Perception as a hierarchical competition: a model that differentiates imagined, veridical, and hallucinated percepts

Alexander A Sulfaro[1], Amanda K Robinson[1,2], Thomas A Carlson[1]

[1]The University of Sydney, New South Wales, Australia
[2]Queensland Brain Institute, The University of Queensland, Queensland, Australia

Abstract

Mental imagery can be understood as the process by which a thought becomes experienced with sensory characteristics. Yet, it's not clear why mental images have diminished sensory features compared to veridical percepts, nor if and how mental images are phenomenologically distinct from hallucinations, another type of non-veridical sensory experience. Current evidence suggests that externally- and internally-generated percepts both depend on similar neural systems. If so, it's plausible that imagined and veridical stimuli should interfere with one another. Here we propose that considering how externally-generated stimuli (i.e. sensory input) and internally-generated stimuli (i.e. thoughts) might compete could be sufficient to differentiate veridical, imagined, and hallucinated percepts. To predict the perceptual consequences of competition, we constructed a computational model of a serially-connected, hierarchical system with bidirectional information flow, emulating the primate visual system. We simulated mental imagery as top-down generative feedback in our model, both with and without competition from bottom-up sensory input. Without competition, imagined stimuli formed the dominant representation at each level of the hierarchy. Yet, when competing sensory input was present, imagined stimuli were outcompeted at lower levels while remaining dominant at higher levels, with an equilibrium point in between. We interpret our findings under a framework whereby low-level, high-dimensional stimulus information (e.g. in the early visual cortex) contributes most to the sensory aspects of perceptual experience, while high-level, low-dimensional stimulus information (e.g. towards temporal regions) contributes most to the abstract, amodal aspects of perceptual experience. Our findings suggest that ongoing bottom-up sensory input during waking life may prevent imagined stimuli from outcompeting and overriding our sensory experience of veridical stimuli. In contrast, dampening or eradicating sensory input could cause internally-generated stimuli to dominate sensory experience. We further discuss how our model can explain individual differences in mental imagery quality, account for phenomena such as daydreaming and hallucinations, and model mental imagery in multiple sensory modalities.

Keywords: mental imagery, visual working memory, inner speech, aphantasia, mind-wandering, hallucinations

1 Introduction

1.1 Mental imagery: a sensory form of thinking

What do we really experience when we think? Sometimes, we may be aware that we are thinking despite no perceivable experience of what we are thinking about. In other cases, we may experience our thoughts as feelings or emotions. Yet, thoughts can also be seen, heard, or in some way perceived with other sensory qualities as well. This latter process, by which thoughts become experienced with sensory qualities, may aptly describe the phenomenon commonly known as *mental imagery*.

The term *mental imagery* is typically associated with the visual modality, where it involves in some way *seeing* what you are thinking about rather than only *knowing* that you are thinking about it. However, the term encapsulates all the senses such that seeming to hear, smell, taste, and even touch what you are thinking about are all forms of mental imagery. Yet, how does the brain create sensory experience from thoughts in any sensory modality? How can we see what we're thinking about when the focus of our thoughts is not present in front of us?

1.2 Modelling mental imagery as a retrohierarchical process

One answer is that mental imagery is a generative process which uses similar neural machinery to that used during veridical perception, yet with an opposite direction of information flow (Figure 1). In veridical visual perception, primitive sensory information is sequentially integrated to form neural representations of more abstract, amodal information. For instance, once visual information enters the visual cortex via the retina and thalamus, the receptive fields of neurons become increasingly large and complex along the ventral stream, moving from a selectivity for specific edge orientations in V1 to curves in V2, shapes in V4, and objects towards the temporal lobe (see Wilson & Wilkinson (2015) for an overview). This process allows knowledge and understanding to be hierarchically extracted from real images in the world. Hence, a visual experience of seeing a red circle on top of a long black line also entails gaining the abstract understanding of knowing we're looking at a traffic light.

