

Non-commitment in mental imagery is distinct from perceptual inattention, and supports hierarchical scene construction

Yichen Li ^{*}, Veronika Hammond [†], Tomer D. Ullman ^{*}

Abstract

When calling an image before the mind’s eye, people often do not include basic properties of the mental scene. For example, when imagining a person knocking a ball off a table, people may neglect to consider the color of the ball. This “non-commitment” in imagery has been part of ongoing debates on the nature of mental imagery in philosophy and cognitive science, but the central question of its origin remains open. It is possible that this lack of detail in the imagination reflects actual non-commitment, as one constructs a hierarchical mental scene. But it is also possible that people do form highly detailed mental images, but fail to notice, encode, or recall properties of the image, for the same reasons that they do not notice, encode, or recall details of real scenes in direct perception. We tested these contrasting hypotheses across three experiments (N=3,895 total). First, we compared non-commitment patterns in mental imagery with the details that participants failed to notice in briefly presented real images. We found no correlation between properties that were non-committed to in the imagination, and those missed under rapid perception, indicating that non-commitment does not simply arise from perceptual noise. Next, we examined the temporal unfolding of mental imagery, and found consistent evidence of a hierarchical pattern: core entities and spatial relations, then physical properties, then surface details like color and texture. Attempts to interrupt or constrain scene construction did not change the hierarchical ordering. We suggest the results support a hierarchical partial-scene construction account of mental imagery.

^{*}Department of Psychology, Harvard University, Cambridge, MA, 02138. Contact: yichenli@fas.harvard.edu

[†]Department of Philosophy, Harvard University, Cambridge, MA, 02138

Introduction

Please imagine the following, as vividly as you can:

“Dropping something next to another thing.”

Hold the image in your mind for a few moments, then read on.

Did you imagine what thing was dropped? Did you see the movement and flow of the objects? How about the color of the thing being dropped? Did you imagine a surface or floor that supported the things? Was there someone doing the dropping?

If you are like most people, you probably answered “yes” to some of these questions, and “no” to others. Moreover, if you are like most people, you probably had the sense that some things were imagined before others. Rather than the whole scene coming to life like a fade-in, perhaps you had the overall sense of *how* a thing was dropped, before you knew *which* thing was dropped.

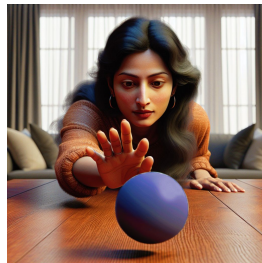
The imagination is often likened to a painting in the mind’s eye, with people varying in the amount of detail and vividness of their mental images ¹. Yet if the imagination is like a painter who paints before the mind’s eye, it seems to be a painter who leaves basic things out: By their self-report, people seem to leave out details in mental images that should be readily apparent in real images (Bigelow et al., 2023).

Why do some basic properties seem to be undefined or nonexistent in imagined scenes? It is possible that imagined scenes are actually rich and detailed when created, but that basic properties are then neglected in the same way that perceptual properties are overlooked in real images, due to lapses in attention or memory. But it is possible that these properties and aspects were never there to begin with (see Figure 1).

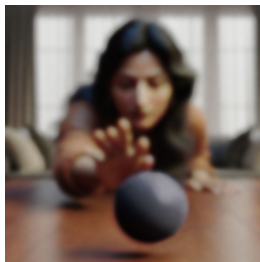
The question of whether some basic properties are truly missing in mental scenes and only filled in on demand, or actually fully imagined to begin with and only later forgotten, is a fundamental matter for mental imagery. In this paper, we propose that the absence of basic details in mental imagery is driven by true non-commitment in a hierarchical scene construction mechanism, rather than being an artifact of perceptual noise or inattention. We empirically test this idea by comparing the details that people omit in the imagination, to the details that people overlook in the brief perception of real scenes, and also by examining the temporal unfolding of scene properties during

¹Cf. Plato: ‘in our souls (there is) a block of wax, in one case larger, in another smaller, in one case the wax is purer, in another more impure and harder, in some cases softer.’

Perception



Inattention, fading,
forgetting, etc.



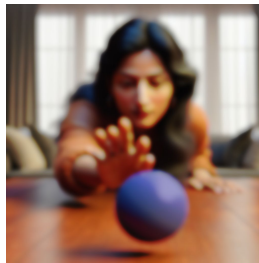
“Did you notice the
color of the ball?”



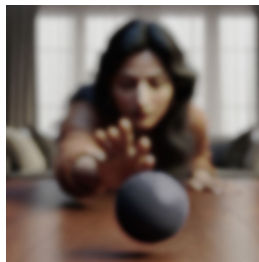
“No”

Imagination

Hypothesis 1:
Detailed Rendering



Inattention, fading,
forgetting, etc.

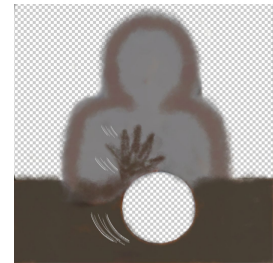


“Did you imagine the
color of the ball?”

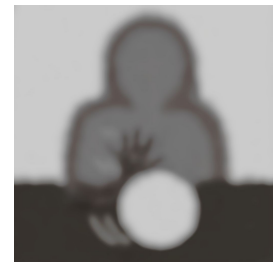


“No”

Hypothesis 2:
Partial Construction



Inattention, fading,
forgetting, etc.



“Did you imagine the
color of the ball?”



“No”

Figure 1: **What drives lack of commitment to some properties in the imagination?** In perception (Left), various details of a scene may be missed due to a variety of processes, such as failures of encoding or retrieval. In the imagination (Right), it may be that people render fully-detailed scenes, which are then subject to the same processes that lead to a lack of detail in perception (**Hypothesis 1**). But, it may be that – in addition to perceptual-like processes – people simply do not generate certain details of the scene to begin with (**Hypothesis 2**).

imagination.

Before detailing our studies and findings, we situate them more broadly in the context of mental imagery, the imagery debate, perceptual inattention, and the recent empirical work on non-commitment in the imagination.

Mental imagery – our ability to “see” with the mind’s eye in the absence of direct perceptual input – has been a topic of much debate in philosophy, psychology, and neuroscience. At the risk of

oversimplification, within this broad debate there are several main camps. Of these, one camp holds a depictive view: mental images have picture-like properties, recruiting analog representations that share an architecture and format with sensory experience, and enabling quasi-perceptual operations (e.g. Kosslyn, 1996, 2006; Pearson, 2019). A different camp holds a propositional view: mental imagery is better understood as existing in a language-like format, lacking any truly functional pictorial characteristics (e.g., Dennett, 2002; Pylyshyn, 2002). There are other camps besides these two. Various views argue for different mixtures of the depictive/propositional formats, or reject the dichotomy as not empirically or theoretically useful (e.g., see for example the replies in Pylyshyn, 2002). Different camps have marshaled diverse sources of evidence, including behavioral tasks, subjective reports, theoretical and computational frameworks, and neuroimaging data, leading to a wide range of models of imagery.

We do not propose to definitively adjudicate between the different camps of the imagery debate. Our focus is instead on the fact that within the general context of this argument, researchers in cognitive science and philosophy have observed the phenomena of “non-commitment” in the imagination (though not necessarily under that specific term). Philosophers have noted that it is of interest that people can imagine scenes without committing to their basic details, like imagining a striped tiger without knowing how many stripes it has, or a purple cow without knowing what shade of purple it is (Shorter, 1952; Dennett, 1986; Dennett and Dennett, 1993). In cognitive science, such examples were marshaled to argue images are not picture-like (Pylyshyn, 1978, 2002). After all, the argument goes, if imagining a striped tiger is like seeing a striped tiger, one should be able to count the stripes.

Defenders of the pictorial view have countered that non-commitment does not rule out a depictive format Kosslyn 2006. The counter-argument points to the fact that visual perception is also incomplete and inexact, that memory is faulty, and that attention is limited. All of these, working together or separately, can lead us to overlook basic details in a scene (Figure 1, left). Failing to specify certain features in an imagined scene simply parallels the selective nature of seeing or remembering. Just as it might be difficult to count the real dots on a complex textured ceiling, so too it may be challenging for mental imagery to retain or process details of a fully rendered image. To be more exact: decades-long research in vision highlights how people routinely fail to register or recall details from scenes. Classic work on inattention blindness Mack and Rock (1998); Mack

(2003) demonstrates that highly salient events can go unnoticed if attention is directed elsewhere; Change blindness studies reveal that observers can miss large alterations in a visual display across brief interruptions (Rensink et al., 1997; Simons, 2000); Attentional blink paradigms show that when two targets appear in quick succession, the second is often not perceived at all (Raymond et al., 1992). These phenomena underscore that perception is far from a full comprehension of the environment: Noise, limited attentional capacity, and rapid decay of visual information all conspire to create a selectively sparse representation of what is “seen” (Most et al., 2001). Even under ideal viewing conditions, people regularly miss details of a scene. So, it may be the case that people create rich, detailed images that share their format, architecture, and level of detail with perception, and then simply fail to attend to the details of those images (Figure 1, middle).

