Serial Reproduction with Multimodal Information

Byron Zhang (zishuoz@princeton.edu)
Department of Computer Science, Princeton University
Princeton, NJ 08540 USA

Raja Marjieh (raja.marjieh@princeton.edu)
Department of Psychology, Princeton University
Princeton, NJ 08540 USA

Thomas L. Griffiths (tomg@princeton.edu)
Departments of Psychology and Computer Science, Princeton University
Princeton, NJ 08540 USA

Abstract

Humans endode bias from prior experiences when perceiving information. Serial reproduction is a standard method used to study these priors by amplifying human bias through a telephone-like psychology experiment, where one subject observes a stimulus and is asked to reproduce it for the next subject, forming a chain of reproductions. While past work in serial reproduction operates on only one modality per chain, information observed by one agent may be passed onto another agent in a different modality. In this study, we examine multimodal serial reproduction by asking subjects who receive a visual stimulus reproduce it in textual format, and vice versa. Our analysis shows noticeable differences between the dynamics and distribution of the unimodal and multimodal chains. In particular, participants in the multimodal study made more errors in the reconstructions phase, and generated visual stimuli with more symmetry and structure than those in the unimodal study, suggesting that language acts as an important factor in the transmission of visual information among humans.

Keywords: serial reproduction, Bayesian inference, multimodal learning

Introduction

Transmission of information among humans is rarely perfectly accurate. As we receive pieces of information such as images, text, and sound, we encode them with noise that relies on our prior beliefs. This can lead to biases, where the reproduced response differs from the true stimulus. Serial reproduction (Bartlett, 1932) is a method used to elicit shared priors coming from biases in human perception and memory through a telephone-game-like experiment. Subjects are asked to pass on a piece of information to one another in sequence, with each subject reproducing the information for the next subject. This process is often used to study how the original stimulus changes as it is passed from one person to the next, and how these changes may be influenced by various factors such as inductive biases and cultural differences.

While serial reproduction is often used to study the cultural transmission of information, current studies do not consider the multimodal nature of information transmission. Most experiments involving human subjects constructing a serial reproduction chain employ only one modality of information. However, in the real world, information observed by one agent may be passed onto another agent in a different modality. For example, if an agent initially observes a photo, then the agent may encode the information from the photo and retell it in the form of a story or a piece of text for another agent.

In this paper, we aim to study the properties of multimodal serial reproduction and compare its differences with unimodal serial reproduction. Specifically, we focus on the modalities of vision and language and how adding language to a serial reproduction chain influences the dynamics and distributions of visual stimuli. To simulate multimodal serial reproduction, participants who observed a visual stimulus were asked to reproduce a textual stimulus for the next participant and vice versa. As a result, the multimodal serial reproduction chain will alternate between the visual and language modalities. To establish the comparison, we collected data from human participants on two serial reproduction chains with the same initial visual stimuli, one for the unimodal case and one for the multimodal case.

We start by providing the theoretical and practical implications of unimodal serial reproduction in cognitive psychology, statistics, and computer science. Then, we present a theoretical model of multimodal case by relating to unimodal serial reproductino. Our experimental results demonstrate multiple differences that the dynamics and distributions elicited by the two serial reproduction chains. Specifically, the multimodal reproductions yielded more errors from the stimulus, and the visual stimuli generated in the multimodal chain exhibited more symmetry and recognizable structure than the unimodal chain.