Visual mental imagery, in contrast, involves reconstructing a visual experience from knowledge and understanding which we already possess. One way to accomplish mental imagery would therefore be to process information about a stimulus retrohierarchically. In line with this idea, neural activity during the early stages of veridical visual perception resembles neural activity during the later stages of visual mental imagery and vice versa (Breedlove et al., 2020; Dentico et al., 2014; Dijkstra et al., 2020; Linde-Domingo et al., 2019), with both processes most likely using the same neural machinery (Cichy et al., 2012; Dijkstra, Bosch, et al., 2019; Dijkstra et al., 2017, 2018; Ishai, 2010; Kosslyn et al., 1993; Xie et al., 2020). Hence, for visual mental imagery, evidence suggests that we might see what we are thinking about using the same neural systems used to think about what we are seeing during veridical visual perception, yet in reverse. However, veridical perception and imagination are unlikely to be entirely symmetrical given that most individuals reliably report that their mental images are experienced with some form of reduced quality (termed *vividness*)

compared to real images (Marks, 1973). Explaining the origin of this difference is therefore necessary for a complete account of the mechanisms underlying mental imagery and consequently forms the focus of our investigation here.

One explanation holds that mental images may not be experienced with the same quality as real images due to the intrinsic consequences of retrohierarchical information flow in the visual system. Breedlove et al. (2020) suggest that, in the absence of visual input, a generative retrohierarchical process should necessarily result in distortions to imagined stimuli compared to real stimuli. During veridical perception, multiple pieces of sensory information (e.g. multiple edges) are integrated to form a single piece of more abstract information (e.g. a single curve). Storing abstract information therefore requires less dimensionality than sensory information (that is, a smaller number of units, such as neurons, are required to represent information about the feature). However, this entails that detailed visual information is lost with each instance of feedforward pooling. Retrohierarchically-generated mental images could then never be recreated with the same level of detail as real images because they are being reconstructed from high-level source material which does not contain sufficiently-detailed sensory information. High-dimensional mental image reconstructions would therefore be limited by the lower dimensionality of abstract, amodal neural representations in a similar way to the generative images produced by machine-learning autoencoders. Accordingly, Breedlove et al. (2020) showed that neural populations which represent mental images of previously seen stimuli in early visual cortices have lower spatial frequency preferences, more foveal receptive fields, and larger receptive fields than when those same images were represented during veridical perception. During mental imagery, earlier visual regions therefore tended to adopt response properties more typical of neural populations involved in later-stage vision, aligning with the expected outcome of an information-deficient, low-dimensional stimulus representation being retrohierarchically mapped into a high-dimensional space.

1.3 Why don't we hallucinate our mental images?

The approach by Breedlove et al. (2020) effectively explains how a stimulus may be represented with less detail at low levels of the visual hierarchy if it arrived there by a feedback process, which could account for the seemingly diminished quality of mental images. However, it's unclear whether the reportedly low vividness of mental images compared to veridical percepts reflects phenomenological differences which are purely reducible to image statistics (e.g. spatial frequency, resolution, size). If mental images are experienced only as distorted versions of their veridical counterparts, then the experience of a mental image should be possible to emulate purely by perceiving an aptly distorted veridical image. Additionally, if mental imagery is modelled as an isolated process without competition from veridical perception, then imagined percepts should be expected to be the only percepts present during imagery, replacing or superseding veridical images in the visual field with an equivalently visible, albeit distorted, non-veridical image. Collectively, such imagery may be better characterised as a hallucination. A hallucination is another form of endogenously-generated, non-

veridical sensory experience which differs from a veridical percept in that it involves non-veridical content despite retaining the same perceptual quality (i.e. vividness) (Chiu, 1989; Ffytche et al., 1998). If mental images are distinct from hallucinations as well as veridical percepts, then additional neural mechanisms must underly mental imagery apart from those discussed by Breedlove et al. (2020).

What other factors might contribute to the unique phenomenology of mental imagery? A complete explanation may require considering that mental imagery does not occur in a vacuum. Although mental imagery can be modelled as an isolated process, it invariably occurs while we are awake and perceiving the external world. Yet, if mental imagery and veridical perception share neural resources, then both processes should interact, if not compete or interfere with one another. Mental and veridical stimuli can interact constructively and destructively, inducing priming or adaptation to veridical stimuli, depending on the individual (Dijkstra et al., 2021; Dijkstra, Hinne, et al., 2019; Keogh & Pearson, 2017). Increasing environmental luminance can also impede the priming effects of mental imagery (Keogh & Pearson, 2017), while modality-specific distractors reduce the recall of modality-specific information (Vredeveldt et al., 2011; Wais et al., 2010), potentially reflecting competitive interference. A framework is therefore necessary to explain how mental images can be phenomenologically differentiated from hallucinations and veridical percepts while accounting for interactions between internally- and externally-generated stimuli.