Does neuroscience help to settle the issue of non-commitment? While neuroscience has contributed greatly to research into mental imagery, current findings seem to accommodate both views. In cognitive neuroscience, a long-established view is that mental imagery operates as a kind of “perception in reverse”: it recruits neural mechanisms and cortical pathways associated with visual perception, but driven by top-down signals rather than immediate sensory input (Kosslyn et al., 2001; Pearson and Westbrook, 2015). Early visual areas (or other sensory areas) become activated during imagery, yet the impetus for this activation comes from higher-level brain regions generating an internal model of what would otherwise be incoming stimuli (Albers et al., 2013; Ishai et al., 2000). In this sense, imagery travels backward along the visual processing hierarchy, effectively re-creating a semblance of perceptual experience. Importantly, this backward flow is subject to noise, cognitive constraints, and individual differences in vividness. But, it remains unclear how one should view non-commitment in this framework. It is possible that non-commitment reflects a lack of activation of relevant details along the top-down pathway (“true” non-commitment, in that one does not know the color of the ball in an imagined scene because this detail was never recruited to begin with, as in Figure 1 right). Alternatively, perhaps because imagery shares considerable overlap with actual perception circuits, it is inherently shaped by attentional priorities, expectations, and stored knowledge – leading to similar attentional blind spots or noise limitations as perception (Figure 1, middle). That is, it is possible a scene *is* fully rendered in the early sensory areas, details are omitted just like they are for perception (e.g., while the imagined scene included a red ball, this detail was not attended to, or not retained).

Despite its importance for how the imagination may be organized and its role in the imagery debate, non-commitment in imagery was only recently studied empirically and at scale (Bigelow et al., 2023). This work found that participants spontaneously fail to commit to many simple properties of various imagined scenes, properties that should be readily in real perception (e.g., color, clothing, shapes). This non-commitment was only weakly correlated (if at all) with the vividness of a participant’s imagination. Non-commitment was also not uniform – some properties were consistently more likely across participants and scenes to be “non-committed”. Also, speaking to the difficulty of using self-report measures, participants seemed to readily confabulate post-hoc details about the scene in a way that could lead to the impression that they imagined rich, detailed mental scenes to begin with. While an important step towards empirically addressing non-commitment in the imagination, that work still leaves open several key questions. Specifically, it remains an open question whether seeming non-commitment is different from perceptual inattentiveness and failures of retrieval, or whether it arises from a actual lack of instantiating certain details in the imagination to begin with. The previous work also does not directly address why some properties are more likely to be committed to than others, or how this process proceeds hierarchically, and over time.

Our present work tackles several of the major open questions about non-commitment head-on. We hypothesize that non-commitment in the imagination goes beyond perceptual lapses. We conducted a series of behavioral studies designed to compare the details people omit in purely imagined scenes, with the details that they fail to notice in brief perception conditions.

To foreshadow the results of our first two experiments, the data supports the view that non-commitment is driven by non-perceptual processes. Non-commitment in mental imagery dissociates from perceptual attention failures: the properties omitted when imagining a scene do not systematically match those overlooked when participants briefly view the similar scenes in reality.

This leads to the second open question: What distinct constructive process is involved in non-commitment, and why are some properties more likely to be committed to than others? To help answer this question, we take inspiration from computer graphics: When engineers construct computer-based scenes for the purposes of real-time simulation, they often go through a hierarchical process (see e.g. Gregory, 2018). Rather than painting pixels one by one, a scene is first constructed by importing necessary object assets, and creating a spatial skeleton of the scene lay-

out. Then, more refined shapes and details are filled in. Finally, when the scene is ready for production, a separate computational operation graphically *renders* the scene, filling in a pixel-based representation of what the scene looks like to a camera placed in a particular position, under specific lighting conditions. Indeed, the software that underlies real-time simulation in games and animation has been proposed as broadly equivalent to the way that human mental physical reasoning works (e.g. Hamrick et al., 2011; Battaglia et al., 2013; Ullman et al., 2017; Lake et al., 2017), using similar constraints and workarounds (Li et al., 2023b; Balaban and Ullman, 2025a; Balaban et al., 2024; Bass et al., 2021; Balaban and Ullman, 2025b).

We propose that non-commitment stems from a hierarchical scene construction mechanism that preferentially represents essential entities, spatial layouts, and basic physical relations, before proceeding to finer-grained surface details. This mechanism aligns with proposals in visual cognition that suggest humans build up scene representations in layers, guided by top-down goals and resource constraints (e.g., Mechelli et al., 2004; Stokes et al., 2009; Dijkstra et al., 2019; Griffiths et al., 2015), and is in line with previous work in mental imagery research, including the broad proposal that sees imagery as “perception in reverse” (Kosslyn and Shwartz, 1977; Kosslyn, 1980; Pearson and Westbrook, 2015, e.g.). Neuroscience evidence showed early visual cortex activation during imagination (Kosslyn and Thompson, 2003; Albers et al., 2013; Bergmann et al., 2016; Pearson, 2019; Senden et al., 2019). Additionally, the hippocampus was suggested as the locus of the “director” underlying scene construction (Hassabis and Maguire, 2009). We hypothesize that as scene construction proceeds from the top of the hierarchy, certain properties and variables can be left “open”, undefined or unevaluated, and do not reach rendering.

Our last experiment examined this proposal of hierarchical scene construction by specifically focusing on the temporal progression of properties. We asked participants to imagine abstract scenes, and interrupted their imagination at various time points, before they were asked to report the order of which different scene properties emerged. The idea was to stop people at different points in their scene construction process, which would lead to varying levels of commitment on specific properties. Although we did not find an effect of time, our data showed a robust pattern of a hierarchical progression: participants consistently commit to spatial and identity attributes first, with finer details often left unfilled. This is in line with the general idea that mental imagery involves hierarchical structuring, but further enriches it by showing how the hierarchy might

progress, omit, or skip properties. By positing that features like color or texture might remain uncommitted unless specifically probed or needed, we offer a principled account of why people systematically omit certain details in mental imagery.

Experiment 1: Perception vs. Imagination

We examined whether the report of unimagined properties is due to detailed scene-construction followed by perceptual-like inattentiveness (Figure 1, Hypothesis 1), or due to certain properties not being generated at all (Figure 1, Hypothesis 2). Hypothesis 1 predicts that there should be a tight correlation between the properties that are not noticed or forgotten in perception, and those that are reported as not imagined. Hypothesis 2 does not make such a prediction. Given this, we compared people’s reports of various properties for various scenes, in both direct perception, and in the imagination.

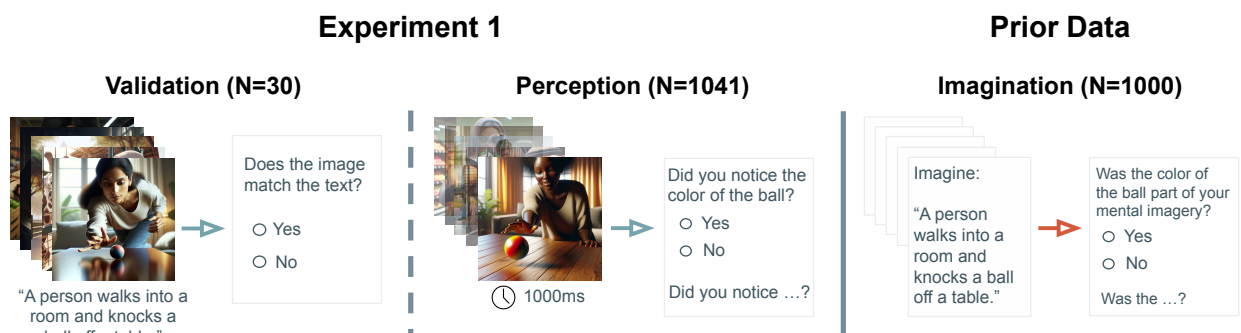


Figure 2: Experiment 1, Procedure: Comparing brief visual perception with imagination. In an initial validation study, participants rated whether each generated image accurately matches a scene description. The images with the highest scores were used in subsequent parts of the study. The “Perception” part asked a different group of participants to observe a (validated) image for 1 second, before answering whether they noticed certain properties. The “Imagination” part of the study relies on data taken from prior work (Bigelow et al., 2023), resulting from participants being asked to imagine a scene as vividly as possible, then reporting on the same properties as in the “Perception” part.

Method

Stimuli

We used 5 different scenes previously used to study non-commitment in the imagination (Bigelow et al., 2023), corresponding to the following text descriptions:

1. A person walks into a room and knocks a ball off a table.
2. A person walks down a supermarket aisle and puts a piece of fruit in their bag.
3. A person walks up to a zoo cage and offers an animal a treat.
4. A person walks into a room and lights a candle in a candle holder.
5. A person stops at a bench and takes off their shoe.

For each scene, we created AI-generated images using DALL-E 3 (Betker et al., 2023)² and the text description for each scene in the prior study of non-commitment (Bigelow et al., 2023). A total of 75 candidate images were created (15 images for each scene). We validated with participants whether the image matches the text, and kept the ones that were highly agreed upon (see the 25 selected images in Supplementary Information). Then, independent groups of participants either viewed the validated image and reported on its properties (Perception), or imagined the scene and reported on its properties (Imagination).

Procedure

In the preliminary **validation** part of the study (Figure 2, left), participants were shown 75 trials of image-text pair (i.e., the AI-generated image paired with its associated scene prompt). For each pair, participants were asked whether the image accurately matches the text, and could provide an optional comment. Participants had 2 practice trials during the instruction phase. The trial order was randomized across participants. For every scene, the top 5 most accurate images from this validation were used in the following perception experiment (i.e. 25 validated images in total).

