Collaborative Encoding of Visual Working Memory

Huang Ham[1], **Evan Russek**[1], **Thomas L. Griffiths**[1,2], **Natalia Vélez**[1]
[1]Department of Psychology, Princeton University
[2]Department of Computer Science, Princeton University
{hamhuang, evrussek, tomg, nvelez}@princeton.edu

Abstract

Collaboration helps humans surmount individual cognitive limitations by distributing information over many minds. However, figuring out when and how to collaborate is not trivial. This study examines whether dyads split up information in a collaborative visual working memory task when doing so improves performance. Participants (N=356) memorized grids of 4, 16, or 36 images both alone and with a partner. We used a visual working memory model to estimate how much dyads would benefit from splitting up a grid of images, rather than each memorizing the grid independently. Our model predicts that participants should split up grids that are neither too easy nor too difficult to benefit from collaboration. Indeed, participants tacitly adopted conventions to split up medium and large grids—and were more accurate in these conditions when they worked together than when they acted alone—but not small grids where individual performance was already at ceiling. Our work provides a first step to understand how decisions about when and how to collaborate are shaped by the adaptive use of cognitive resources.

Keywords: collaboration; social cognition; coordination; visual working memory; computational modeling

Introduction

Individuals have limited capacity to store, process and utilize information. Collaboration provides a means for humans to surmount these limitations by distributing information over many minds (Griffiths, 2020; Vélez, Christian, Hardy, Thompson, & Griffiths, 2023). There is a rich literature that suggests that memories are not stored solely within individual minds, but are also maintained through social interactions (Wegner, Erber, & Raymond, 1991; Coman, Momennejad, Drach, & Geana, 2016; Momennejad, Duker, & Coman, 2019). To reap the benefits of distributing information, it is important to form a reusable *convention* to determine who keeps track of what (Lewis, 2008). For example, one member of a household might keep track of when the plants need watering, while the other might remember to put out the trash (Wegner et al., 1991). How do people arrive at conventions to distribute information in memory, and when is it beneficial to do so?

Prior work has largely examined how people adopt such conventions in the face of *external* constraints, such as in search tasks where each participant can check a limited number of locations (Goldstone, Andrade-Lotero, Hawkins, & Roberts, 2024; Andrade-Lotero & Goldstone, 2021). Prior studies have found that people are more likely to form stable conventions when the payoffs at stake are high (Hawkins & Goldstone, 2016), which suggests that people may decide whether to collaborate by weighing the benefits of doing so.

However, it is an open question whether people are also sensitive to *internal* cognitive constraints when deciding whether and how to split up tasks. In many domains, the amount of cognitive effort people are willing to spend depends on the potential benefit (Lieder & Griffiths, 2019). If this is the case, participants may be less likely to collaborate when tasks are so easy that performance is high even if people act alone, or so hard that performance is low even if people collaborate.

Methods

Participants: To test this prediction, participants ($N = 356$) completed a visual working memory task both alone and with a partner through Prolific. An additional 534 participants were excluded because they or their partner dropped out before completing at least 50% of trials, following preregistered exclusion criteria ($N = 534$ excluded; preregistration available at `https://aspredicted.org/p624-s347.pdf`). This attrition rate is typical of online multiplayer studies and did not differ by condition ($\chi^2(2) = 2.3, p = .3$).

Procedure: Participants were assigned to three between-subjects conditions where studied grids of 4, 16, or 36 images. Each square on the grid contained an image of a face, house, limb, or object selected from the stimulus set in Stigliani, Weiner, and Grill-Spector (2015). To track which images participants studied, we covered each image with a gray square that participants could remove by hovering their mouse over the square. Each trial consisted of an *encoding phase* where participants studied the grid for 10s and a *retrieval phase* where participants were asked to report the image that was hidden behind a randomly-cued square.

Each participant completed this task both alone (solo trials) and with another participant (dyadic trials). Participants were allowed 2 min. to communicate via chat before being shown task instructions, which enabled participants to build rapport without explicitly discussing strategy. In dyadic trials, participants saw an orange border that indicated which image their partner was currently studying. Thus, participants could tacitly arrive at a convention to split up the grid by avoiding squares that their partner studied, e.g., by sticking to the left side of the grid while their partner studied the right. Participants were rewarded for their own responses in solo trials and for the first response submitted in dyadic trials. Participants completed 20 solo trials, 40 dyadic trials, and 6 catch trials where all squares contained the same image.