We suggest that assessing how effectively internally- and externally-generated stimuli compete for high- and low-level representational resources in the brain could sufficiently account for the reportedly reduced vividness of mental images compared to real images or hallucinations. Consequently, we constructed a computational network with serial, bidirectional information flow to simulate the hierarchical interaction of internally- and externally-generated stimulus representations through feedforward and feedback pathways (Figure 1). We used this model to evaluate the degree to which internally-generated stimulus representations are able to spread throughout a hierarchical system, with and without competition from externally-generated stimuli.

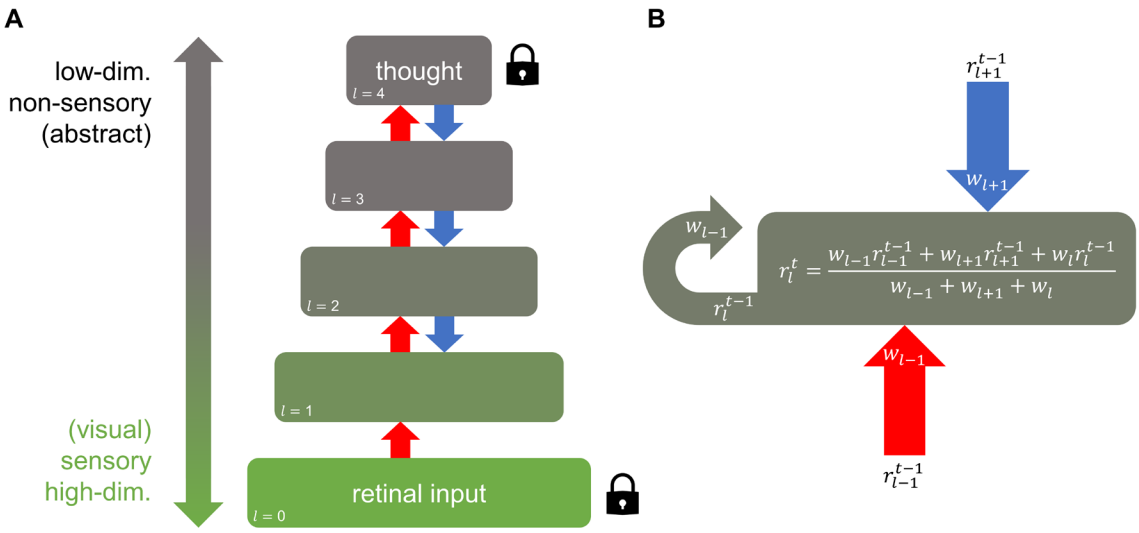
Our findings suggest that the quasi-sensory experience of mental imagery could be explained by imagined stimuli outcompeting veridical stimuli in neural regions which support the more abstract contents of perceptual experience at the same time as veridical stimuli outcompete imagined stimuli at neural regions supporting the more sensory components of perceptual experience. Hence, during visual mental imagery, even though we may be thinking of something that is not there, the most visual aspects of our experience might still be dominated by the images coming in from our eyes. The quality of a mental image may therefore not purely be a product of its image statistics, but also of how well an imagined stimulus can outcompete a veridical stimulus. We therefore put forward hierarchical competition as a major factor differentiating the perceptual experiences of real and imagined content, as well as imagined and hallucinated content. We further discuss how the model presented here can be applied to explaining variation in mental imagery quality, as well as daydreaming, hallucinations, and mental imagery in other sensory modalities (e.g. auditory imagery).

2 Methods

2.1 Computational model

Figure 1

A hierarchical network modelling interference between externally- and internally-generated stimuli



Note. **A.** A hierarchical network model of visual perception. Layer size indicates the relative dimensionality (dim.) of the encoded representation. Information is pooled via feedforward processes (red) and extrapolated via feedback processes (blue). Lock icons indicate input layers with clamped representations during simulations where mental imagery is attempted during veridical perception. Clamped representations do not evolve with time. Note that a "thought" here could be any internally-generated representation. **B.** Evolution rule for unclamped layers. Each value *r* in the representational matrix of each layer *l* evolves at each timepoint *t* according to a weighted average of the adjacent layers and the current layer at the previous timestep with weights *w*. The ascending value array was convolved with an averaging kernel before being weighted to facilitate dimensionality reduction.