In the main **perception** part of the study, new participants were recruited, and each participant was randomly assigned to one of the 5 scenes (Figure 2, middle). Participants were told that

²This model may continue to evolve while maintaining the same title, we used the January 2024 version.

their task was to pay close attention and observe an image for only 1 second, and then report honestly about the details of the image. Participants were instructed that there were no right or wrong answers to the questions, and went through 2 practice trials. In the main study, participants were briefly shown a random image selected from the 5 validated images of the scene for 1000ms. Immediately afterwards, they were automatically directed to the reporting page, where participants were asked whether or not they noticed a set of properties on the image (presented in random order). The properties mirrored those used in prior work on imagination (Bigelow et al., 2023).

At the end of the experiment, participants completed check questions and an optional demographics questionnaire asking for their age and gender. We excluded data from participants who failed to answer any of the comprehension check questions correctly (“Please briefly describe your task in this study.”, “Please write down one of the questions you were asked during the main study.”, “What is the color of the sky?”, “What is 3+2?”).

The **imagination** part of the study was from prior work (Bigelow et al., 2023). Participants were asked to read a text prompt that described a given scene, and then imagine the scene as vividly as they could (with unlimited time). Then, they were asked to report whether or not a set of properties was part of their mental image.

Participants

All participants were U.S.-based desktop users on Prolific, and had at least an approval rate of 95% and 100 prior submissions.

We recruited 30 participants for the Validation experiment. The main completion time was 25 minutes, and the median was 17 minutes. Participants were reimbursed for \$5 (about \$12/hour) for their time. Participants’ mean age was 39.4, and the median was 37.0. 17 participants identified as female, and 10 as male, and 1 preferred to not say. All participants passed the attention and comprehension checks.

We recruited 1041 participants (~ 200 per scene) for the main Perception experiment, following the sample size from prior work (Imagination). The mean completion time was 4.61 minutes, median was 3.65 minutes. Participants were reimbursed for \$1 upon completion (about \$15/hour). Participants’ mean age was 39.6, the median was 37. 543 participants identified as female, 467 as male, 22 as non-binary/third gender, 3 preferred not to say, and 6 participants did not respond.

After applying the exclusion criteria, data from 1020 participants remained for analysis.

Results

We compared the data we collected from the Perception part of the study to existing Imagination data from prior work (Bigelow et al., 2023). If participants generate full mental images during imagination but fail to report all details, there should be a strong correlation between properties not committed to in imagination and properties not noticed during brief perception. We computed the percentage of participants who reported “No, I did not notice that” for each property, and correlated with the percentage of participants who reported “No, it was not part of my mental image” in prior work’s imagination experiment.

As shown in Figure 3, aggregating across scenes, there was no significant correlation between the reporting of properties in Perception and in Imagination ($p = 0.17$, $R^2 = 0.043$). When examining the scenes individually, four of the five scenes showed no correlation, and one showed only a modest correlation ($p = 0.038$, $R = 0.48$).

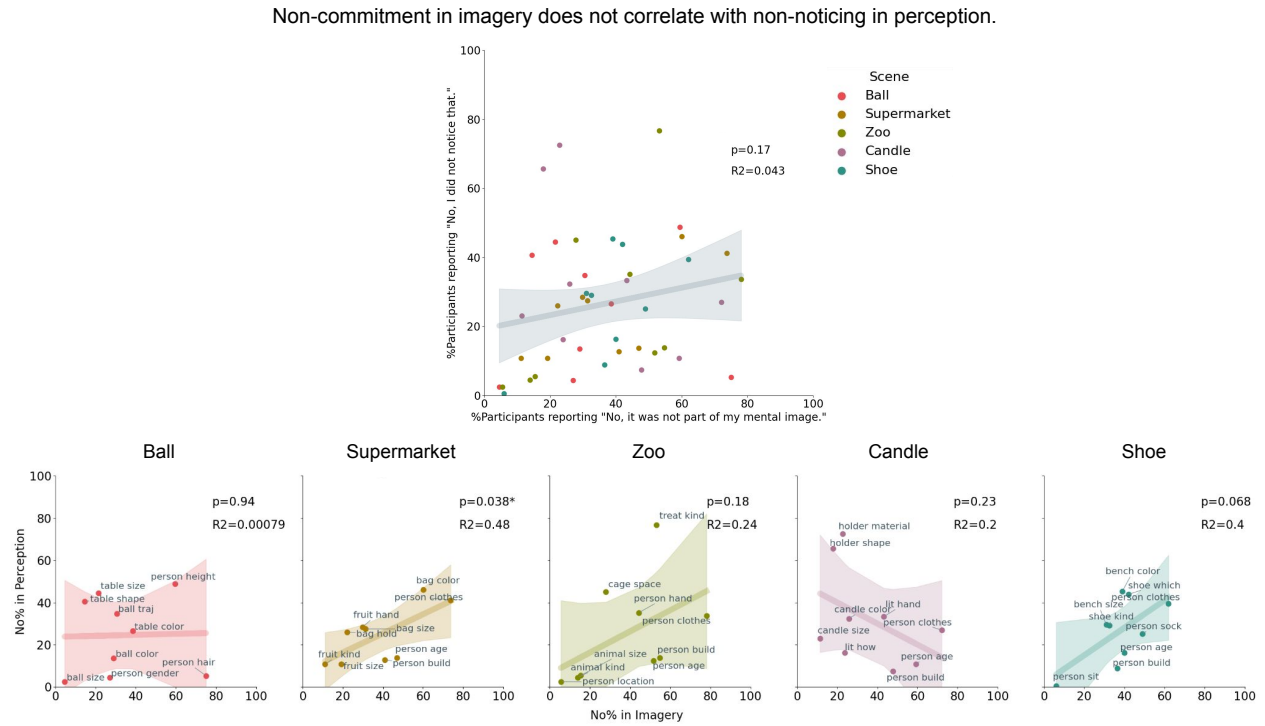


Figure 3: **Experiment 1, Results:** The proportion of participants reporting not noticing a given property in brief perception (y-axis) versus the proportion of participants reporting non-commitment to the same property in imagination (x-axis).

The weak-to-no-correlation results speak against Hypothesis 1, indicating that the properties people do not commit to in the imagination are not the same ones they fail to notice in brief perception. This strengthens the suggestion that non-commitment is not driven by perceptual-like rendering followed by perceptual-like encoding, retention, or retrieval failures.

While the results generally support Hypothesis 2 (Figure 1), the experimental design leaves open another possibility. In the *Imagination* condition, people were given a text prompt that described a given scene. In the *Perception* condition, people were shown an image without the associated text. While the images were validated independently as corresponding to the text, it is possible that the use of certain language (e.g., “ball”, “person”) primed people to pay attention to some details or objects and not others. In other words, it is theoretically possible that in brief visual perception people may freely attend to items that grab attention in a way that differs from where there attention would focus when given a text-based description of the scene. The next experiment tested this confound by providing a textual prompt of the scene in the Perception condition before participants viewed the image.

Experiment 2: Controlling for Prompt-Based Attention

To address the possibility that the text-based prompt used in Imagination condition focused their attention in a different way compared to the more free-form viewing of the Perception condition in Experiment 1, we ran another study. Experiment 2 repeated the set-up of Experiment 1, but used a text-based prompt prior to showing the image stimuli in the Perception condition. Beyond addressing the outstanding concern about text-driven attention priming, this experiment also serves broadly as a replication of Experiment 1.

Method

Stimuli

The same 25 validated images as in Experiment 1 were used in this experiment. Every image was paired with a text prompt depicting the scene. We adopted the exact text prompts from prior work.



Figure 4: **Experiment 2, Procedure: Controlling for potential attention-based priming effects due to text.** Experiment 2 repeated the procedure of Experiment 1, but with additional text displayed before the presentation of images in the “Perception” part. The text corresponded to the same text given to people in the “Imagination” part of the study.

Procedure

Participants went through the same instructions and practice as in Experiment 1. In addition, they were informed about the content of the image prior to observing the image (e.g., “On the next page, you will see an image of ‘A person walks into a room and knocks a ball off a table’”), displayed in highlighted text. Participants again had 1000ms to view the image, then were automatically directed to the reporting page, where they were asked the same list of properties about the image. The rest of the procedures was identical to Experiment 1.

Participants

To match the sample size from the previous experiment, we recruited 1004 participants on Prolific. The mean completion time was 5.08 minutes, and the median was 4 minutes. Participants were reimbursed with \$1 upon completing the experiment (around \$15/hour). The mean age was 40.0, the median was 37; 489 participants identified as female, 489 as male, 20 as non-binary/third gender, and 2 preferred not to say. After applying the exclusion criteria, data from 994 participants remained for analysis.

Results

As in Experiment 1, we found no significant correlation between non-commitment patterns in Imagination and unnoticed details in Perception ($p = 0.17$, $R^2 = 0.044$, Figure 5). Breaking

Non-commitment in imagery does not correlate with non-noticing in perception, after accounting for prompt-based attention.