Background

Unimodal Serial Reproduction

The concept of serial reproduction was first proposed by Bartlett (1932) to study how bias from human's previous experiences influences how they perceive new experiences. In a unimodal serial reproduction study, the original stimulus is presented to the first participant, who then reproduces it for the second participant, and so on until the final participant reproduces the stimulus. Formally, serial reproduction can be interpreted as a Gibbs Sampler over a pair of variables (x, μ), where x represents the distribution of stimuli in the world, and *mu* represents the states people infer from those stimuli. The Gibbs Sampler takes the form of a Markov Chain $...x_t \to \mu_t \to x_{t+1} \to \mu_{t+1}...$. There are two existing mathematical interpretations of this Markov Chain. One interpretation by Xu and Griffiths (2008) treats x_t as a noisy stimulus, and assumes that humans have accumulated a prior distribution $p(\mu)$ about the true world through previous experiences. To reconstruct the observed stimulus x_t, humans try to estimate the true state of the world μ_t by sampling the posterior distribution $p(\mu_t|x_t)$. Another interpretation used by Jacoby and McDermott (2017); Langlois, Jacoby, Suchow, and Griffiths (2021) treats x_t as a true stimulus. Humans encode stimulus x_t as a noisy percept mu_t through the likelyhood $p(\mu_t|x_t)$. To reconstruct the x_t, the percept μ_t is decoded by sampling the posterior distribution $p(x_{t+1}|\mu_t)$.

Unimodal serial reproduction has been widely used to study perceptual and spatial priors of the human visual system. For example, Langlois et al. (2021) conducted a large-scale serial reproduction experiment in which participants are given a point on an image, then asked to reproduce the location of that point. This experiment allowed the authors to analyze the allocation of visual encoding resources to specific locations in a visual stimulus. Xu and Griffiths (2008) asked participants to be first trained to distinguish the width of fishes, then asked to take part in a serial reproduction experiment where they reproduce the width of fishes for future iterations.

In addition to applications in cognitive psychology, unimodal serial reproduction also draws a parallel with the paradigm of diffusion probablistic models (Song, Meng, & Ermon, 2021), which are becoming prominent in generative modeling in computer science and artificial intelligence. Marjieh, Sucholutsky, Langlois, Jacoby, and Griffiths (2022) proposed to use the mechanisms of serial reproduction as a tool to theoretically study the effect of different noise classes on a diffusion chain. Understanding the dynamics of multimodal serial reproduction may inspire lines of work in diffusion-based text-to-image synthesis.

Related Paradigms

Serial reproduction falls into a broader category of elicitation methods that enable cognitive psychologists, statisticians, and computer scientists to study human priors. These methods are all iterative in nature, requiring human participants to observe and reproduce some stimulus for the next participant and amplifying the bias from prior experiences as the iteration increases. Representative methods include iterated learning (Kirby, Griffiths, & Smith, 2014), Markov-Chain Monte-Carlo with People (MCMCP) (Sanborn & Griffiths, 2007), and Gibbs Sampling with People (GSP) (Harrison et al., 2020).

While these techniques share similar theoretical grounding, it is important to draw the distinction with serial reproduction for the clarity of our study and future work. To the best of our knowledge, there are also very few studies that involve multimodal stimulus on these prior elicitation methods. Iterated learning is used to study priors that humans use when learning new knowledge, while MCMCP and GSP are more general methods that can elicit priors in multiple psychological structures of interest. For example, GSP was employed by Kumar et al. (2022) to elicit human prior distribution over binary boards. Our experiments in multimodal serial reproduction will also utilize binary boards as channels of visual representations.

Multimodal Serial Reproduction

To best illustrate the effect of switching between modalities during information transmission, we set up our theoretical analysis and experiment with alternating modalities for every iteration, and we consider only two modalities. The original stimulus is presented to the first participant, who will then be asked to reproduce it in a different modality. In other words, in a multimodal serial reproduction study, the stimulus at time step t and $t+1$ will be different, while the stimulus at t and $t+2$ will be the same.