Our goal was to model the degree and extent to which internally-generated stimuli are able to spread throughout a hierarchical system, with and without competition from externally-generated stimuli. To do this, a dynamic, serially-connected, five-layer computational network was constructed to approximate a simplified perceptual processing hierarchy. Each layer *l* encoded a representation *r* consisting of a square matrix of values, decreasing linearly in size from 256 x 256 elements at the lowest layer (*l* = *0*) to 208 x 208 elements at the higher layer (*l* = *4*). Each element in the matrix could vary in value along a single dimension, from 0 to 1, encoding the variation of an arbitrary stimulus feature (e.g. brightness in Figure 2). Each layer was therefore able to store information about a stimulus, with the degree of information stored limited by the layer's dimensionality (the number of elements available to represent the

stimulus at that layer). The lowest layer with the highest dimensionality acted as a stimulus input layer for the network, analogous to the retina providing high-dimensional sensory input to the visual system.

The network allowed representations to propagate between layers, and for bottom-up processes to interact with top-down processes. Hence, the representation r of each layer l was updated at each timestep t according to a weighted average of the ascending representation r_{l-1} from the layer below, descending representation r_{l+1} from the layer above, and representation r_l of the current layer, each at the previous timestep $t-1$ (Figure 1B).Each layer was updated sequentially, first as part of a feedforward sweep, then as part of a feedback sweep, and then alternating with every subsequent iteration. Weightings w_{l-1}, w_{l+1}, and w_l for the ascending, descending, and current representations respectively were each equal to one unless otherwise specified. However, if a layer acted as a constant source of information in the network (representing either an internally- or externally-generated stimulus), the values in its representational matrix were locked in place (clamped) by setting the weightings of ascending and/or descending inputs to that layer to zero. Given that the retina is always receiving visual input in an awake state, the lowest layer of the network was always clamped to some predetermined set of input values. Drawing from modelling by Breedlove et al. (2020), the highest layer in the network was also clamped to a fixed set of input values to simulate an internally-generated stimulus (i.e. a thought) during mental imagery.

To simulate feedforward information pooling in the visual ventral stream, ascending representations were combined using a square convolutional kernel such that each element of the layer above received the average of inputs from multiple elements from the layer below. Hence, from $l = 0$ to $l = 4$, the receptive field size of each element in each representation increased while layer dimensionality decreased. Because forming a mental image of an object first requires at some point seeing the object or its constituent parts, internally-generated (e.g. imagined) stimulus inputs were formed from externally-generated (e.g. retinal) stimulus inputs processed through a single complete feedforward sweep from layer 0 to layer 4.

To assess the degree to which mental imagery was able to compete with veridical sensory input, we allowed the network to evolve under three main sets of initial conditions: veridical perception without imagination, imagination without veridical perception, and imagination during veridical perception. In the first case, to simulate a scenario where externally-generated visual input is present without competition, only the bottom layer was clamped to a stimulus representation (veridical perception without imagination). Secondly, to simulate a scenario where an internally-generated stimulus is present without competition, the top layer was clamped to an imagined stimulus representation (imagination without veridical perception). The ascending input from the base layer was down-weighted to zero in this scenario to enforce a state of sensory disconnection. Thirdly, in what we consider a more realistic mental imagery scenario, both the top and bottom layers were simultaneously clamped to different stimulus representations to simulate mental imagery occurring in the presence of competing sensory input (imagination during veridical perception). All intervening or non-input layers were initialised with noise.

In each case, at the steady state, we measured the proportion of the imagined stimulus represented at each layer relative to the veridical stimulus. This was achieved by taking the average steady-state value of all elements in each layer for the case where the initial low-level or high-level input representations were set homogenously to 0 or 1, respectively. To allow for the findings to be visually interpretable, each scenario was replicated using images as inputs such that each element of a representation corresponded to one pixel in the image (Figure 2). A 256 x 256 greyscale image of a camel was used for sensory input, and a 256 x 256 greyscale image of a hat was used for imagined input after conversion to a reduced dimensionality of 208 x 208 via feedforward pooling. Both were sourced from the BOSS image database (Brodeur et al., 2010). We also separately explored how varying the ascending and descending weighting schemes affected competition between simulated imagined and veridical stimulus representations (Figure 3).