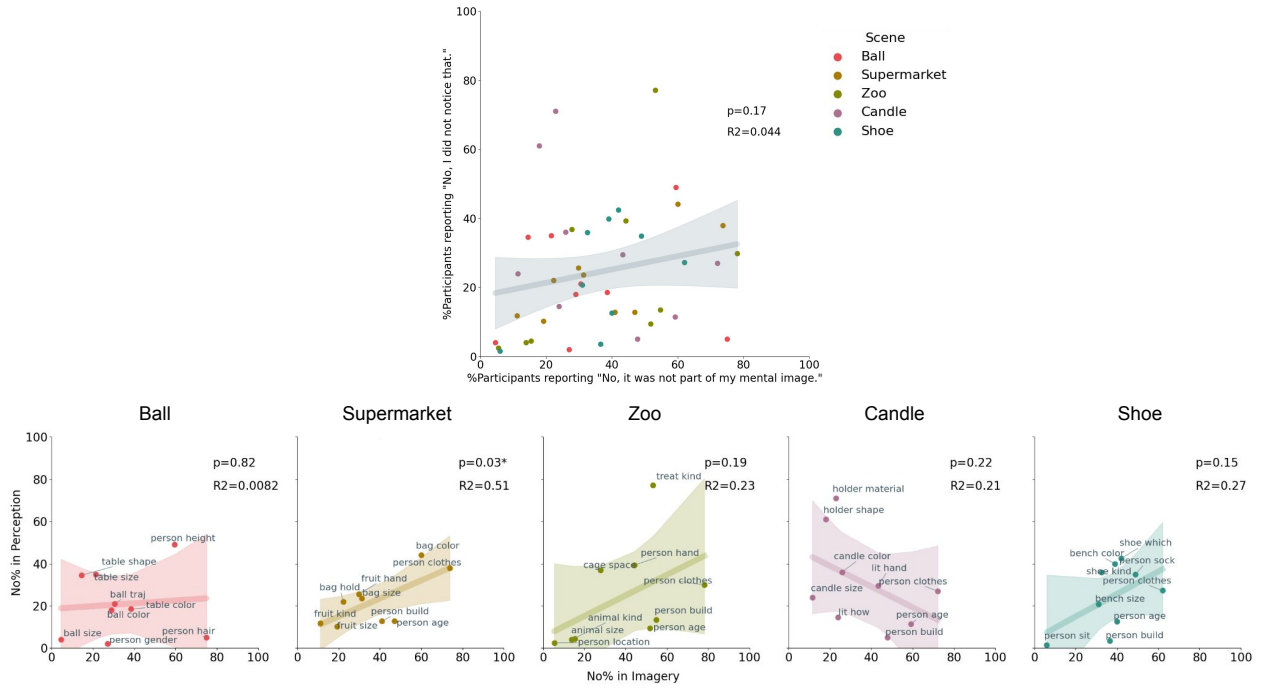


Figure 5: **Experiment 2, Results:** The proportion of participants reporting not noticing a given property in brief perception (y-axis) versus the proportion of participants reporting non-commitment to the same property in imagination (x-axis).

it down by scenes, again, four out of the five scenes showed no correlation, and only one scene showed a moderate correlation ($p = 0.030$, $R = 0.51$).

The weak-to-no-correlation results replicate the overall findings of Experiment 1, and speak against Hypothesis 1 (Figure 1). Attention directed by a text prompt did not appear to be the source of the difference between perception and imagination inattention, which strengthens the idea that failing to report details in imagination is not explained by perceptual or attentional lapses alone. The results suggest that a lack of detail in the imagination is driven by people not encoding certain properties in the first place during scene construction. This raises the question of how this process proceeds, and why some properties are more likely to be committed to compared to others. We turn to this in our next experiment.

Experiment 3: Hierarchical Scene Construction in Imagination

Experiments 1 and 2 suggest that the lack of commitment in the imagination is not due simply to perceptual-based processes. This is in line with the proposal in (Bigelow et al., 2023), that people report some properties as not being part of their imagined scenes because they were not thought of to begin with. This leaves open the question of why some properties are more likely to be committed to. We build on the proposal that mental imagery uses a hierarchical scene construction process, similar to how game engines create a scene by progressing from abstract entity and spatial physical skeleton, to finer-grained rendering details. We examined this hypothesis through the lens of non-commitment and the temporal progression of properties in the imagination process.

In this experiment, we asked participants to imagine an abstract scene for varying amounts of time, and report whether certain properties were in their imagery, and if so, how early they showed up in their imagination. We expect that the ordering that people report follows the hierarchical scene construction process: Identity and spatial physical properties appear before color and material details. Additionally, we expected to be able to stop the scene construction process at varying points, and observe varying levels of commitment to the properties under different times allotted for the imagination (e.g., surface features might show an increasing level of commitment when more time was given to construct and render the scene).

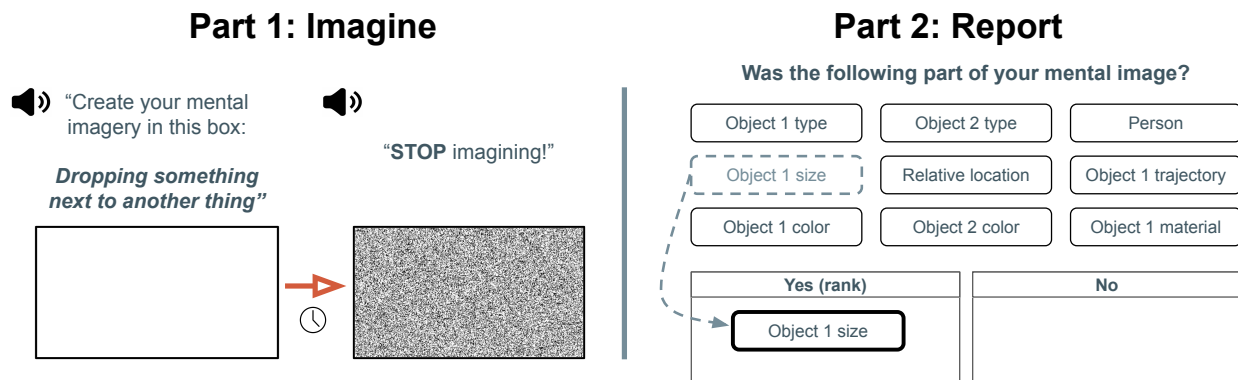


Figure 6: **Experiment 3, Procedure:** Participants were asked to imagine an abstract scene for a short time before they were interrupted. Next, participants were asked to choose and rank the properties of their mental image.

Method

Stimuli

Departing from the previous studies, and motivated by the action recognition literature in AI (Goyal et al., 2017; Mahdisoltani et al., 2018), we used more abstract descriptions as our scene prompt. The idea was to leave open many details as unspecified, and let mental scene construction naturally fill them in. The scene prompts we used include:

1. Dropping something next to another thing.
2. Pushing something with another thing.
3. Pulling something out of another thing.

For every scene, we asked 9 properties which are organized into categories: (1) Identity (i.e. what type of object is it); (2) Spatial/Physical (e.g., relative location, size of objects, trajectory or movement); (3) Surface features (e.g., color, material of object); and (4) Person/Table/Floor (i.e. exploratory features of whether a non-essential entity exists in the scene). Participants did not see which category each property belong to. The properties are randomly shuffled for every participant.

Procedure

After the consent, participants viewed the instructions and practiced with the drag-and-rank response interface. In the main study (Figure 6), participants listened to an audio clip describing one of the abstract scenes and were given one of the four time intervals to imagine the scene: 250ms, 500ms, 1000ms, or 2000ms. The audio asked participants to imagine the scene as vividly as they can and create their mental image in the rectangular box on the screen. After the time elapsed, an interruption (visual scramble noise pattern in the rectangular box and the audio “stop imagining”) redirected them to the reporting page, where they had to decide whether or not each property was part of their mental image. For the properties that were part of their mental image, participants were asked to drag and rank them according to how early the properties showed up in their imagination. Participants who did not pass the practice or failed to answer any of the comprehension checks reasonably (“Please describe your task in this study.”, “Please write down one of the properties you were asked to drag into the boxes.”) were excluded from the analysis. At the end of

the experiment, participants were asked to complete an optional demographics questionnaire about their age and gender in free-form response boxes.

Participants

We recruited 1850 participants from Prolific (about 600 participants each scene and 150 in each timing condition). All participants were U.S.-based desktop users, and had a 95%+ approval rate and 100+ prior submissions. The mean completion time was 6.81 minutes, median was 5.53 minutes. Participants were reimbursed for \$1 upon completion (about \$10.37/hour). For those who responded to the demographics questionnaire, participants' mean age was 41.5, median age was 39; 947 participants identified as “female” or “woman”, 866 as “male” or “man”, and 32 gave a different response. After applying the exclusion criteria, data from 1,020 participants remained for analysis.

Results

Looking at the percentage of participants who answered “Yes, it was part of my mental imagery”, Surface properties showed an overall low level of commitment (Figure 7 top), however, the effect of time was minor. Generalized linear model showed a strong main effect of the Surface category ($p < 0.01$) corresponding to lower probability of commitment, and only a minor interaction effect between Surface category and timing conditions (250ms baseline, 500ms $p = 0.0873$, 1000ms $p = 0.0207$, 2000ms $p = 0.0442$). This may be due to a runaway process for imagination that is difficult to stop. That is, our interruption cues (audio and visual blocking) may be processed bottom-up, whereas imagination may be top-down and thus occupy different cognitive resources. Additionally, participants had to read instructions and consider 9 different properties during the reporting phase, which may lead to memory degradation, continued or reconstructed imagery. Experimental refinements, however, did not produce any clear timing effects (see Supplementary Information). This suggests methodological challenges in halting the ongoing construction process and the noise level in online experiments for time-sensitive designs, as we will discuss in more depth in Discussion.

