UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Influences of both prior knowledge and recent historyon visual working memory

Permalink

https://escholarship.org/uc/item/3wk1b0qn

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

DeStefano, Isabella
Vul, Edward
Brady, Timothy F.

Publication Date

2020

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Influences of both prior knowledge and recent history on visual working memory

Isabella DeStefano (idestefa@ucsd.edu),
Edward Vul (evul@ucsd.edu), Timothy F. Brady (timbrady@ucsd.edu)
University of California, San Diego, Department of Psychology
9500 Gilman Dr., La Jolla, CA 92093 USA

Abstract

Existing knowledge shapes and distorts our memories, serving as a prior for newly encoded information. Here, we investigate the role of stable long-term priors (e.g. categorical knowledge) in conjunction with priors arising from recently encountered information (e.g. 'serial dependence') in visual working memory for color. We use an iterated reproduction paradigm to allow a model-free assessment of the role of such priors. In Experiment 1, we find that participants' reports reliably converge to certain areas of color space, but that this convergence is largely distinct for different individuals, suggesting responses are biased by more than just shared category knowledge. In Experiment 2, we explicitly manipulate trial n-1 and find recent history plays a major role in participants' reports. Thus, we find that both global prior knowledge and recent trial information have biasing influences on visual working memory, demonstrating an important role for both short- and long-term priors in actively maintained information.

Keywords: working memory, serial dependence, prior knowledge, iterated learning, reconstructive memory

Introduction

When we perceive the world, and particularly when we store information in memory, we do not do so independently of our knowledge and expectations. Instead, our existing knowledge influences our memory (Bartlett, 1932) — resulting in systematic biases in which information we remember best, and distorting our memories toward our "priors" or schemas (e.g. Brewer & Treyens, 1981). This occurs because our prior knowledge provides an independent source of information that can help reduce uncertainty about the world, particularly for memory which is imperfect and noisy. Such memory distortions can be formalized as Bayesian inference where new noisy information is incorporated with prior expectations that originate from our knowledge about the world (Huttenlocher, Hedges, & Vevea, 2000; Hemmer & Steyvers, 2009) or about the current context (e.g. Brady & Alvarez, 2011). To understand the nature of memory representations; how they are limited; and how they support our decisions and behavior, we need to unravel the role of priors in these processes.

Isolating people's "priors" is not a straightforward task, as they are likely quite complex and high-dimensional. One strategy is to borrow methods from the rich tradition in machine learning and statistical inference; for example, past work has shown that by modeling tasks after the iterative sampling process known as Monte Carlo Markov Chain (MCMC) we can utilize systematic biases in behavior to measure people's priors (Sanborn & Griffiths, 2008). Similarly, iterated learning or iterated reproduction are expected to converge to people's prior: by having subjects reproduce or evaluate the retrieved information from a different subject, it is possible to measure Gestalt priors in working memory across people (Lew & Vul, 2015).

To date, these attempts have largely assumed that the biases caused by such priors are relatively stable — both across and within individuals. Notably, it is often assumed that memory judgements are relatively unchanged by recent history, and so the use of such iterated designs taps into stable long-term priors. However, the extent to which this is true remains an open question, as recent history often has a strong influence on judgments about current stimuli (Huang & Sekuler, 2010; Fischer & Whitney, 2014). In addition, past iterated learning studies investigating memory biases have used different individuals in each iteration (Lew & Vul, 2015), thus investigating only long-term priors that are invariant across people. This technique may not be suitable for all cases since the extent to which individuals differ in their prior knowledge will depend both on the temporal stability of such priors and on the domain of study.

In the present study we adapt an iterated reproduction paradigm to investigate how idiosyncratic individual priors play a role in working memory biases, and do so while considering not only stable long-term priors of individuals but also possible trial-by-trial influences of recent history.

