject effects are often inferred from group-level univariate statistics, here we demonstrate between-subject *generalization* by using a leave-one-subject-out cross-validation approach. This novel evidence for generalizability of memory-driven attentional bias may thus reflect a common, shared cognitive mechanism for how WM impacts attention.

The multivariate nature of the current approach facilitated the decoding of specific contents of WM from incidental attentional bias. A standard univariate analysis would not have achieved such results (see Supplement S5). While the primary multivariate signal driving classification performance was likely the changing pattern of validity effects, inputting response times on the scale of individual searches further maximized the potential for classifier to pick up on patterns of response time variability. The additional inclusion of quadratic and square-root features (see Supplement S4) and residual differences in the magnitude of validity effects across memory colors (see Supplement S1) may have also contributed to classification performance. The more complex machine learning methods (i.e., logistic regression and linear SVM) also performed better than the simplest distance analysis for within-subject classification, supporting these classifiers' ability to fit and leverage small, trial-level differences within subjects. While a simple, theoretically motivated rule achieved an appreciable fraction of the best classifier performance in our current task, it is important to note that in many other tasks where no such rule is known, or where nonlinearities dominate, our approach remains valid. Thus, this method links exploratory data mining and confirmatory modeling approaches and may generalize across cognitive paradigms.

In cognitive neuroscience, multivariate pattern-based analyses of brain activity have been used to reveal a person's mental states—such as what the individual is perceiving, attending to, or remembering (e.g., Harrison & Tong, 2009; see Tong & Pratte, 2012, for a review). Specifically, the contents of WM have been successfully decoded from within-subject neural activity, with variable ranges of accuracy—for example, Serences, Ester, Vogel, and Awh (2009) classified red versus green stimuli in early visual cortex (V1; 62 voxels) with nearly 65% accuracy (compared to 50% chance), while Albers, Kok, Toni, Dijkerman, and de Lange (2013) classified three grating orientations in V1–V3 (360 voxels) with 54% accuracy (compared to 33% chance), both across different time periods within the WM delay. While it is difficult to directly compare classification accuracies across different feature inputs and task designs, our within-subject results (31% compared to 25% chance) are derived from inexpensive, purely behavioral inputs. Furthermore, we demonstrate that decoding of WM content generalizes reliably to *completely novel* individuals, a notoriously difficult problem for neuroimaging-based classifiers due to the methodological hurdle of normalizing morphological idiosyncrasies across participants (see Haxby, Connolly, & Guntupalli, 2014).

Moreover, the multivariate behavioral inputs for WM decoding in this study are derived from an *unrelated* attentional task. While multivariate pattern analyses have been previously applied to behavioral eye-movement statistics to classify the type of viewing task that a participant is engaging in (Borji & Itti, 2014; Henderson, Shinkareva, Wang, Luke, & Olejarczyk, 2013; Kardan, Berman, Yourganov, Schmidt, & Henderson, 2015; but see Greene, Liu, & Wolfe 2012), these studies took advantage of systematic differences in a task-relevant behavior (i.e., goal-directed eye movements) to identify different task strategies—in contrast, this study focuses on classifying specific mental (WM) representations, as opposed to entire task sets, and is based on behavioral data from an *unrelated* attentional task: the appearance of memory-related visual objects (i.e., colored circles) in visual searches was task-irrelevant; they had no bearing on the search task itself (i.e., searching for a target line), and the same 12 combinations of two-color and target-distractor arrangements were always presented within every trial. Thus, there was a decoupling between the particular memory color and the subsequent validity relationships, such that there was no strategic incentive to attend to the changing pattern of validity conditions within each trial (also see Soto et al., 2008).

In conclusion, this study demonstrates for the first time that memory-based incidental attentional biasing effects are so robust that the specific single-trial contents of WM can be reliably inferred from patterns of attentional bias in an *unrelated* search task. Furthermore, predictions about memory-based attentional bias were successfully generalized to completely new subjects, suggesting a highly generalizable relationship between WM and attention, likely due to a shared cognitive architecture—specifically one in which the maintenance of any information in WM activates memory-matching sensory features, providing an automatic and robust advantage for further attentional processing in the visual field (Desimone & Duncan, 1995). Together, the current results emphasize the specificity and the generalizability of a tight link between what we keep in mind and what we attend to, contributing to a fuller theoretical understanding of the interactions between WM and visual attention.

Acknowledgments This work was supported in part by National Institute of Mental Health Grant R01 MH087610 (to T. Egner) and National Institutes of Health Big Data to Knowledge Career Development Award K01-ES-025442-01 to (to J. Pearson).

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology, 23*(15), 1427–1431.

Borji, A., & Itti, L. (2014). Defending Yarbus: Eye movements reveal observers' task. *Journal of Vision, 14*(3), 29–29.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7,* 1–30.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., & Weingessel, A. (2006). *e1071: Misc functions of the department of statistics (e1071)*.

Dowd, E. W., Kiyonaga, A., Beck, J. M., & Egner, T. (2015). Quality and accessibility of visual working memory during cognitive control of attentional guidance: A Bayesian model comparison approach. *Visual Cognition, 23*(3), 337–356.

Dowd, E. W., Kiyonaga, A., Egner, T., & Mitroff, S. R. (2015). Attentional guidance by working memory differs by paradigm: An individual-differences approach. *Attention, Perception & Psychophysics, 77*(3), 704–712.

Franz, V. H., & von Luxburg, U. (2015). No evidence for unconscious lie detection: A significant difference does not imply accurate classification. *Psychological Science, 26*(10), 1646–1648.

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision Research, 62*(C), 1–8.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458*(7238), 632–635.

Hastie, T., & Junyang, Q. (2014). *Glmnet vignette*. Retrieved from https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience, 37*(1), 435–456.

Haynes, J.-D. (2015). A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron, 87*(2), 257–270.

Henderson, J. M., Shinkareva, S. V., Wang, J., Luke, S. G., & Olejarczyk, J. (2013). Predicting cognitive state from eye movements. *PLoS ONE, 8*(5), e64937. doi:10.1371/journal.pone.0064937

Kardan, O., Berman, M. G., Yourganov, G., Schmidt, J., & Henderson, J. M. (2015). Classifying mental states from eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 41*(6), 1502–1514.

Kiyonaga, A., Egner, T., & Soto, D. (2012). Cognitive control over working memory biases of selection. *Psychonomic Bulletin & Review, 19*(4), 639–646.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science, 20*(2), 207–214.

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance, 31*(2), 248–261.

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences, 12*(9), 342–348.

Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences, 18*(2), 82–89.

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage, 65*(C), 69–82.

Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology, 63*(1), 483–509.