People consistently committed less to features in the Surface category compared to the Spa-

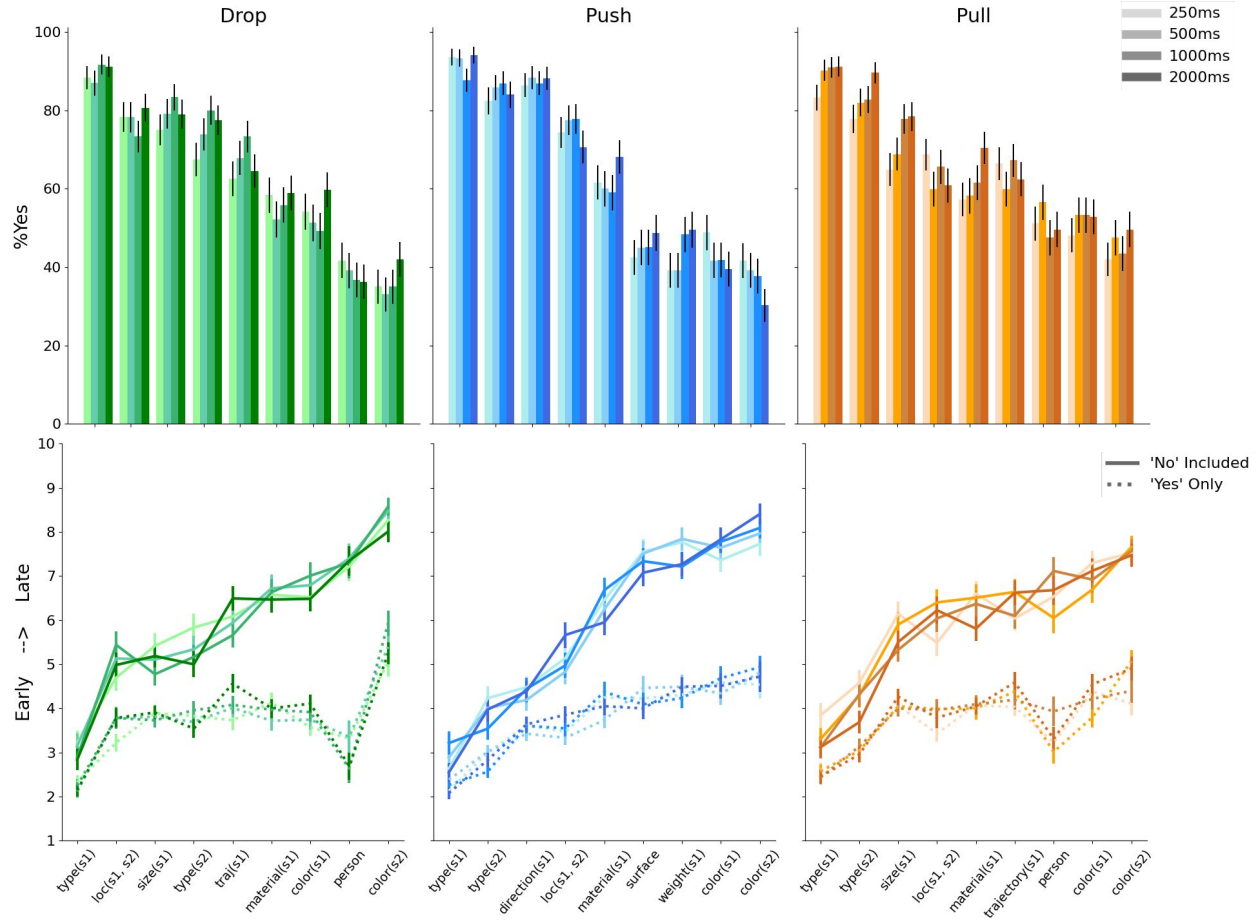


Figure 7: **Experiment 3, Results:** Participants reported whether or not a property was part of their mental image (**top**), and if so, they ranked the properties by how early they showed up in imagination (**bottom**). Properties such as entity type and spatial location are more likely to appear and occur earlier in the construction sequence than surface properties like color or material.

tial/Physical category and the Identity category (Figure 8). Tukey's multiple comparison showed $p < 0.01$ after adjustments for every pairwise comparison among the three categories. For the exploratory category Person/Table/Floor (i.e. non-essential entity existence), the level of non-commitment was similar to the Surface category in 2 of the 3 scenes (Tukey's adjusted $p = 0.7912$ in scene "Push", and $p = 0.8612$ in scene "Pull"), and in scene "Drop", the Person/Table/Floor category showed even lower commitment than the Surface category ($p < 0.01$). Generalized regression model showed a significant main effect of the Person/Table/Floor category on non-commitment compared to the Identity and Spatial/Physical categories ($p < 0.01$). There was no interaction effect between the table/floor entity and the timing conditions, but a significant interaction between person entity and longer times (1000ms $p = 0.0238$, 2000ms $p = 0.0097$) for

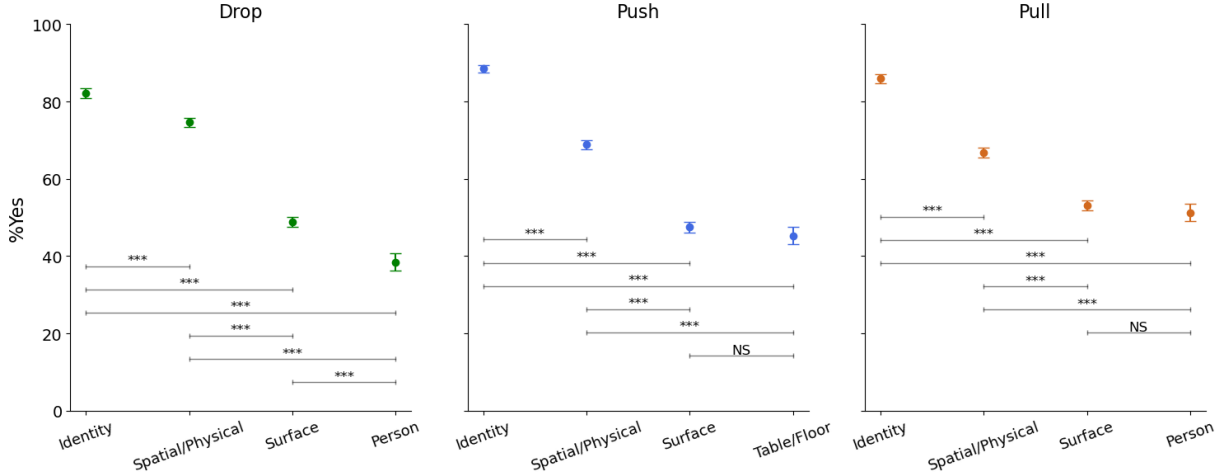


Figure 8: **Experiment 3, Results:** Commitment of properties aggregated for each category.

having greater non-commitment. This suggests that the non-essential entities are not established in the first iteration of the scene construction process, and might be similarly optional as the Surface properties.

For the properties that people committed to, we further calculated the average ranking participants assigned (Figure 7 bottom) and performed ordinal regression to compare among categories. We observed an emerging hierarchical pattern of the properties: Identity properties were ranked as appearing first, followed by Spatial/Physical properties ($p < 0.01$), with Surface details appearing last ($p < 0.01$). Interestingly, for those participants who reported imagining a non-essential entity such as person, table or floor, the ranking they gave for the entity was similar to the rank for the Spatial/Physical properties, though it still occurred later than the Identity properties of the main objects ($p = 0.026$ for person entity, $p < 0.01$ for the table/floor entity). This could mean that most participants did not create non-essential entities at all, but for those who did, they committed to these entities very early on. Future work can test the emergence of these non-essential entities in more detailed ways. In general, the overall pattern of progression we observed supports the hypothesis that mental images develop incrementally, emphasizing identity of core objects and spatial-physical skeleton before finer-grained surface attributes.

Discussion

When we imagine something, it can seem as though a curtain was lifted over a richly decorated scene. But past research (Bigelow et al., 2023) shows the stage of our imagined world might actually be quite sparse. Is this sparseness the result of a sharp decay of the mental image, like a light shining but briefly over a fully furnished stage, that is then lost to memory? Or is the stage bare because no one filled it to begin with?

Our experiments compared imagination to brief visual perception, and also probed the order in which different properties emerge during imagination. Across experiments, people reported the high-level structure of imagined scenes (e.g., what core entities are present, how they relate physically, and where they are) while consistently showing less commitment to surface features (e.g., color, texture, precise shape). The probability of non-commitment in the imagination did not correlate with memory degradation for perceptual scenes. This suggests that non-commitment reflects a top-down, hierarchical scene construction process: The mind first establishes an abstract scaffold of entities and spatial relations, and only later, if at all, fills in fine-grained surface details.

The fact that people showed more non-commitment to surface details, and that this non-commitment did not pattern with perceptual degradation, can be explained by a resource-rational (Griffiths et al., 2015), probabilistic (Chater et al., 2006; Chater and Oaksford, 2008; Griffiths et al., 2010; Bonawitz et al., 2014), and generative modeling framework for mental scene construction (Clark, 2013; Hassabis and Maguire, 2009; Hassabis et al., 2014; Schacter et al., 2007). On this view, mental imagery operates over a probabilistic and structured internal representation, akin to a scene graph or hierarchical generative model, rather than a fully rendered picture. The top of the hierarchy specifies broad elements and relations (e.g., a person standing next to a table), while lower-level nodes encode finer attributes (e.g., the table’s color, the person’s facial expression). Participants can leave selected nodes undefined or represented as distributions over possible values, until constraints or questions require resolution (i.e., the sampling of a specific value for the variable). Regardless of the specific framework one adheres to regarding the imagination, leaving properties of mental images unspecified makes sense from a computational perspective, as it offers both speed and flexibility (Petrov and Du, 2024; Breedlove et al., 2020; Sundh et al., 2023), and has parallels to “lazy evaluation” in computer science (Henderson and Morris Jr, 1976) that

operates in similar ways for similar reasons.