Formally, a participant receives a stimulus x_t in a modality X, and then is asked to reproduce x_t in a different modality Y to yield y_{t+1}. The next participant observes y_{t+1}, and is tasked with reproducint it in the form of x_{t+2}, and so on. Similar to unimodal serial reproduction, we can interpret multimodal serial reproduction as a Gibbs sampler over the distribution of the pair of variables (x,y), forming a Markov Chain $...x_t \to_{t+1} \to x_{t+2} \to_{t+3} ...$. Without loss of generality, we assume that we want to study the probability distribution over stimuli in the modality X. Since serial reproduction is often used to study *shared* priors among communities of agents, we also assume agents asked to reproduce x and y share the prior $p(x)$. Our interpretation of the Gibbs sampler is similar to Jacoby and McDermott (2017), as we treat the stimuli y_t as an explicit representation of the noisy percept mu_t, generated by the likelihood $p(y_{t+1}|x_t)$. The subjects then reproduce the stimulus by sampling the posterior distribution $p(x_{t+2}|y_{t+1})$. Note that for simplicity of notations, we can also combine timesteps t into one iteration τ, such that theoretical analysis from Jacoby and McDermott (2017); Langlois et al. (2021) will follow.

In this study, we examine the modalities of vision and language, primarily focusing on the differences between unimodal and multimodal serial reproduction in the *visual modality* through a simple pattern memorization game. To construct the multimodal serial reproduction chain, participants will either be asked to describe a visual pattern in succinct language or reproduce a visual pattern using language descriptions.

Method

Participants

Participants are recruited from Amazon Mechanical Turk and prescreened with two tasks to ensure quality of data. The Lexical Decision Task (Lemhöfer & Broersma, 2012) evaluates fluency in English by asking each participant to determine the validity of 10 English words. The Colorblindness Task eval-

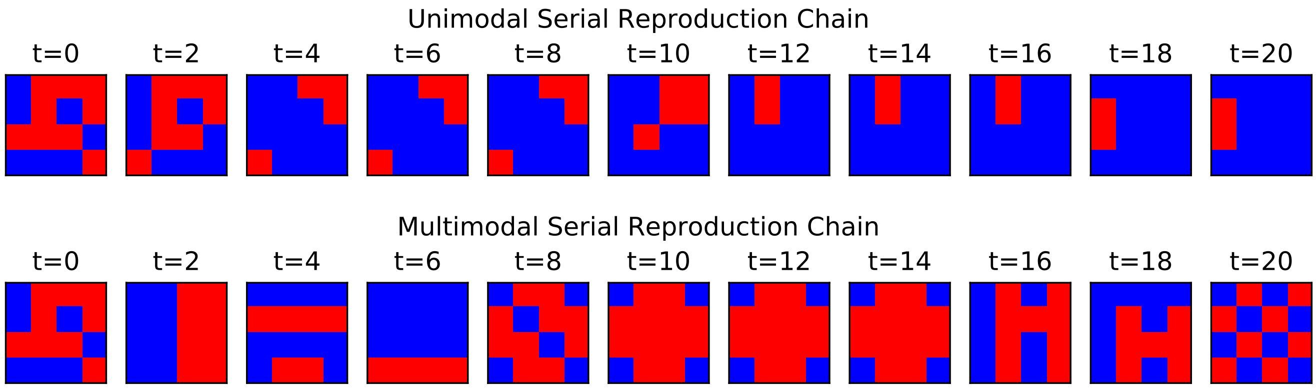

Figure 1: **Example reproduction chains.** The board at iteration t=0 is the initial stimulus given to the first participant. Note that all boards after t=0 in the unimodal serial reproduction chain are constructed through directly memorizing boards at $t-1$, while those in the unimodal serial reproduction chain are constructed through language instructions from the participant at $t-1$.

uates participants' ability to distinguish colors by asking participants to identify the numbers on five Pseudoisochromatic plates. Participants must score 8/10 on the Lexical Decision Task and 4/5 on the Colorblindness Test.

Materials

The stimulus space is set to be consistent with Kumar et al. (2022), including all 4x4 boards, thus 16 dimensions. Each dimension accepts binary values that represent the color of the tile, which can be either red or blue. The initial stimulus for each serial reproduction chain is uniformly sampled board from the stimulus space. We sampled 50 boards to collect experimental results.