We focus specifically on the the case of visual working memory for color. This domain provides a rich arena for study, as it has been well characterized; is related to important broader concepts (like fluid intelligence; Cowan, Fristoe, Elliott, Brunner, & Saults, 2006) and allows for rigorous psychophysical techniques designed to precisely understand memory representations (Wilken & Ma, 2004). While the precise and seemingly perceptual nature of this memory system has led some to conclude it is relatively unaffected by priors and biases (e.g., Lin & Luck, 2012), there is some evidence that this is not the case. For example, there are systematic biases that show up on average across individuals, and seem to align with the use of linguistic color categories (Bae, Olkkonen, Allred, & Flombaum, 2015; Allred & Flombaum, 2014). This group-level color category memory bias is consistent with Bayesian integration, and can be mechanistically modeled with a drift-diffusion process wherein color memories drift towards psychophysical attractor states that correspond to color categories (Panichello, DePasquale, Pillow, & Buschman, 2019). In addition to such long-term priors, there

©2020 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

are, at least in some circumstances, effects of immediate trial history in generating interference (Makovski & Jiang, 2008) and biasing memory (Huang & Sekuler, 2010). Thus, color memory is an important domain and one in which both short- and long-term priors seem to influence memory retrieval.

There remain many important unanswered questions about the usage of priors in this domain. For example, the role that individual differences in color categories play in memory has remained relatively unexplored. Above and beyond biological factors that affect color perception (e.g., color blindness), individual differences in color categories have been demonstrated in perceptual matching tasks (Webster & Kay, 2012), and cultural and linguistic differences in color categories are well established (e.g., Regier, Kay, & Khetarpal, 2007). In addition, the relative importance of short-term (Makovski & Jiang, 2008) vs. more stable (Bae et al., 2015) priors remains unknown. Thus, the known presence of both short- and long-term biases and individual differences means the case of visual working memory for color is well suited for studying how biases from recent history interact with individual differences and long-term category priors. Therefore, in the current study we used an iterated reproduction paradigm designed to shed light on these issues, including two core research questions: 1) How much do individual differences affect visual working memory? and 2) How do global category biases interact with short-term biases from recent trials?

Experiment 1

Experiment 1 is designed to capture individual differences in how color categories affect visual working memory. We use a novel iterated reproduction paradigm to reveal participants' priors.

Methods

Participants N=70 undergraduate students were recruited to participate in this study (48 female, mean age 20.8) in exchange for course credit. All subjects gave informed consent, and the study was approved by the UC San Diego Institutional Review Board. All subjects had normal or corrected-to-normal visual acuity and normal color vision as assessed with Ishihara's test of color deficiency(Ishihara, 1987).

Procedure Participants were repeatedly asked to reproduce a single color from memory using a continuous color wheel. Color stimuli were drawn from a circle in CIE L*a*b* color space, centered at (L=54, a=21.5, b=11.5) with a radius of 49. Thus, each color can be considered an angle – and errors in reproduction can be quantified by the angular distance between the studied color and reproduced color.

On each trial, a single color stimulus was presented for 1000ms in one of four possible locations (with equal probability). After a 1000ms delay, subjects were then cued to report the color with the continuous report color wheel. The color wheel was randomly rotated and flipped trial to trial so that location preferences would not systematically impact color reports.

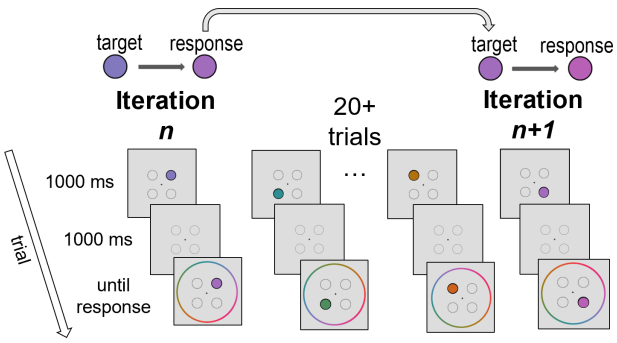

Figure 1: Iterated design: On each trial, participants reproduced 1 color after a delay of 1 second. The color reported on one trial was later shown as the memoranda on a (much) later trial.

The crucial feature of the design was that the colors shown on some trials were identical to subjects' reproductions from previous trials, making an iterated design (Figure 1). These critical trials were interleaved with an equal number of filler trials so that this manipulation was not apparent to subjects.

Due to this iterated design, throughout the experiment there were 'chains' of trials: the response from iteration n in a given chain was used as the stimulus in the $n+1$ iteration of that chain. This process continued until 15 iterations were obtained for each chain. Each chain began with a predetermined 'seed color' displayed on the first iteration. Ten seed colors evenly spaced 36 degrees apart on the color wheel were used for all subjects. These initial positions were chosen without regard to the location of color category centers or boundaries in this color space, which are not evenly spaced or uniform on the color wheel (see Figure 4). Each subject completed two chains for each of 10 unique seed colors, resulting in 20 total chains per subject. Subjects completed blocks of 20 trials (10 chain-iteration trials and 10 filler trials) that alternated between the 2 sets of iterated chains such that no two chains within a block began with the same seed color. This alternation between unique chains ensured that presentation of subsequent iterations were sufficiently delayed so that the previous iteration was wiped from working memory by intermediary trials.