The proposed scene-construction hierarchy in which spatial/physical layouts and relations are prioritized over deferred surface properties also aligns with how current engineered systems create simulations of the environment. Nearly all of the computer programs that power video games and animations separate “simulation/physics” (tracking objects, positions and physics) from “rendering/graphics” (textures, colors, lighting) for computational efficiency. These “game engine” programs can simulate a ball rolling across a table and predict how it will drop off the edge or explain where the person is relative to the ball, without picturing the shading or texture or color of the scene, or indeed how the scene “looks” from a certain perspective. The overall separation between “simulation/physics” and “rendering/graphics” in mental simulation is an ongoing topic of current research (Ullman et al., 2017; Balaban and Ullman, 2025b). Within this overall framework of mental simulation, non-commitment joins other forms of principled approximations in simulation that speed up or save computational resources, including partial simulation (Bass et al., 2021), coarse object representation (Li et al., 2023b), finite simulation budgets (Wang and Ullman, 2025), and a dynamic trade-off between more cumbersome simulation and abstractions or rules (Sosa et al., 2025; Smith et al., 2023).

Converging neuroscience evidence supports a top-down hippocampal-cortical hierarchy for scene construction, with abstract layout and relations for “simulation” assembled first, and fine sensory “rendering” recruited only as needed. Mental scene construction recruits overlapping brain regions for episodic memory, such as the (anterior) hippocampus, retrosplenial cortex, and the frontoparietal network, which coordinate the instantiation of spatial layouts and entities (Hassabis and Maguire, 2007; Hassabis et al., 2007; Buckner and Carroll, 2007; Addis et al., 2007; Klein et al., 2002; Angeli et al., 2025; Zeidman and Maguire, 2016; Zeidman et al., 2015; Milton et al., 2021; Liu et al., 2023), upon the ventromedial prefrontal cortex seeding schematic/goal signals (e.g., McCormick et al., 2018; Barry et al., 2019; Monk et al., 2021). High-level temporal visual cortex (e.g., left fusiform gyrus, parahippocampal gyrus) carries abstract visual semantics and is driven by top-down inputs during imagery (Spagna et al., 2021; Hajhajate et al., 2022; Liu et al., 2023; Thorudottir et al., 2020; Carricarte et al., 2024; Boccia et al., 2021; Pearson et al., 2015). Although some imagery content is decodable in visual cortex (e.g., Pearson and Kosslyn, 2015; Albers et al., 2013; Stokes et al., 2009; Harrison and Tong, 2009; Vetter et al., 2014; Lee et al.,

2012), compared with perception, these imagery representations are more compressed, emphasizing overall spatial and broad categorical information (Roy et al., 2025), and showing biases toward lower spatial frequencies and dilated receptive fields (Breedlove et al., 2020). Importantly, early visual cortex engagement is not obligatory (Kosslyn and Thompson, 2003; Dijkstra, 2024), even in visual imagery (Bartolomeo, 2002; Bartolomeo et al., 2020, 1998; Spagna et al., 2021). This separation between simulation and rendering aligns with our behavioral observations: gist and layout are committed to early, but low-level details are often omitted. Additionally, because imaged scenes must be maintained and manipulated, they occupy visual working memory, which likewise shows a hierarchical and probabilistic structure (Utochkin and Brady, 2020; Brady et al., 2011; Brady and Tenenbaum, 2013), blending summary/ensemble with item-specific details of unequal fidelity. The aforementioned neuroscience findings offer insights into a top-down cascade loop for scene construction: high-level cortical areas in the frontoparietal network initiate the construction (e.g., “a room with a table” at the gist-level), the hippocampus integrates selected elements and their visual semantics in the temporal network into coherence (e.g., the spatial layout of the room, the presence of objects), and only then may finer perceptual details potentially be filled in by early visual cortex to the degree that the context requires. This process may involve multiple cycles of probabilistic iterations, in which the generative program first instantiates an initial scene hierarchy for efficiency and global coherence, later coupled with detail elaboration and consistency checking.

There are at least two caveats to our methodologies. First, there is heterogeneity along multiple dimensions in imagery, and the stimuli in Experiments 1–2 may not match every individual’s idiosyncratic imagery. We tried to account for this by aggregating across participants, validating our stimuli independently, and by using simplified but varied displays that avoid unnecessary distractions or ambiguities. But it remains possible that there were systematic idiosyncratic differences left unaccounted for. Beyond this, during imagination, participants were given as much time as needed. Under a full imagery hypothesis, unlimited imagination time should yield more detailed commitments than a brief glance; instead, we observed the opposite – different commitment patterns between imagination and imagery, and lower commitment to surface details in imagination.

The second caveat concerns the null effects of timed interruptions in Experiment 3. Participants’ scene construction seemed either essentially complete within the time-frames considered, or not at all affected by the interruption timing. This result itself is informative about the dynamics

of scene building. One possible reason is the remarkable speed with which a coarse scene scaffold can be set up: perception can deliver global-property and basic-level category information within 100ms (Greene and Oliva, 2009), and conceptual information at even briefer exposures (Potter et al., 2014). Neuroimaging finds high-level imagery signals in the brain emerge early within a few hundred milliseconds (Dijkstra et al., 2018; Maier et al., 2020). By analogy, imagination may establish the broad spatial framework on which details are later hung within a short time period. In our case, if a coarse scaffold is in place by 250 ms, our interruptions would be too late to catch incomplete scenes. Another (non mutually exclusive) possibility is on-demand completion. We asked participants to project their mental image onto the rectangular box (Azañón et al., 2025) before they were interrupted. After being interrupted, they can post-hoc “finish” the image upon realizing they needed to report it, masking an earlier incompleteness of the image. The very act of querying a detail can trigger its generation. This is a common issue for behavioral paradigms. Disentangling these possibilities will require measures other than subjective report, such as binocular rivalry paradigm (Pearson, 2014; Pearson et al., 2008), pupil dilation (Laeng and Sulutvedt, 2014; Kay et al., 2022), priming and interference paradigm (Nanay, 2021; Wong et al., 2024; Purkart et al., 2025), time-resolved MEG (Dijkstra et al., 2018), or EEG (Maier et al., 2020).

Looking ahead, we think it would be valuable to explore how individual differences in vividness relates to our simulation-vs-rendering account of scene construction, though we note that the vividness construct (often measured using self-report, such as the VVIQ Marks (1973)) has been criticized as theoretically under-specified (Kind, 2017). Prior work shows that non-commitment correlates little with vividness – high-vividness imagers can still leave basic visual features unspecified, and low-vividness imagers can still know structural layouts of the scene (Bigelow et al., 2023). In our framework, vividness may relate to the *rendering* stage sensory strength, rather than the informational content of the underlying simulation. We predict that structural commitments (spatial layout, physical relations, identity at coarse level) persist even at low vividness, whereas sensory rendering markers scale with vividness, and may lead to stronger effects in binocular rivalry tests (Pearson et al., 2008; Keogh and Pearson, 2018) and greater pupillary responses (Laeng and Sulutvedt, 2014; Kay et al., 2022). Neuroscience and behavioral evidence of aphantasia aligns with our view that aphantasia reflects under-engagement of the rendering pipeline that sits on top of an intact simulation scaffold (Balaban and Ullman, 2025b): imagery attempts show atypical early

visual responses (Meng et al., 2023) and reduced frontoparietal-fusiform functional connectivity (Liu et al., 2025; Milton et al., 2021), yet mental rotation reasoning (Kay et al., 2024), spatial imagery (Keogh and Pearson, 2018; Zeman et al., 2020), standard memory (Milton et al., 2021), and spatial features in involuntary dreams (Zeman et al., 2020) are often preserved. Although aphantasia and hyperphantasia may reflect opposite extremes of the engagement of rendering pipeline, their resource-rational non-commitment remains adaptive in both. Future empirical work that orthogonalizes vividness and task demands – separating questions about structure from surface details and pairing them with physiological and neural readouts – can help uncover this potential mechanistic dissociation.

An additional exciting direction to explore includes the similarities and differences between the hierarchical, partially-defined scene representations of the human imagination, and how current generative AI models produce images or dynamic scenes. State-of-the-art vision generative systems ultimately create fully-rendered pixel images; they do not inherently leave parts of an image unspecified or undefined. Autoregressive models such as the original DALL-E (Ramesh et al., 2021) and diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2020; Dhariwal and Nichol, 2021) likewise commit to a concrete pixel array by the end of sampling, which is not in line with non-commitment. Furthermore, while diffusion models create images through iterative refinement from noise to concrete, this process is not guided by an over-arching object-level scene representation. To “keep the layout but swap one object,” current methods usually require extra conditioning (bounding boxes, edges/poses, depths, masks) rather than manipulating nodes in the hierarchical structure (e.g., Li et al., 2023a; Zhang et al., 2023; Avrahami et al., 2023). That said, current generative models improve when they separate high-level structure from fine detail. For example, latent diffusion (Rombach et al., 2022) first operates in a compressed latent space that preserves semantics (e.g., composition of objects and relations in the scene) before decoding into pixels; emerging hybrid and object-centric approaches introduce slot-based and compositional scene representations that better capture scene structures (Locatello et al., 2020; Niemeyer and Geiger, 2021; Liu et al., 2022; Huang et al., 2023). Linking these approaches to non-commitment in humans could sharpen what internal structure is required for human-like flexible and efficient imagination.