Procedure

The serial reproduction interface was programmed using the PsyNet platform. We collected both unimodal and multimodal serial reproduction chains of length 20 for each initial stimulus, as described below.

Unimodal Serial Reproduction A participant asked to memorize a stimulus board that is shown for 10 seconds, and then tasked with reproducing the board. The new board will serve as the stimulus for the next iteration in the chain.

Multimodal Serial Reproduction A participant can be shown either a stimulus board, or a string of textual description for 5 seconds. If shown a board, then they are asked to memorize the board and provide an accurate textual description of the board. If shown a string of textual descriptions, then they are asked to reproduce a board that most accurately illustrates the textual description. The new board or text will serve as the stimulus for the next iteration in the chain.

To produce the stimulus, participants will click on a tile in an online interface, and the tile will alternate colors with every click. To adjust for coloring bias, the initial color of associated with the first click in both unimodal and multimodal serial reproduction studies will be randomized.

Results

Board Dynamics

Through a qualitative analysis presented in Figure 1, we see immediately that the unimodal serial reproduction chain appears to be more stable, although the patterns are less diverse and more sparse than the multimodal serial reproduction chain.

We also analyzed the dynamics of the serial reproduction chain through two quantitative measures. First, the average error of the boards from the original stimulus measures the number of tiles with different colors from the tiles in the board at $t = 0$. As shown in Figure 2, the differences in error between unimodal and multimodal serial reproduction are small, except for iteration 2. Since the initial stimulus is generated randomly, it is more difficult for participants to describe and reproduce it using natural language than if they memorized the board directly, as the stimulus most likely lacks apparent structure. In addition, a high error at an earlier iteration will likely persist because participants in later iterations will be asked to reproduce the erroneous boards.

Then, we examine the average error of boards over a time horizon, which measures the number of differently colored tiles between boards at time t and $t+h$, where h is the time horizon. Figure 2 shows a clear pattern that larger time horizons result in larger differences between the boards, since more noise is injected into the chain as time passes. In this case, we observe consistently noticeable differences between errors in unimodal and multimodal serial reproduction chains. The signal from unimodal chains seems to persist for a longer time horizon, for which we suspect that more noise from prior beliefs is encoded in the multimodal serial reproduction chain.

As a note on the analyses on the board dynamic, we are only concern about the *pattern* of the boards instead of individual color values. For example, a board with all red tiles would be considered the same as a board with all blue tiles,