In our experiment each chain was completed by a single subject in order to capture individual differences. To avoid lapses from derailing the entire chain, participants reported value was silently 'rejected' if it was an unlikely value for the chain. In particular, we used a fixed rejection rule which rejected responses with an absolute error greater than 22.5 degrees. This value was chosen as to include 80% of the error distribution measured from previous studies using the same stimuli and set size (Schurgin, Wixted, & Brady, 2018). Ultimately 15% of iteration trials had error greater than 22.5 and therefore were rejected. If a response from a particular iteration was rejected, the chain would not advance, and the next trial from that chain would have the same color stimulus value as the previous one. Because of this rejection criterion there was not a fixed number of trials per subject. Instead,

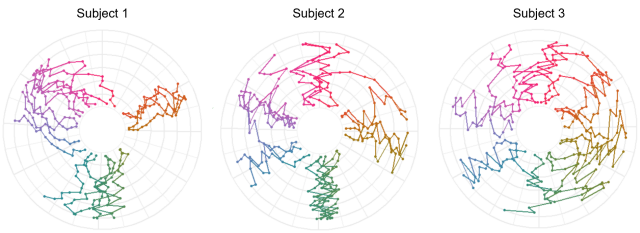

Figure 2: Examples of individual subject iterated chains. Chains begin in the center of the circle with seed colors evenly spaced. Iteration 1 is shown in the center of the circle, and iteration increases as chains move outwards. The color of each line changes with the color reproduced on that iteration which is also indicated by the line's position around the circle. As can be seen, multiple chains from each participant tend to converge to similar areas of color space, but these areas are not consistent across subjects.

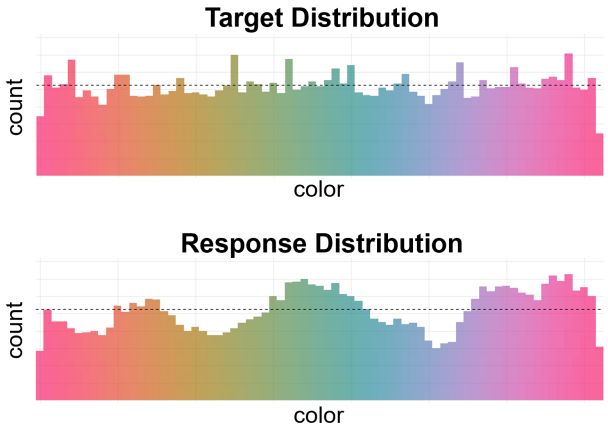

Figure 3: Participants report some colors more than others, even though those colors are not over-represented in the shown colors. This suggests reliable across-participant priors, perhaps based on color categories.

blocks were added to the end of the experiment until there were 15 iterations in all 20 chains for that subject. All subjects completed at least 30 blocks consisting of 20 trials each, alternating between the different sets of seeds. Additional blocks were added to compensate for the rejection of trials as needed. Subjects were given 1 minute breaks after every quarter of the experiment and a 2 minute break at the half-way point.

The filler trials were selected semi-randomly in order to smooth the distribution of colors presented (e.g., to ensure that subjects always see examples of each color on the wheel even if their own chains drift to a subset of the color wheel). These trials ensure that a consistent and uniform set of colors are shown to all subjects regardless of what occurs in the subjects' chains.

Results

Example response sequences are shown in Figure 2. We begin by collapsing across both chains and individuals (Figure 3). Doing so reveals that there are global biases across participants. That is, in the overall response distribution, collapsing across all subjects and all trials, the reported colors tend to cluster in particular regions of color space even when these regions are not over-represented in the distribution of colors shown (Figure 3). To quantify this tendency, we use Shannon entropy, which is maximized when a distribution is uniform. Thus, smaller values of Shannon entropy indicate more clustering of responses in certain areas of the color wheel. For each subject we calculated the Shannon entropy for the frequency distribution of colors shown and the frequency distribution of responses binned into ten degree intervals in color space, and found that the response distribution had significantly lower entropy than the target distribution, $t(69) = 17.3, p < 0.001$. This suggests that participants tend to gives responses focused around a limited number of colors rather than purely reproducing the target distribution.