Our results suggest that mental imagery is not a fleeting, fading snapshot, but a fast and frugal

program, that is ready to refine on demand. We related our findings to a general framework in which imagery is resource-rational, probabilistic, and hierarchical, with depictive content embedded in a simulation scaffold that need not be fully specified at every level, or rendered at all.

References

- Addis, D. R., Wong, A. T., and Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45(7):1363–1377.
- Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., and De Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, 23(15):1427–1431.
- Angeli, P. A., DiNicola, L. M., Saadon-Grosman, N., Eldaief, M. C., and Buckner, R. L. (2025). Specialization of the human hippocampal long axis revisited. *Proceedings of the National Academy of Sciences*, 122(3):e2422083122.
- Avrahami, O., Hayes, T., Gafni, O., Gupta, S., Taigman, Y., Parikh, D., Lischinski, D., Fried, O., and Yin, X. (2023). Spatext: Spatio-textual representation for controllable image generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18370–18380.
- Azañón, E., Pounder, Z., Figueroa, A., and Reeder, R. R. (2025). Individual variability in mental imagery vividness does not predict perceptual interference with imagery: A replication study of cui et al.(2007). *Journal of Experimental Psychology: General*.
- Balaban, H., Smith, K. A., Tenenbaum, J. B., and Ullman, T. D. (2024). Electrophysiology reveals that intuitive physics guides visual tracking and working memory. *Open Mind*, 8:1425–1446.
- Balaban, H. and Ullman, T. D. (2025a). The capacity limits of moving objects in the imagination. *Nature Communications*, 16(1):5899.

- Balaban, H. and Ullman, T. D. (2025b). Physics versus graphics as an organizing dichotomy in cognition. *Trends in Cognitive Sciences*.
- Barry, D. N., Barnes, G. R., Clark, I. A., and Maguire, E. A. (2019). The neural dynamics of novel scene imagery. *Journal of Neuroscience*, 39(22):4375–4386.
- Bartolomeo, P. (2002). The relationship between visual perception and visual mental imagery: A reappraisal of the neuropsychological evidence. *Cortex*, 38(3):357–378.
- Bartolomeo, P., Bachoud-Lévi, A.-C., De Gelder, B., Denes, G., Dalla Barba, G., Brugières, P., and Degos, J.-D. (1998). Multiple-domain dissociation between impaired visual perception and preserved mental imagery in a patient with bilateral extrastriate lesions. *Neuropsychologia*, 36(3):239–249.
- Bartolomeo, P., Hajhajate, D., Liu, J., and Spagna, A. (2020). Assessing the causal role of early visual areas in visual mental imagery. *Nature Reviews Neuroscience*, 21(9):517–517.
- Bass, I., Smith, K., Bonawitz, E., and Ullman, T. (2021). Partial Mental Simulation Explains Fallacies in Physical Reasoning.
- Battaglia, P. W., Hamrick, J. B., and Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110(45):18327–18332.
- Bergmann, J., Genç, E., Kohler, A., Singer, W., and Pearson, J. (2016). Smaller primary visual cortex is associated with stronger, but less precise mental imagery. *Cerebral cortex*, 26(9):3838–3850.
- Betker, J., Goh, G., Jing, L., Brooks, T., Wang, J., Li, L., Ouyang, L., Zhuang, J., Lee, J., Guo, Y., et al. (2023). Improving image generation with better captions. *Computer Science*. <https://cdn.openai.com/papers/dall-e-3.pdf>, 2(3):8.
- Bigelow, E. J., McCoy, J. P., and Ullman, T. D. (2023). Non-commitment in mental imagery. *Cognition*, 238:105498.

- Boccia, M., Sulpizio, V., Bencivenga, F., Guariglia, C., and Galati, G. (2021). Neural representations underlying mental imagery as unveiled by representation similarity analysis. *Brain Structure and Function*, 226(5):1511–1531.
- Bonawitz, E., Denison, S., Griffiths, T. L., and Gopnik, A. (2014). Probabilistic models, learning algorithms, and response variability: Sampling in cognitive development. *Trends in cognitive sciences*, 18(10):497–500.
- Brady, T. F., Konkle, T., and Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of vision*, 11(5):4–4.
- Brady, T. F. and Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological review*, 120(1):85.
- Breedlove, J. L., St-Yves, G., Olman, C. A., and Naselaris, T. (2020). Generative feedback explains distinct brain activity codes for seen and mental images. *Current Biology*, 30(12):2211–2224.
- Buckner, R. L. and Carroll, D. C. (2007). Self-projection and the brain. *Trends in cognitive sciences*, 11(2):49–57.
- Carricarte, T., Iamshchinina, P., Trampel, R., Chaimow, D., Weiskopf, N., and Cichy, R. M. (2024). Laminar dissociation of feedforward and feedback in high-level ventral visual cortex during imagery and perception. *Isience*, 27(7).
- Chater, N. and Oaksford, M. (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. OUP Oxford.
- Chater, N., Tenenbaum, J. B., and Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in cognitive sciences*, 10(7):287–291.
- Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, 36(3):181–204.
- Dennett, D. C. (1986). *Content and consciousness*. Routledge.

- Dennett, D. C. (2002). *Content and consciousness*. Routledge.
- Dennett, D. C. and Dennett, D. C. (1993). *Consciousness explained*. Penguin uk.
- Dhariwal, P. and Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794.
- Dijkstra, N. (2024). Uncovering the role of the early visual cortex in visual mental imagery. *Vision*, 8(2):29.
- Dijkstra, N., Bosch, S. E., and van Gerven, M. A. (2019). Shared neural mechanisms of visual perception and imagery. *Trends in cognitive sciences*, 23(5):423–434.
- Dijkstra, N., Mostert, P., Lange, F. P. d., Bosch, S., and van Gerven, M. A. (2018). Differential temporal dynamics during visual imagery and perception. *Elife*, 7:e33904.
- Goyal, R., Ebrahimi Kahou, S., Michalski, V., Materzynska, J., Westphal, S., Kim, H., Haenel, V., Fruend, I., Yianilos, P., Mueller-Freitag, M., et al. (2017). The” something something” video database for learning and evaluating visual common sense. In *Proceedings of the IEEE international conference on computer vision*, pages 5842–5850.
- Greene, M. R. and Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological science*, 20(4):464–472.
- Gregory, J. (2018). *Game engine architecture*. AK Peters/CRC Press.
- Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., and Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in cognitive sciences*, 14(8):357–364.
- Griffiths, T. L., Lieder, F., and Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in cognitive science*, 7(2):217–229.
- Hajhajate, D., Kaufmann, B. C., Liu, J., Siuda-Krzywicka, K., and Bartolomeo, P. (2022). The connectional anatomy of visual mental imagery: evidence from a patient with left occipito-temporal damage. *Brain Structure and Function*, 227(9):3075–3083.

- Hamrick, J., Battaglia, P., and Tenenbaum, J. B. (2011). Internal physics models guide probabilistic judgments about object dynamics. In *Proceedings of the 33rd annual conference of the cognitive science society*, volume 2. Cognitive Science Society Austin, TX.
- Harrison, S. A. and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238):632–635.
- Hassabis, D., Kumaran, D., Vann, S. D., and Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, 104(5):1726–1731.
- Hassabis, D. and Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in cognitive sciences*, 11(7):299–306.
- Hassabis, D. and Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521):1263–1271.
- Hassabis, D., Spreng, R. N., Rusu, A. A., Robbins, C. A., Mar, R. A., and Schacter, D. L. (2014). Imagine all the people: how the brain creates and uses personality models to predict behavior. *Cerebral Cortex*, 24(8):1979–1987.
- Henderson, P. and Morris Jr, J. H. (1976). A lazy evaluator. In *Proceedings of the 3rd ACM SIGACT-SIGPLAN Symposium on Principles on Programming Languages*, pages 95–103.
- Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851.
- Huang, L., Chen, D., Liu, Y., Shen, Y., Zhao, D., and Zhou, J. (2023). Composer: Creative and controllable image synthesis with composable conditions. *arXiv preprint arXiv:2302.09778*.
- Ishai, A., Ungerleider, L. G., and Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron*, 28(3):979–990.
- Kay, L., Keogh, R., Andrillon, T., and Pearson, J. (2022). The pupillary light response as a physiological index of aphantasia, sensory and phenomenological imagery strength. *Elife*, 11:e72484.

- Kay, L., Keogh, R., and Pearson, J. (2024). Slower but more accurate mental rotation performance in aphantasia linked to differences in cognitive strategies. *Consciousness and cognition*, 121:103694.
- Keogh, R. and Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, 105:53–60.
- Kind, A. (2017). Imaginative vividness. *Journal of the American Philosophical Association*, 3(1):32–50.
- Klein, S. B., Loftus, J., and Kihlstrom, J. F. (2002). Memory and temporal experience: The effects of episodic memory loss on an amnesic patient’s ability to remember the past and imagine the future. *Social cognition*, 20(5):353–379.
- Kosslyn, S. (2006). The case for mental imagery. *Oxford Psychology Series*, 39.
- Kosslyn, S. M. (1980). *Image and mind*. Harvard University Press.
- Kosslyn, S. M. (1996). *Image and brain: The resolution of the imagery debate*. MIT press.
- Kosslyn, S. M., Ganis, G., and Thompson, W. L. (2001). Neural foundations of imagery. *Nature reviews neuroscience*, 2(9):635–642.
- Kosslyn, S. M. and Schwartz, S. P. (1977). A simulation of visual imagery. *Cognitive science*, 1(3):265–295.
- Kosslyn, S. M. and Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological bulletin*, 129(5):723.
- Laeng, B. and Sulutvedt, U. (2014). The eye pupil adjusts to imaginary light. *Psychological science*, 25(1):188–197.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, 40:e253.
- Lee, S.-H., Kravitz, D. J., and Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *Neuroimage*, 59(4):4064–4073.