2.2 Translating model outcomes to subjective experience

Our model can be interpreted as analogous to a simplified model of the human visual processing hierarchy, with a basal layer for externally-generated retinal input and an apical layer for internally-generated stimuli corresponding to more abstract stimulus representations encoded by neural structures towards the parietal and temporal lobes (Binder, 2016). However, the model could be applied to any sensory processing hierarchy where two-way communication is plausible and a gradient from high-dimensional to low-dimensional representation exists. Here we present a framework for interpreting our model in the visual modality.

In veridical visual perception, knowing whether you're looking at a cat or a dog requires not just recognition, but perceiving the animal's constituent body parts, and the constituents of those parts also: shapes, curves, edges, and most fundamentally, a spatial pattern of light. A single perceptual experience can therefore contain a distribution of information all at once, spanning from highly-sensory (modal) in nature to highly-abstract (amodal). Therefore, to interpret the simulation outcomes in terms of phenomenological experience, we assume that each representational layer in the visual hierarchy which receives both feedforward and feedback input contributes information to the construction of conscious perceptual experience. We also assume that the type of information provided by each layer aligns with the layer's dimensionality such that high-dimensional representations contribute most to the sensory aspects of the experience while low-dimensional representations contribute most to the non-sensory aspects of the experience. Hence, low-level representations would be most influential in determining the literal image content of a visual experience (e.g. the exact location, orientation, and contrast of edges), while high-level representations would contribute most to the abstract understanding of what is being thought about or observed (e.g. knowing whether the edges in a scene form a cat or a dog).

Crucially, this implies that a perceptual experience *gains* sensory or abstract properties depending on how well neural regions with the corresponding high- or low-dimensional structure, respectively, are recruited during that experience. Therefore,

the further an internally-generated stimulus is able to spread towards the primary visual cortex, which encodes high-dimensional stimulus information, the more it may be experienced with properties approaching that of an actual visual image (i.e. a true mental image) rather than something that is pondered without sensory qualia (i.e. a non-sensory thought). This approach can therefore maintain consistency with traditional accounts implicating the primary visual cortex in mental imagery (Kosslyn, 2005; Kosslyn & Thompson, 2003; Pearson, 2019; Pearson & Kosslyn, 2015). Note that, given that hallucinatory visual experience is possible without corresponding retinal activity, we consider the lowest level in the hierarchy purely as a non-conscious input layer to the rest of the system. We also note that while there's compelling evidence that some high-level representations are stored in a largely amodal form (e.g. Popham et al., 2021), they are not necessarily completely non-sensory. Late-stage object representations can still contain some degree of sensory information (e.g. see DiCarlo & Cox, 2007), albeit in a considerably more low-dimensional form relative to primary sensory areas.