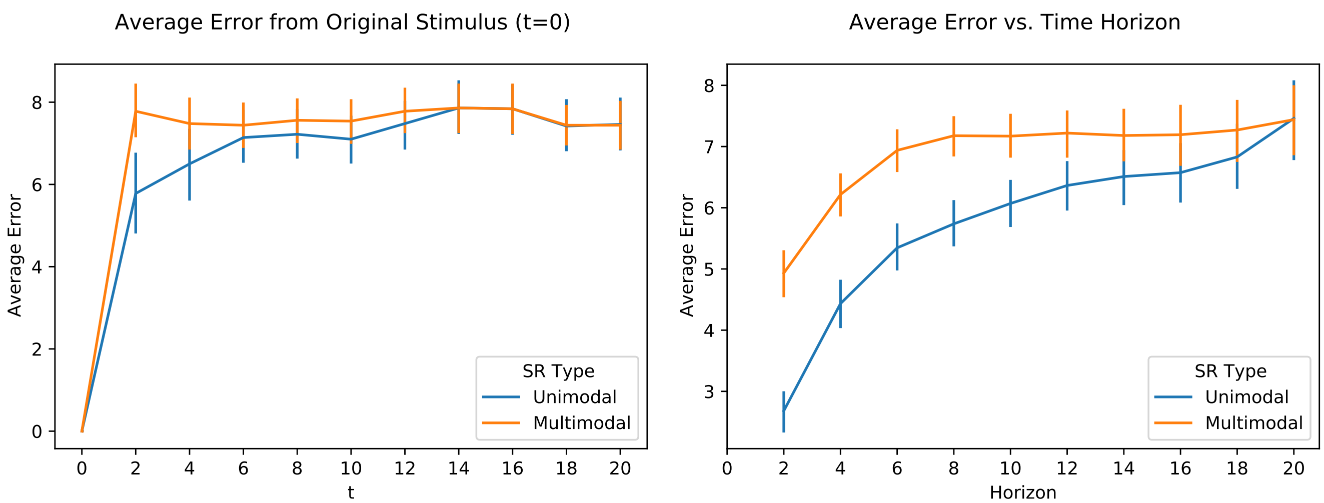

Figure 2: **Error dynamics between time steps.** *Left.* The average error of between boards at iteration t and iteration 0 for all chains. Participants in unimodal serial reproduction shows slightly less error from the original board at the earlier iterations, behaves similarly toward later iterations. *Right.* The average error between boards at different time horizons. Participants in unimodal serial reproduction showed noticeably less error for most iterations, but the error rate becomes similar to that of multimodal serial reproduction on iteration 20.

as the pattern between these boards is the same. To calibrate the differences in colors, we always designate the top-left tile (row 0, column 0) as blue. In other words, we invert the color of the entire board if its top left tile is red.

Board Distribution

We also examined the distributions formed by the boards by inspecting the most frequent boards produced across all iterations, as shown in Figure **??**. It is apparent that the distribution of boards elicited by unimodal serial reproduction is often sparse, asymmetrical, and does not exhibit any patterns describable by simple text, while the distribution of boards elicited by multimodal serial reproduction is highly symmetrical and structural. We hypothesize that this difference is attributed to the shape bias from the human visual system (Canini, Griffiths, Vanpaemel, & Kalish, 2011; Tartaglini, Vong, & Lake, 2022), which describes the tendency for people to attribute certain characteristics to objects based on their shape. Because natural language allows humans to explicitly spell out these characteristics, multimodal serial reproduction more effectively elicits the shape bias. As we will see in the next section, shapes are frequently employed as descriptors of boards in the multimodal serial reproduction chain.

Similar to board dynamic, the analysis on board distribution also concerns only the pattern of the boards instead of individual values. We assume consider the top left tile is blue, and invert the color of the entire board if that tile is red.

Text Characteristics

While our study focuses on the visual stimulus, we would also like to examine what textual descriptions the participants were producing and what elicited the boards in the multimodal serial reproduction chain.

Table 1: Top 20 Most Frequent Nouns in Multimodal SR

Word	Count	Word	Count
red	740	center	65
blue	556	right	62
color	255	L	52
box	223	board	42
square	144	corner	42
top	137	shape	42
bottom	102	middle	36
left	95	horizontal	34
row	75	side	31
line	66	I	26
center	65	T	25

Table 1 shows the most frequent nouns and adjectives the participants used to describe boards in multimodal serial reproduction, and two apparent biases are presented. First, color bias is apparent, as frequency of color descriptors (red and blue) are significantly higher than all other words. However, part of the high frequency can be attributed to the binary nature of the colors on the board, because describing only two colors is simple. Second, shape bias is illustrated more explicitly by the tendency of participants to describe many board patterns as a letter (L, I, T in the table), identifiable features such as "corner," or simply using the word "shape."

We only consider nouns and adjectives because we are only interested in what descriptors are used to *qualify* the boards.