The non-uniformity observed in the aggregate data is exaggerated in the distribution of responses from the final iteration of the iterated chains (Figure 4). The iterated design allows us to treat this distribution from the final iteration as the convergence of the chains to people's priors. To assess whether the convergence regions of the iterated chains can be explained by global linguistic color category priors (e.g., Bae et al., 2015; Persaud & Hemmer, 2014), we collected color naming data in two separate tasks. Both tasks were administered online, so some variance may be accounted for by differences in screen color calibration. Category labels used in both tasks were *red, orange, yellow, brown, green, blue, purple*, and textitpink which were selected from the set of basic color terms (Berlin & Kay, 1969), excluding *black* and *white*. In the first task, N=124 participants were asked to choose a prototypical color from the color wheel used in Experiment 1 for a given color term. In the second task, N=94 participants were shown a color and asked to choose a color term (from the set of terms stated above) to describe that color, and did this for each of 360 colors on our color wheel. The raw data from task 1 was then scaled by the frequency with which each color term was chosen in task 2 to produce the histogram shown in Figure 4.

The distribution of responses on the final iteration is not completely explained by color categories. While there are some consistencies, there are also large discrepancies (e.g., in green/blue, purple/pink). In the second color naming task, these regions were frequently classified as belonging to both of the flanking categories (green *blue*, and purple *pink* respectively). This was not the case for other categories (e.g., red, *orange* and yellow), where classifications had very little overlap. Thus, the failure to converge to linguistic color categories could be in part caused by ambiguity in classification of certain regions of the color wheel.

While the final iteration suggested that there are regions of color space that are, at least to some extent, converged to across participants, the extent to which individuals converge to these reliably is not clear from this analysis. To determine if individual's iterated chains are converging, we looked at how likely subjects are to report very similar colors on different trials of the same iteration. To quantify this we used the average nearest neighbor index (ANN). The nearest neighbor

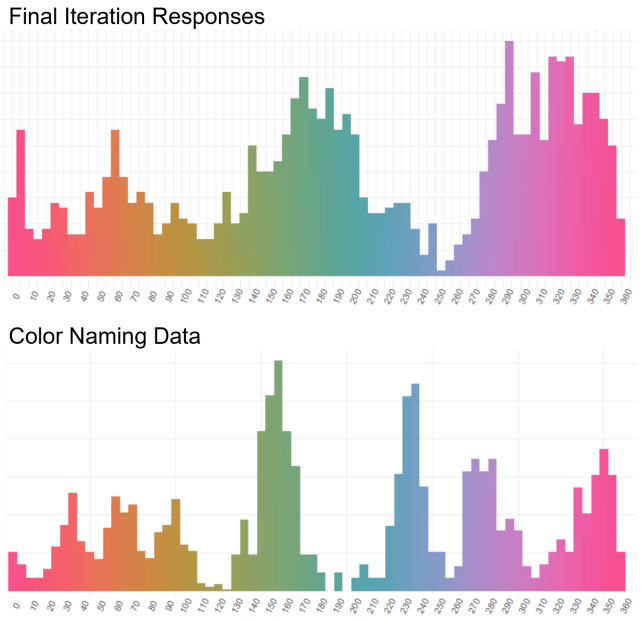

Figure 4: top: The distribution of responses for the final iteration of the iterated-chains in Experiment 1. bottom: Color naming data collected from two color term matching tasks.

distance for a single response within an iteration is the distance (in degrees around the color wheel) to the closest other response in that iteration; we then take the mean of the nearest neighbor distance for each response within an iteration to get a single number (the ANN) to approximate 'clustering' of responses in certain areas of color space across trials. Since the clustering will be relative to the number of responses within an iteration, we simulated the ANN distance that would occur by chance alone with 10,000 samples from a uniform distribution and used this as a baseline. The ratio of the ANN distance within iteration to that which would occur by chance will be greater than 1 if the responses are more separated than would occur by chance (i.e., if they were regularly spaced) and will be less than 1 if the responses are more clustered than would occur by chance. Note that the 'seed' colors shown to participants at the beginning of the experiment are maximally anti-clustered, since they are spaced equally along the color wheel (the ANN ratio for the starting seeds was equal to 2).