3 Results

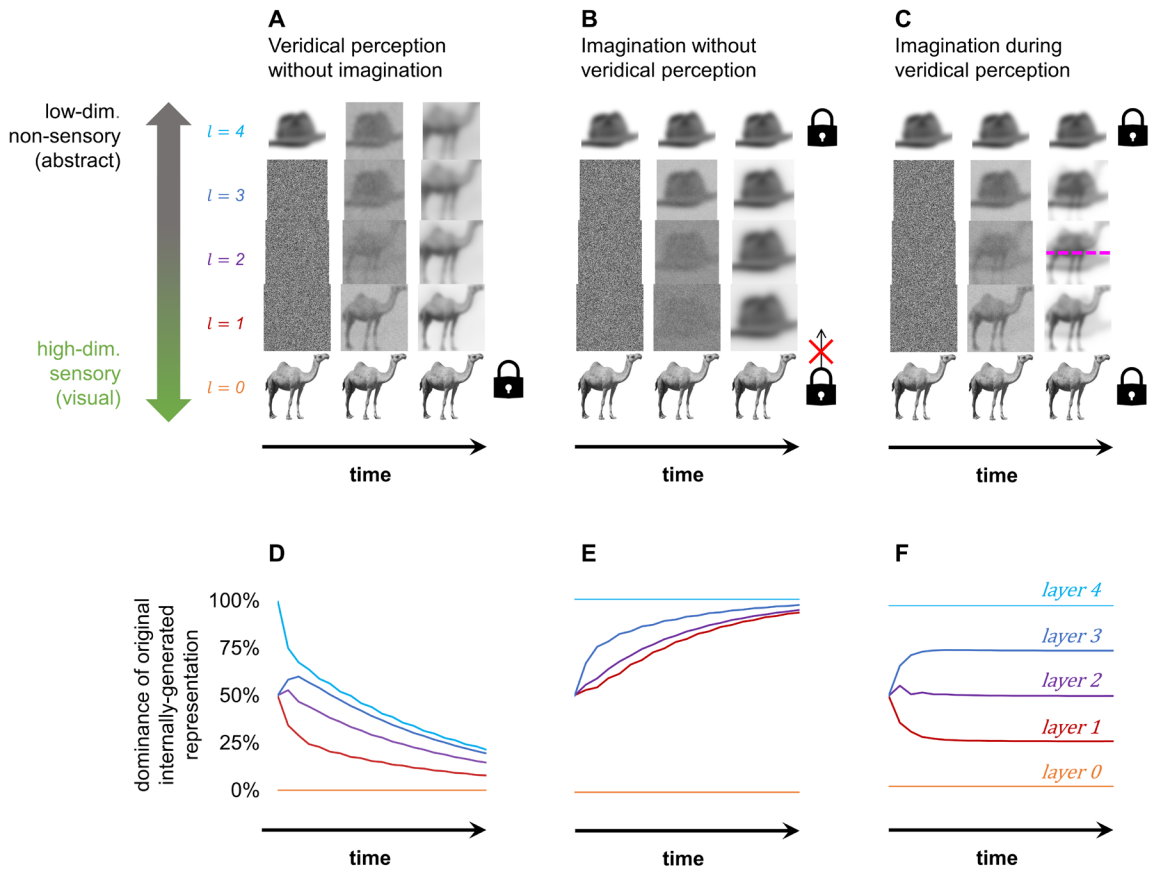
3.1 Perception without competition

For both simulations where only one stimulus was attended to, either external sensory input or an internally-generated imagined stimulus, only one fixed source of perceptual information was present within the network. Consequently, all unclamped layers in the network converged asymptotically to a representation dominated by the attended, clamped stimulus (Figure 2A & B). As all unclamped representations are eventually dominated by the source stimulus, both scenarios result in states corresponding to having a visual experience (seeing) and having an amodal experience (thinking about abstractly) of the same stimulus. Hence without competition from imagination, the system evolves to emulate seeing a camel and knowingly thinking about the camel. Likewise, where the internally-generated stimulus is uncontested, the sensory characteristics of the imagined stimulus would be expected to completely dominate visual experience, resulting in the perception of what was imagined and not what was present in the environment. Accordingly, without competition from veridical perception, the system evolves towards knowingly thinking about a hat and also seeing the hat.

Although in both cases, layer $l = 1$ is entirely dominated by one stimulus, each state occurred through a distinct formation route. Note that the top-down, but not bottom-up, formation route resulted in substantial distortions at lower layers, reflecting the findings of Breedlove et al. (2020). Yet, these distortions only affect the *content* of the imagined stimulus, not the degree to which that stimulus is able to commandeer a hierarchical layer in this model.

Figure 2

Veridical stimuli outcompete imagined stimuli at low-level sensory representations, but not high-level abstract representations



Note. Hierarchical spread of representations of an imagined camel and veridical hat, with and without competition. **A-C.** Representations at each hierarchical layer over time using images as inputs. Each value in a layer's representational matrix here corresponds to the brightness of a specific pixel as it evolves over 0, 2, and 20 alternating feedforward or feedback sweeps. The initial input representations at the apex ($l = 4$) and base ($l = 0$) layers do not vary across scenarios (thinking about a hat while receiving a retinal image of a camel). All mid-layers were initialised with noise. Lock icons indicate clamped representations. The crossed arrow indicates disconnected sensory input ($w_{l=0} = 0$). All weightings are set to 1 except for at clamped layers where $w_{l-1} = w_{l+1} = 0$. **A.** All layers except the base layer are free to evolve. **B.** Mid-layers are free to evolve only. Sensory input is present but disconnected. **C.** Mid-layers are free to evolve only, and sensory input is present and connected. The magenta line indicates the *representational equilibrium point*—the point where the internally- and externally-generated stimuli equally contribute to a layer. **D-F.** One-dimensional simulations of each scenario above in A-C illustrating how the character of each layer evolves over time. Values were obtained by running the corresponding multi-dimensional simulation with an initial apical representational matrix of ones (all white), basal matrix of zeros (all black), and intervening layers of noise, then averaging across all pixels in a layer. The dominance of the original externally-generated stimulus representation at $l = 0$ can be obtained by subtracting the internally-generated stimulus dominance from 100%.