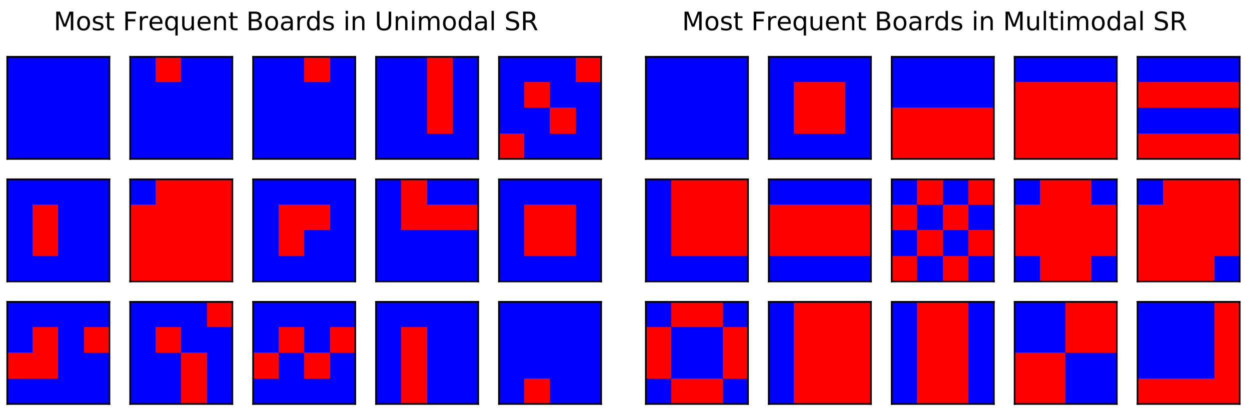

Figure 3: **Distribution of human priors.** *Left.* Top 15 most frequently reproduced patterns in unimodal serial reproduction. *Right.* Top 15 most frequently reproduced patterns in multimodal serial reproduction. The patterns in the multimodal serial reproduction chain exhibit more structure and symmetry than the ones from the unimodal serial reproduction chain.

Although participants very frequently use words such as "on" and "the" occur in their descriptions, these words are the least related to how humans internally represent the boards.

Discussion

Previously, serial reproduction has been used to elicit shared human priors and study the theory of cultural transmission. However, natural information can often be represented by different modalities, and stimulus from one modality can be passed on from one agent to another in a different modality. The current paradigm in serial reproduction and similar iterative methods fail to consider the multimodality nature of information transmission. This paper presents a preliminary study on incorporating multimodal information in the serial reproduction method, providing both theoretical and empirical analysis on multimodal serial reproduction. Although tools to theoretically analyze unimodal serial reproduction can be applied to multimodal serial reproduction, our experiments demonstrated that the dynamics and distribution over their respective chains exhibit multiple modes of difference. First, multimodal serial reproduction tends to be less stable and stays accurate for shorter time horizons. Second, multimodal serial reproduction generated more symmetrical, recognizable patterns, whereas unimodal serial reproduction generated more sparse, asymmetrical patterns. Next, we will discuss several limitations of the study and potential directions to expand on this new paradigm in serial reproduction.

Limitations

The main limitation of this study is the small amount of data we were able to collect due to resource constraints. We only collected chains of length 20, which may not be enough for us to observe what patterns either chain will converge to, and if it would be possible for the chains to converge. In addition, since multimodal serial reproduction alternates between modalities, we could only obtain half of the visual stimulus that was generated by unimodal serial reproduction, which does not provide the fairest comparison between the two reproduction chains.

We were also unable to avoid noise from Amazon Mechanical Turk workers who did not follow the instructions presented in the study. For example, boards that were entirely blue or red were much more common than any other boards, by a large margin. This phenomenon occurs because when one worker paints an entire board as a single color, this pattern is extremely simple to reproduce, either through text or direct visuals. As a result, all subsequent iterations of the chain will collapse to a board that is entirely red or blue.

Future Directions

Since multimodal serial reproduction is a new, underexplored paradigm, there are many possible future directions to continue investigating, spanning multiple applications.