So that the ANN is not confounded by having two responses originating from the same seed, we find the ANN ratio across iteration for each set of chains separately within each subject. We find that the ANN ratio decreases as a function of iteration, with iteration being a significant predictor of ANN distance, $F = 1485.8, p < 0.001$ (Fig. 4). This suggests that within individuals, responses tend to cluster in particular regions of the color wheel, and this tendency increases over iteration. In the last iteration we find that the ANN ratio is significantly less than 1, $t = 6.0, p < 0.001$, meaning that the responses are more clustered than what would be expected by chance (despite having started out anti-clustered).

The clustering is largely driven by within-subject consistency, as opposed to the global consistency across subjects (as in Fig. 3). This is revealed by simulating what global clustering would exist if subject were not a factor. We simulated 10,000 samples shuffling across subjects and find that the ANN ratio of the last iteration of the shuffled null hypothesis is significantly higher than that of the original data, $t = 4.7, p < 0.001$. Thus, this procedure removes nearly all clustering, making the results look like they converge to near randomness (Fig. 4). Thus, people converge reliably to smaller parts of the color space, but these parts are not entirely the same across all individuals. The global effects (Fig. 3, Fig. 4) thus account for only a small part of the reliable pattern of convergence within individuals.

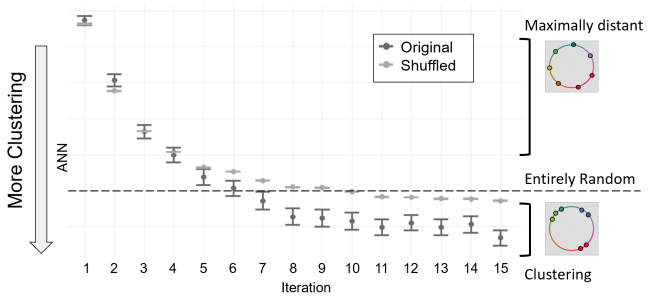

Figure 5: The distribution of seed colors starts out maximally non-clustered (equally-spaced), and the amount of clustering increases over time within-subject. The amount of clustering also increases over time between subjects ("shuffled"), but to a lesser degree than within-, and ultimately a large amount of the clustering that is reliable within-subject ("original") is not present between subject ("shuffled").

To provide a preliminary assessment of whether these individual differences were accounted for by language and color category differences, we divided the subjects into self-identified native English speakers (N=37) and non-native English speakers (N=33). We computed the Kullback–Leibler (KL) divergence between the distribution of responses from the iterated chains of native and non-native English speakers. KL divergence can be interpreted as the number of extra bits needed to encode one distribution given you already know the second distribution; it will be zero if the distributions are exactly the same. Since KL divergence is not symmetric we computed both the KL divergence between native and non-native English speakers and visa versa.

We found that $KL(native||non-native) = 0.0576$ and that $KL(non-native||native) = 0.0588$. Given that the entropy of these distributions are $H(native) = 8.35$ bits and $H(non-native) = 8.34$ bits, these are extremely low values of KL divergence, showing the two distributions barely differ, on average. Thus, the significant heterogeneity in which areas of the color wheel participants responses converge to is not well accounted for by differences in native languages.

Discussion

What causes this heterogeneity across participants? Why do participants converge to particular areas of the color wheel, but not necessarily the same ones as each other? There are

several possibilities, including individual differences in color categories, and language differences too subtle to be picked up by our analysis of native vs. non-native English speakers.

If it were the case that inconsistencies in the convergence locations of individuals could be explained by variation in individual's color categories, we would expect the final iteration responses to roughly resemble the color naming data, as both are pooled across a large number of individuals. Thus variation in color categories should still result in convergence centered around the modal colors of basic color terms. The discrepancy between the color naming data and the distribution of final chain iteration colors suggests that the convergence patterns arise from something more complicated than just individual, cultural, or linguistic variation in color categories.

An alternative explanation for this pattern of convergence is that the global prior guiding convergence of the chains is not derived from linguistic color categories, but from perceptual inhomogeneities (which may correspond to some extent to linguistic color categories); that is, from the color wheel not being perfectly perceptually uniform. Preliminary analyses of perceptual similarity data, beyond the scope of the current paper, suggest that this may to some extent explain the shape of the final iteration response distribution.