Computational model: To estimate the benefits of collaboration, we used a visual working memory model adapted from Suchow and Griffiths (2016). The model assumes that participants have a fixed resource budget of N discrete units, which they must allocate across K stimuli arranged on a grid. When prompted to recall one of the K items, the probability of successful recall depends on the amount of resource allocated to that item, following a concave function $f(Q) = \frac{-1}{2} + \frac{\frac{3}{4}}{1+e^{-0.5Q}}$. If the agent plays alone or ignores the presence of a partner, the optimal strategy is to distribute resources equally across all stimuli. In this case, each stimulus receives N/K units. In contrast, if the agent coordinates with a partner using a shared convention—such as each remembering only half the stimuli—they can allocate their N units more narrowly across $K/2$ stimuli, assigning $N/(K/2) = 2N/K$ units to each. The benefit of collaboration is then the difference in recall probability between these two cases, $f(\frac{2N}{K}) - f(\frac{N}{K})$.

Figure 1A shows the benefit of collaborating at varying grid sizes, using a more general form of this model that allows for individual participants to differ in their resource capacity and thus split the grid non-equally where the participant with a larger memory capacity takes on a larger portion of the grid. When K is small, both $f(N/K)$ and $f(2N/K)$ are close to 1, so the difference is small. As K increases, both probabilities decrease, but $f(2N/K)$ decreases slower, leading to a larger benefit. However, when K becomes very large, both terms approach 0, and the benefit again diminishes. Thus, collaboration brings the greatest benefits at intermediate grid sizes.

To apply this model to our task, we estimated each player's resource capacity, N, based on their mouse-click and recall behavior in solo trials. We fit the model using a version of simulation-based inference (Russek, Callaway, & Griffiths, 2024; Rmus, Pan, Xia, & Collins, 2024), training GRUs to estimate N from behavior on simulated data, and then applying the trained network to participant's actual data.

Results

Consistent with model predictions, participants benefited the most from collaborating in intermediate grid sizes (dyadic vs. solo performance in grid size 16: $t(237.90) = -4.28, p < 0.001$; Fig. 1B). Conversely, we saw no benefit for small grids, where solo performance was already at ceiling ($t(172.50) = -0.56, p = 0.58$). Interestingly, we also found that participants benefited from collaborating in the largest grids (dyadic vs. solo performance in grid size 36: $t(244.32) = -3.66, p < 0.001$), which suggests that our model may not have been perfectly calibrated to human performance.

To measure whether dyads split up the grid, we measured how their movement patterns differed between solo and dyadic trials. Our measure of collaboration, **spatial overlap**, reflects how much time participants spent studying the same images. We derived a vector representation of each player i's time allocation, $\mathbf{t}_i$, by measuring the proportion of the 10-second encoding phase that they spent hovering over each tile k. Spatial overlap measures how similar player 1 and 2's time allocations are to each other: $\sum_k \frac{\min(t_1(k),t_2(k))}{10}$. Because spatial overlap can vary based on grid size, we computed *relative* spatial overlap by taking the ratio of each dyad's average spatial overlap in solo and dyadic trials. If participants split up the grid, we would expect to see less spatial overlap in dyadic trials (relative spatial overlap < 1). Consistent with this prediction, participants' spatial overlap dropped in dyadic trials for medium and large grid sizes (mean rel. spatial overlap in grid size 16: 0.81, grid size 36: 0.80; grid size 16 vs. 36: $t(118.50) = 0.08, p = 0.94$)—that is, in precisely the conditions where participants benefited from doing the task together—but not for small grid sizes (mean in grid size 4: 1.24; grid size 4 vs. 16: $t(65.49) = 2.87, p = 0.005$).