3.2 Perception with competing imagined and veridical stimuli

Where two fixed sources of information are simultaneously present within the network, neither imagined nor veridical stimulus can dominate every unclamped layer in the system (Figure 2C). Instead, the network represents two distinct stimuli simultaneously at the steady state. However, the further a given layer is from an input layer, the smaller the contribution of the corresponding input to the representation held by the given layer, as shown in Figure 2F for the case where feedforward and feedback weightings are equal ($w_{l-1}/w_{l+1} = 1$).

Consequently, where imagination occurs in the presence of sensory input, the externally-generated veridical stimulus tends to dominate low-level representations while the internally-generated imagined stimulus tends to dominate high-level representations. Such a system is therefore representing two stimuli simultaneously, yet at largely separate hierarchical levels. As lower levels are more suited for encoding sensory information, and higher levels for abstract, this state corresponds to a situation where an individual has largely non-sensory thoughts about something distinct from the visual content they are experiencing. Hence, the simulation suggests that sensory input outcompetes imagination at the low-level regions most likely to contribute to the most sensory aspects of perceptual experience.

3.3 Modelling variation in imagery ability (e.g. aphantasia and hyperphantasia) as a shift in representational equilibrium

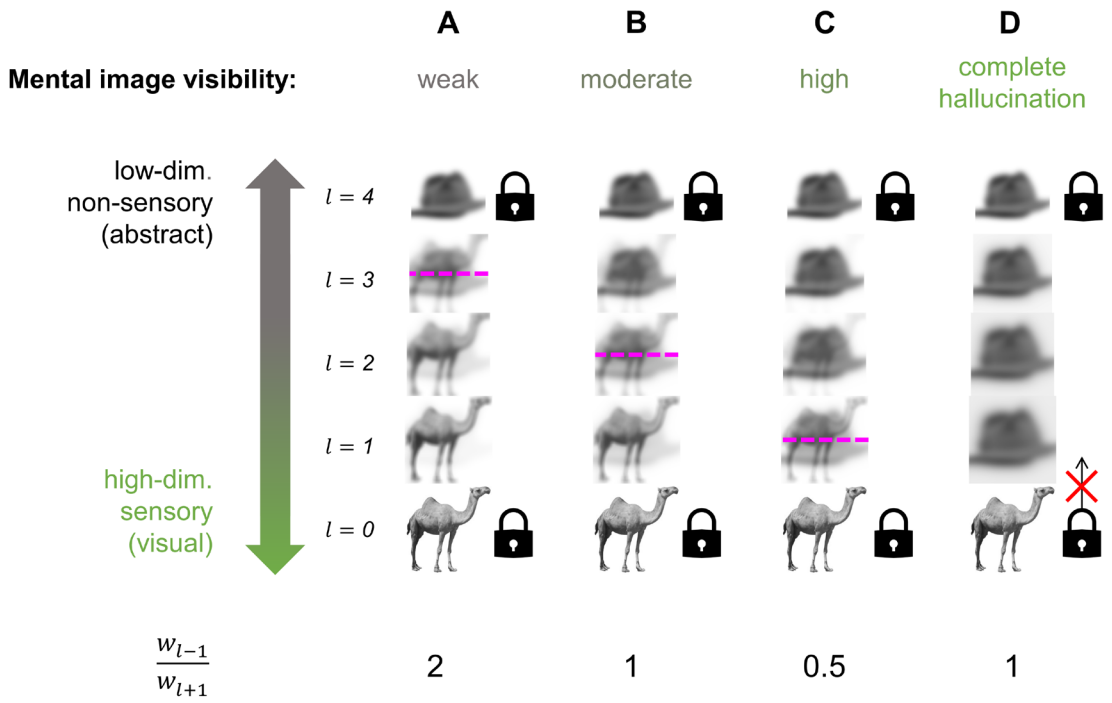
This model can also be used to understand states where thoughts are experienced in a highly sensory way (hyperphantasia) or in a non-sensory or weakly-sensory way (aphantasia) (Figure 3). When sensory input and imagination compete, our simulation showed that spanning bottom-up along the perceptual hierarchy corresponded to transitioning from primarily representing a veridical stimulus to primarily representing an imagined stimulus. The exact point where the two stimuli are represented in equal proportion is here dubbed the *representational equilibrium point* for the system. Representations contain more veridical stimulus character below this point and more imagined stimulus character above it.