Since there are other contextual factors that also bias memory, this pattern of convergence could also arise from an interaction between the global priors and contextual information. It may be the case that the structure of the colors seen within an experiment has a large influence on subsequent memory, significantly obscuring the influence of stable category structure. For example, there may be local serial dependence such that the color from the previous trial has a major attraction effect on the current trial (e.g. Fischer & Whitney, 2014), which would lead responses in an iterated design like ours to converge over time but in idiosyncratic ways. In this paradigm, we used filler trials to smooth the distribution of colors seen by participants. The clustering of iterated chains in certain regions in color space resulted in the filler trials in later blocks being systematically different from the iteration trials. If there are indeed strong local effects of context, this systematic difference between filler and iteration trials in later blocks could have altered the convergence path of iterated chains. In Experiment 2 we explore how recent history affects convergence to see if this could have contributed to the individual differences in convergence patterns observed in Experiment 1.

Experiment 2

In Experiment 2 we performed a similar iterated color reproduction experiment but explicitly manipulated the colors participants saw on trial $n-1$ to examine the effect of recent trial-by-trial history.

Methods

Participants Seventy (N=70) undergraduate students were recruited to participate in this study (57 female, mean age 20.2) in exchange for course credit. All subjects gave informed consent, and the study was approved by the relevant Institutional Review Board. All subjects had normal or corrected-to-normal vision acuity and normal color vision as assessed with Ishihara's test of color deficiency(Ishihara, 1987).

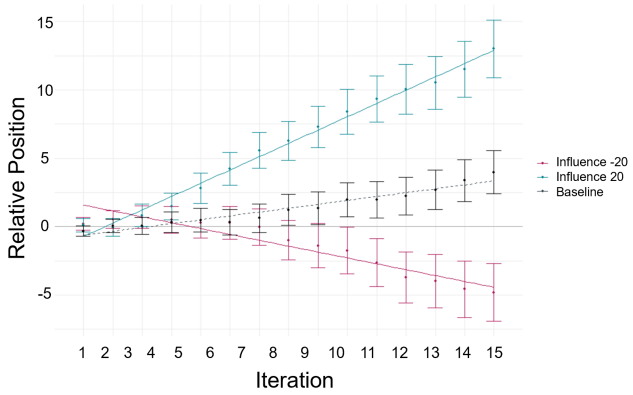

Figure 6: The relative position of each iterated chain, relative to the starting color of that chain, is strongly modulated by the direction of the prior influence trial — chains preceded by clockwise influence trials tended to drift clockwise (positive direction), and vice versa. The baseline drift is caused by having only 5 seed locations.

Procedure The memory reproduction paradigm used in Experiment 2 was almost identical to that of Experiment 1. The same iterated reproduction procedure was used to collect a set of chains. However, only a subset of the original seed colors was used: Five seed colors evenly spaced 72 degrees apart on the color wheel were shown on the first chain-iteration trial for every participant. Each participant again completed 2 chains for each seed color, resulting in 10 iterated chains per participant. Participants completed blocks of 15 trials (5 iteration trials and 10 non-iteration trials), which again alternated between the 2 sets of iterated chains such that no two chains within a block began with the same seed color.

Non-chain trials came in two types: filler and influence trials. The filler trials were generated in a similar manner to those in Experiment 1 to smooth the distribution of color stimuli. Influence trials were trials that immediately proceeded the iteration trials, and whose color was manipulated to always be exactly 20 degrees clockwise or counterclockwise of the chain-iteration trial that would follow. The iteration trial always appeared in the same spatial location as the influence trial, with the position of influence trials and filler trials remaining random. We manipulated the influence trials such that of the two chains for each seed color, one was always proceeded by a clockwise influence trial and the other by a counterclockwise influence trial.