First, multimodal serial reproduction can potentially be used as a means to sample higher dimensional stimulus spaces. While Kumar et al. (2022) elicited human prior distribution over 4x4 boards using GSP (Harrison et al., 2020), they found that GSP was unable to converge for boards of any larger sizes. In the future, it may be worth investigating the behavior of multimodal serial reproduction over a diverse set of larger stimuli spaces. For example, in the case of the boards in our experiments, we can examine the chain dynamics on boards of larger dimensions or with more color options.

Second, serial reproduction is one of many iterative methods used to study human priors. It would be interesting to observe the behavior of other chains of multimodal information transmission using techniques such as GSP (Harrison et al., 2020), and MCMCP (Sanborn & Griffiths, 2007), and compare the differences between the unimodal and multimodal cases.

Lastly, as mentioned previously, serial reproduction relates closely to diffusion probabilistic models employed in text-to-image synthesis (Marjieh et al., 2022). However, most text-to-image synthesis pipelines perform diffusion directly in the latent space of the shared vision and language representation, which is not possible through means of unimodal or multimodal serial reproduction. Would multimodal serial reproduction provide an alternative, cognitively-inspired way to train and deploy diffusion models?

Acknowledgments

We would like to thank to Raja Marjieh for valuable guidance on brainstorming the topic, running experiments on Amazon Mechanical Turk, and streamlining the analysis; Dr. Nori Jacoby for setting up and providing documentation on the PsyNet interface; and Professor Tom Griffiths for insightful comments on the direction of the project.

References

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* New York, NY, US: Cambridge University Press.

Canini, K., Griffiths, T., Vanpaemel, W., & Kalish, M. (2011, 01). Discovering inductive biases in categorization through iterated learning.

Harrison, P., Marjieh, R., Adolfi, F., van Rijn, P., Anglada-Tort, M., Tchernichovski, O., … Jacoby, N. (2020). Gibbs sampling with people. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 10659–10671). Curran Associates, Inc.

Jacoby, N., & McDermott, J. H. (2017, January). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Curr Biol*, *27*(3), 359–370.

Kirby, S., Griffiths, T., & Smith, K. (2014, July). Iterated learning and the evolution of language. *Curr Opin Neurobiol*, *28*, 108–114.

Kumar, S., Correa, C. G., Dasgupta, I., Marjieh, R., Hu, M., Hawkins, R. D., … Griffiths, T. L. (2022). Using natural language and program abstractions to instill human inductive biases in machines. In A. H. Oh, A. Agarwal, D. Belgrave, & K. Cho (Eds.), *Advances in neural information processing systems.*

Langlois, T. A., Jacoby, N., Suchow, J. W., & Griffiths, T. L. (2021). Serial reproduction reveals the geometry of visuospatial representations. *Proceedings of the National Academy of Sciences*, *118*(13), e2012938118. doi: 10.1073/pnas.2012938118

Lemhöfer, K., & Broersma, M. (2012, June). Introducing LexTALE: a quick and valid lexical test for advanced learners of english. *Behav Res Methods*, *44*(2), 325–343.

Marjieh, R., Sucholutsky, I., Langlois, T. A., Jacoby, N., & Griffiths, T. L. (2022). *Analyzing diffusion as serial reproduction.* arXiv. doi: 10.48550/ARXIV.2209.14821

Sanborn, A., & Griffiths, T. (2007). Markov chain monte carlo with people. In J. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems* (Vol. 20). Curran Associates, Inc.

Song, J., Meng, C., & Ermon, S. (2021). Denoising diffusion implicit models. In *International conference on learning representations.*

Tartaglini, A. R., Vong, W. K., & Lake, B. M. (2022). *A developmentally-inspired examination of shape versus texture bias in machines.* arXiv. Retrieved from `https://arxiv.org/abs/2202.08340` doi: 10.48550/ARXIV.2202.08340

Xu, J., & Griffiths, T. (2008). How memory biases affect information transmission: A rational analysis of serial reproduction. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems* (Vol. 21). Curran Associates, Inc.