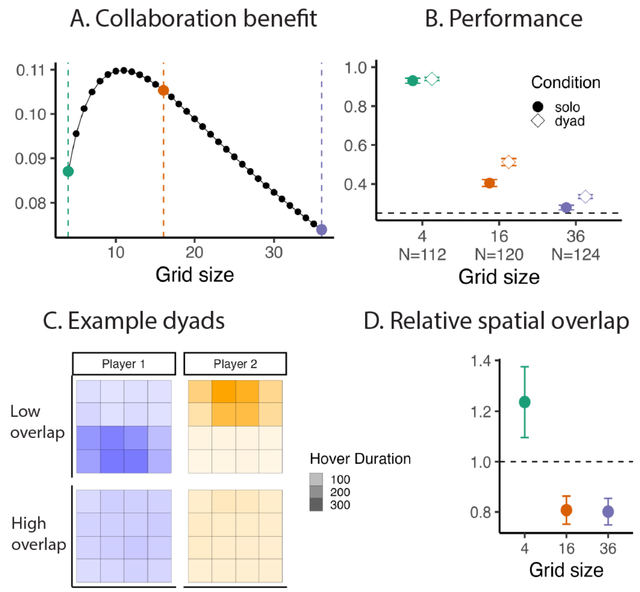

Figure 1: A: Collaboration benefit, operationalized as the model-predicted difference in performance if agents divide grids optimally vs. each memorize the grid independently. B: Average human performance in solo and dyadic trials. Dashed line shows chance performance. C. Two sample dyads with low vs. high spatial overlap in dyadic trials; heatmaps show average time allocations $\mathbf{t}_i$ across all dyadic trials. D. Average relative spatial overlap (dyadic/solo overlap) by grid size. Error bars denote standard error of the mean.

Conclusion

Decisions about when and how to collaborate are decisions about how to best allocate cognitive resources. We examined the extent to which these decisions are adaptive: do people collaborate more when collaboration is more beneficial? We found that participants collaborated more on a visual working memory task when the task was too hard to successfully memorize the stimuli on one's own, supporting the idea that decisions about collaboration involve the adaptive use of cognitive resources. However, participants even collaborated in tasks that were so difficult that our model predicted no benefit to collaboration. Since people actually benefited from collaboration in these settings, these results suggest that our model may not have been perfectly calibrated to human performance. As we continue this work, we plan to refine our model of human performance to better understand when two minds are better than one.

Acknowledgments

We thank Miles Sugarman and the support team of Empirica for their assistance on the design and implementation of the task. We thank Jingjing Li, Ti-fen Tian, Fred Callaway, Robert Hawkins, Qiawen (Ella) Liu, Alexandra Witt, Charley Wu, Harrison Ritz, Caroline Jahn, Lily Tsoy, Xiaoyi Liu, and Zhengang

Lu for helpful discussion and feedback. This work was supported by grant #20648 from the Templeton World Charity Foundation to TLG and NV.

References

Andrade-Lotero, E., & Goldstone, R. L. (2021, 07). Self-organized division of cognitive labor. *PLOS ONE*, *16*(7), 1-22. doi: 10.1371/journal.pone.0254532

Coman, A., Momennejad, I., Drach, R. D., & Geana, A. (2016). Mnemonic convergence in social networks: The emergent properties of cognition at a collective level. *Proceedings of the National Academy of Sciences*, *113*(29), 8171–8176.

Goldstone, R. L., Andrade-Lotero, E. J., Hawkins, R. D., & Roberts, M. E. (2024). The emergence of specialized roles within groups. *Topics in Cognitive Science*, *16*(2), 257–281.

Griffiths, T. L. (2020). Understanding human intelligence through human limitations. *Trends in Cognitive Sciences*, *24*(11), 873–883.

Hawkins, R. X., & Goldstone, R. L. (2016). The formation of social conventions in real-time environments. *PloS one*, *11*(3), e0151670.

Lewis, D. (2008). *Convention: A philosophical study*. John Wiley & Sons.

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. doi: 10.1017/S0140525X1900061X

Momennejad, I., Duker, A., & Coman, A. (2019). Bridge ties bind collective memories. *Nature communications*, *10*(1), 1578.

Rmus, M., Pan, T.-F., Xia, L., & Collins, A. G. (2024). Artificial neural networks for model identification and parameter estimation in computational cognitive models. *PLOS Computational Biology*, *20*(5), e1012119.

Russek, E. M., Callaway, F., & Griffiths, T. L. (2024). Inverting cognitive models with neural networks to infer preferences from fixations. *Cognitive Science*, *48*(11), e70015.

Stigliani, A., Weiner, K. S., & Grill-Spector, K. (2015). Temporal processing capacity in high-level visual cortex is domain specific. *Journal of Neuroscience*, *35*(36), 12412–12424.

Suchow, J. W., & Griffiths, T. (2016). Deciding to remember: Memory maintenance as a markov decision process. In *38th annual meeting of the cognitive science society.*

Vélez, N., Christian, B., Hardy, M., Thompson, B. D., & Griffiths, T. L. (2023). How do humans overcome individual computational limitations by working together? *Cognitive science*, *47*(1), e13232.

Wegner, D. M., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of personality and social psychology*, *61*(6), 923.