Results

For each participant we again calculated the Shannon entropy for the frequency distribution of colors shown and the frequency distribution of responses (binned into 10 degree intervals in color space). As in Experiment 1, we found that the re-

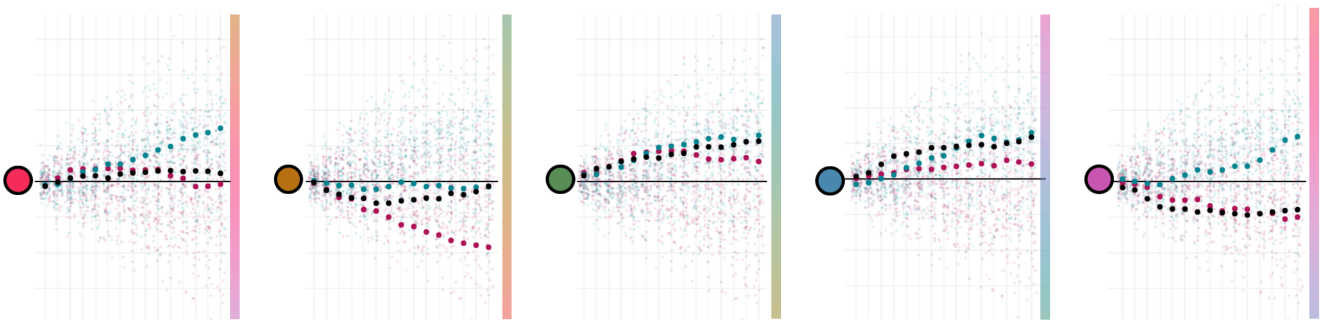

Figure 7: The axes are the same as Figure 6 (iteration x relative position, as a function of baseline vs. influence direction), but here the results are broken down by the particular seeds. The effect of the influence trials depends on the starting seed color.

sponse distributions had significantly lower entropy than the target distributions, $t(63) = 4.2, p < 0.001$.

To determine if the influence trials indeed had an effect on the trajectory of the iterated chains, we look at the relative position of the chain-iteration responses to their starting seed across iteration (Fig. 6). We used a repeated measures ANOVA and found that iteration, influence direction, and their interaction are all significant predictors of relative position (Table 1).

Next, we examined the effect of influence trials relative to no influence trials. There were no iterated chains without the influence manipulation in Experiment 2. Therefore, to make the comparison, we used Experiment 1 as a baseline (where iteration trials were preceded by random trials). We used only the iterated chains from Experiment 1 with seeds identical to those used in Experiment 2. The +20 and -20 influence conditions deviated by the same magnitude from the Experiment 1 baseline (albeit different in different directions).

The degree to which global category biases and contextual dependence biases interact can be seen when we look at the relative position of chain-iteration trials within a starting seed. The ability of the prior trial to influence the trajectory of the chain appears dependent on where the chain began. For some seeds, the influence trials have little to no effect, but for others the direction of the influence trial is large. To quantify this interaction, we used a repeated-measures ANOVA with a full model including fixed effects of iteration, influence direction, and seed. We found that all fixed effects and their 2-way and 3-way interactions were significant (Table 1).

Predictor	F-statistic	p-value
Iteration	21.5039	$p < 0.0001$
Influence Direction	221.0766	$p < 0.0001$
Seed	141.8787	$p < 0.0001$
Iteration*Influence Direction	143.4129	$p < 0.0001$
Influence Direction*Seed	16.0732	$p < 0.0001$
Iteration*Seed	28.2703	$p < 0.0001$
Iteration*Influence Direction*Seed	3.4631	$p < 0.01$

Table 1: Repeated measures ANOVA statistics and significance for all fixed effects.

General Discussion

We examined the role of recent history and long-term priors (color categories) in visual working memory. We used a strategy of iterated reproduction to elicit these effects. We find strong effects of both: while there are across-subject, global effects of color categories, there are also significant individual differences in the convergence behavior we elicit within subject — both because they may differ in their stable priors (e.g., color categories) and because of the overriding effect of recent history on where their memories converge to.

Previous work using iterated reproduction has focused solely on eliciting stable long-term priors (e.g. Sanborn & Griffiths, 2008). Similarly, mechanistic accounts of attractor dynamics assume that attractors are a fixed property of a stimulus space (Panichello et al., 2019). However, memory and cognition are filled with examples where people are unduly influenced by recent history — for example, in decision making (Yu & Cohen, 2009), perception (Fischer & Whitney, 2014) or memory (Huang & Sekuler, 2010). Here we show that such recent trial history has a strong effect on memory, in some cases overpowering the effect of long-term category priors. Existing work has also found that contextual effects from additional items (e.g., if asked to store 3 colors in mind at once) can similarly overpower long-term category priors (Brady & Alvarez, 2015). In understanding working memory it will be important to consider a hierarchical range of priors operating at different time scales — both the most stable, long-term priors (our general world knowledge), as well as priors based on recent history and even those based on the general structure of the current visual input (e.g., ensembles; Chong & Treisman, 2003). To the extent the world is generally stable, long-term priors should dominate; to the extent that the world changes over time, it is optimal to give more weight to recent history; and to the extent nearby items tend to arise from the same generative process, priors based on local context should be important.