- Li, Y., Liu, H., Wu, Q., Mu, F., Yang, J., Gao, J., Li, C., and Lee, Y. J. (2023a). Gligen: Open-set grounded text-to-image generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22511–22521.
- Li, Y., Wang, Y., Boger, T., Smith, K. A., Gershman, S. J., and Ullman, T. D. (2023b). An approximate representation of objects underlies physical reasoning. *Journal of Experimental Psychology: General*.
- Liu, J., Zhan, M., Hajhajate, D., Spagna, A., Dehaene, S., Cohen, L., and Bartolomeo, P. (2023). Ultra-high field fmri of visual mental imagery in typical imagers and aphantasic individuals. *BioRxiv*, page 2023.
- Liu, J., Zhan, M., Hajhajate, D., Spagna, A., Dehaene, S., Cohen, L., and Bartolomeo, P. (2025). Visual mental imagery in typical imagers and in aphantasia: A millimeter-scale 7-t fmri study. *cortex*, 185:113–132.
- Liu, N., Li, S., Du, Y., Torralba, A., and Tenenbaum, J. B. (2022). Compositional visual generation with composable diffusion models. In *European conference on computer vision*, pages 423–439. Springer.
- Locatello, F., Weissenborn, D., Unterthiner, T., Mahendran, A., Heigold, G., Uszkoreit, J., Dosovitskiy, A., and Kipf, T. (2020). Object-centric learning with slot attention. *Advances in neural information processing systems*, 33:11525–11538.
- Mack, A. (2003). Inattentional blindness: Looking without seeing. *Current directions in psychological science*, 12(5):180–184.
- Mack, A. and Rock, I. (1998). Inattentional blindness: Perception. *Visual attention*, (8):55.
- Mahdisoltani, F., Berger, G., Gharbieh, W., Fleet, D., and Memisevic, R. (2018). On the effectiveness of task granularity for transfer learning. *arXiv preprint arXiv:1804.09235*.
- Maier, M., Froemer, R., Rost, J., Sommer, W., and Rahman, R. (2020). Time course and shared neurocognitive mechanisms of mental imagery and visual perception. *bioRxiv*.

- Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British journal of Psychology*, 64(1):17–24.
- McCormick, C., Ciaramelli, E., De Luca, F., and Maguire, E. A. (2018). Comparing and contrasting the cognitive effects of hippocampal and ventromedial prefrontal cortex damage: a review of human lesion studies. *Neuroscience*, 374:295–318.
- Mechelli, A., Price, C. J., Friston, K. J., and Ishai, A. (2004). Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cerebral cortex*, 14(11):1256–1265.
- Meng, M., Chang, S., Zhang, X., and Pearson, J. (2023). Imageless imagery in aphantasia: decoding non-sensory imagery in aphantasia.
- Milton, F., Fulford, J., Dance, C., Gaddum, J., Heuerman-Williamson, B., Jones, K., Knight, K. F., MacKisack, M., Winlove, C., and Zeman, A. (2021). Behavioral and neural signatures of visual imagery vividness extremes: Aphantasia versus hyperphantasia. *Cerebral cortex communications*, 2(2):tgab035.
- Monk, A. M., Dalton, M. A., Barnes, G. R., and Maguire, E. A. (2021). The role of hippocampal–ventromedial prefrontal cortex neural dynamics in building mental representations. *Journal of cognitive neuroscience*, 33(1):89–103.
- Most, S. B., Simons, D. J., Scholl, B. J., Jimenez, R., Clifford, E., and Chabris, C. F. (2001). How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychological science*, 12(1):9–17.
- Nanay, B. (2021). Unconscious mental imagery. *Philosophical Transactions of the Royal Society B*, 376(1817):20190689.
- Niemeyer, M. and Geiger, A. (2021). Giraffe: Representing scenes as compositional generative neural feature fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11453–11464.
- Pearson, J. (2014). New directions in mental-imagery research: the binocular-rivalry technique and decoding fmri patterns. *Current Directions in Psychological Science*, 23(3):178–183.

- Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nature reviews neuroscience*, 20(10):624–634.
- Pearson, J., Clifford, C. W., and Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Current Biology*, 18(13):982–986.
- Pearson, J. and Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the national academy of sciences*, 112(33):10089–10092.
- Pearson, J., Naselaris, T., Holmes, E. A., and Kosslyn, S. M. (2015). Mental imagery: functional mechanisms and clinical applications. *Trends in cognitive sciences*, 19(10):590–602.
- Pearson, J. and Westbrook, F. (2015). Phantom perception: voluntary and involuntary nonretinal vision. *Trends in Cognitive Sciences*, 19(5):278–284.
- Petrov, A. A. and Du, Y. (2024). Are abstract relational roles encoded visually? evidence from priming effects. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 46.
- Potter, M. C., Wyble, B., Haggmann, C. E., and McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, 76(2):270–279.
- Purkart, R., Delem, M., Ranson, V., Andrey, C., Versace, R., Cavalli, E., and Plancher, G. (2025). Are there unconscious visual images in aphantasia? development of an implicit priming paradigm. *Cognition*, 256:106059.
- Pylyshyn, Z. W. (1978). Imagery and artificial intelligence. In Savage, W., editor, *Perception and Cognition*, pages 105–115. University of Minnesota Press.
- Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and brain sciences*, 25(2):157–182.
- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., and Sutskever, I. (2021). Zero-shot text-to-image generation. In *International conference on machine learning*, pages 8821–8831. Pmlr.

- Raymond, J. E., Shapiro, K. L., and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of experimental psychology: Human perception and performance*, 18(3):849.
- Rensink, R. A., O’regan, J. K., and Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological science*, 8(5):368–373.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695.
- Roy, T. S., Breedlove, J., St-Yves, G., Kay, K., and Naselaris, T. (2025). A transformation from vision to imagery in the human brain. *bioRxiv*, pages 2025–09.
- Schacter, D. L., Addis, D. R., and Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature reviews neuroscience*, 8(9):657–661.
- Senden, M., Emmerling, T. C., Van Hoof, R., Frost, M. A., and Goebel, R. (2019). Reconstructing imagined letters from early visual cortex reveals tight topographic correspondence between visual mental imagery and perception. *Brain Structure and Function*, 224:1167–1183.
- Shorter, J. (1952). Imagination. *Mind*, 61(244):528–542.
- Simons, D. J. (2000). Current approaches to change blindness. *Visual cognition*, 7(1-3):1–15.
- Smith, K., Battaglia, P., and Tenenbaum, J. (2023). Integrating heuristic and simulation-based reasoning in intuitive physics.
- Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. pmlr.
- Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2020). Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*.

- Sosa, F. A., Gershman, S. J., and Ullman, T. D. (2025). Blending simulation and abstraction for physical reasoning. *Cognition*, 254:105995.
- Spagna, A., Hajhajate, D., Liu, J., and Bartolomeo, P. (2021). Visual mental imagery engages the left fusiform gyrus, but not the early visual cortex: A meta-analysis of neuroimaging evidence. *Neuroscience & Biobehavioral Reviews*, 122:201–217.
- Stokes, M., Thompson, R., Cusack, R., and Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *Journal of Neuroscience*, 29(5):1565–1572.
- Sundh, J., Sanborn, A. N., Zhu, J.-Q., Spicer, J., León-Villagr , P., and Chater, N. (2023). Approximating bayesian inference through internal sampling.
- Thorudottir, S., Sigurdardottir, H. M., Rice, G. E., Kerry, S. J., Robotham, R. J., Leff, A. P., and Starrfelt, R. (2020). The architect who lost the ability to imagine: The cerebral basis of visual imagery. *Brain Sciences*, 10(2):59.
- Ullman, T. D., Spelke, E., Battaglia, P., and Tenenbaum, J. B. (2017). Mind games: Game engines as an architecture for intuitive physics. *Trends in cognitive sciences*, 21(9):649–665.
- Utochkin, I. S. and Brady, T. F. (2020). Individual representations in visual working memory inherit ensemble properties. *Journal of Experimental Psychology: Human Perception and Performance*, 46(5):458.
- Vetter, P., Smith, F. W., and Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. *Current Biology*, 24(11):1256–1262.
- Wang, Y. and Ullman, T. D. (2025). Resource bounds on mental simulations: Evidence from a liquid-reasoning task. *Journal of Experimental Psychology: General*.
- Wong, C. Y., Cedillo, A. H., and Morsella, E. (2024). The priming of stimulus-elicited involuntary mental imagery. *Acta Psychologica*, 246:104250.
- Zeidman, P. and Maguire, E. A. (2016). Anterior hippocampus: the anatomy of perception, imagination and episodic memory. *Nature Reviews Neuroscience*, 17(3):173–182.

- Zeidman, P., Mullally, S. L., and Maguire, E. A. (2015). Constructing, perceiving, and maintaining scenes: hippocampal activity and connectivity. *Cerebral Cortex*, 25(10):3836–3855.
- Zeman, A., Milton, F., Della Sala, S., Dewar, M., Frayling, T., Gaddum, J., Hattersley, A., Heuerman-Williamson, B., Jones, K., MacKisack, M., et al. (2020). Phantasia—the psychological significance of lifelong visual imagery vividness extremes. *Cortex*, 130:426–440.
- Zhang, L., Rao, A., and Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3836–3847.