Our findings agree with previous work that to some extent, these visual working memory biases may arise from inhomogeneities in the underlying stimulus space; specifically, CIELab color space, and other similarly constructed "uniform" color spaces, do not capture the true underlying psy-

chophysical landscape in which we represent color (Bae et al., 2015; Panichello et al., 2019). However, this is not solely because participants rely on a single set of established priors that are derived from perceptual categories (e.g. Hemmer & Steyvers, 2009). Instead, our results show that recent history can impact our memory for a recently experienced color and interact with stable long-term effects, suggesting non-stationary attractor states can be induced by contextual information. Importantly, our research suggests that while recent exposures have a strong influence on memory, they do not entirely dominate global biases (see Fig. 7).

Overall, our work has demonstrated that considering the multicausality of biases is crucial in studying memory. Memory is necessarily impacted by priors. Stable, long-term priors provide the basis for the psychophysical similarity that memory builds upon (Schurgin et al., 2018), and local priors such as recent history and spatial context provide useful constraints that help — in a spatially and temporally "smooth" world — to construct reliable representations with the very limited working memory capacity that we are afforded. Thus, while our work is a case study in visual working memory, we believe that the implications of this study extend to a much broader body of work. In memory, perception, and decision making, it is necessary to take into account prior knowledge, including both long-term, global priors and local priors arising from recent exposure, which combine and interact to inform these inherently noisy cognitive processes.

References

Allred, S. R., & Flombaum, J. I. (2014). Relating color working memory and color perception. *TiCS*, *18*(11), 1562-565.

Bae, G., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *JEP: General*, *144*(4), 744.

Bartlett, F. (1932). *Remembering*. Cambridge.

Berlin, B., & Kay, P. (1969). *Basic color terms: their universality and evolution.*

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

Brady, T., & Alvarez, G. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *J Vision*, *15*, 6–6.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cog. Psych.*, *13*(2), 207-230.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision research*, *43*(4), 393-404.

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition*, *34*(8), 1754–1768.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuro.*, *17*(5), 738-743.

Hemmer, P., & Steyvers, M. (2009). A bayesian account of reconstructive memory. *Topics Cogn. Sci.*, *1*(1), 189-202.

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: Two classes of attractors at work. *J Vision*, *10*(2), 24.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *JEP: General*, *129*(2), 220.

Ishihara, S. (1987). Ishihara's tests for colour-blindness (concise ed.). *Tokyo, Japan: Kanehara & Co*.

Lew, T., & Vul, E. (2015). Structured priors in visual working memory revealed through iterated learning. In *Proc. 37th Cognitive Science Society* (p. 1332-1337).

Lin, P.-H., & Luck, S. J. (2012). Proactive interference does not meaningfully distort visual working memory capacity estimates in the canonical change detection task. *Frontiers in Psychology*, *3*, 42.

Makovski, T., & Jiang, Y. V. (2008). Proactive interference from items previously stored in visual working memory. *Memory & Cognition*, *36*(1), 43–52.

Panichello, M. F., DePasquale, B., Pillow, J. W., & Buschman, T. J. (2019). Error-correcting dynamics in visual working memory. *Nature Comm.*, *10*(1), 1-11.

Persaud, K., & Hemmer, P. (2014). The influence of knowledge and expectations for color on episodic memory. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 36).

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *PNAS*, *104*(4), 1436–1441.

Sanborn, A., & Griffiths, T. L. (2008). Markov chain monte carlo with people. In *NIPS* (p. 1265-1272).

Schurgin, M., Wixted, J., & Brady, T. (2018). Psychological scaling reveals a single parameter framework for visual working memory. *BioRxiv*, 325472.

Webster, M. A., & Kay, P. (2012). Color categories and color appearance. *Cognition*, *122*(3), 375–392.

Wilken, P., & Ma, W. J. (2004). A detection theory account of visual short-term memory for color. *J Vision*, *4*(8), 150.

Yu, A. J., & Cohen, J. D. (2009). Sequential effects: superstition or rational behavior? In *NIPS* (pp. 1873–1880).