The location of the representational equilibrium point can therefore summarise how far an imagined representation can penetrate towards low layers before it is outcompeted by the veridical stimulus, or how far a veridical stimulus representation can penetrate towards upper layers before it is outcompeted by the imagined stimulus. Where ascending and descending information is weighted equally ($w_{l-1}/w_{l+1} = 1$; Figure 3B), the equilibrium point lies at the middle layer in the centre of the hierarchy such that lower layers (e.g. towards the primary visual cortex) are dominated by externally-generated sensory input and upper layers (e.g. towards the temporal cortex) are dominated by internally-generated thoughts, symmetrically. However, overweighting feedforward information ($w_{l-1}/w_{l+1} > 1$) elevates the equilibrium point such that the imagined stimulus is restricted to only the highest, least-modal hierarchical layers, corresponding to a state of aphantasia (Figure 3A). In contrast, overweighting feedback information ($w_{l-1}/w_{l+1} < 1$) lowers the equilibrium point such that the imagined

stimulus dominates the majority of the hierarchy, corresponding to a state of hyperphantasia (Figure 3C). Feedforward-to-feedback weighting ratio variation could therefore account for between-person and moment-to-moment differences in the apparent sensory quality of mental imagery experiences.

Figure 3

Overweighting feedback information induces more sensory mental images



Note. Multiple hierarchical competition simulations of thinking about a hat while receiving a retinal image of a camel using different feedforward-to-feedback weighting ratios. Each value in a layer's representational matrix here corresponds to the brightness of a specific pixel after 20 alternating feedforward or feedback sweeps. The initial input representations at the apex ($l = 4$) and base ($l = 0$) layers were the same for each scenario. All mid-layers were initialised with noise. Lock icons indicate clamped representations. The crossed arrow indicates disconnected sensory input ($w_{l = 0} = 0$). In each scenario, all mid-layers feedforward or feedback weightings were ≤ 1 with the specified ratio, and did not change across layers. For clamped layers, $w_{l-1} = w_{l+1} = 0$. The magenta line indicates the *representational equilibrium point*—the point where the internally- and externally-generated stimuli contribute equally to a layer. **A-C.** Globally varying the ratio of bottom-up (w_{l-1}) and top-down (w_{l+1}) weightings affects representational equilibrium and the extent to which thoughts should be experienced visually (e.g. weak, moderate, or high). **D.** All unclamped layers converge towards the original internally-generated stimulus representation once sensory input is disconnected.

4 Discussion

4.1 Distinguishing imagined and veridical percepts

In this study, we aimed to explain the apparent differences in vividness between veridical, imagined, and hallucinated percepts. To do so, we first conceptualised perception as an experience composed of information varying in nature from highly-sensory to highly-abstract. We then proposed that the degree to which perceptual information can be represented by our brains in a high-dimensional space may determine the degree to which that information is experienced as a sensory, rather than non-sensory, percept. Under this framework, we showed that, for a hierarchical system with bidirectional information flow, the presence of competing sensory input prevented internally-generated stimuli from dominating low-level regions most suited for supporting a modal, highly-sensory representation. We interpret this finding as suggesting that, when competing sensory input is present, the sensory content of imagination does not supersede the sensory content of veridical perception. That is, even if we imagine a hat, our mental image of the hat does not replace the image of a real camel in our visual field: the camel is still more visible. In contrast, the attenuation or abolition of competing external sensory input could facilitate internally-generated stimuli becoming the dominant contributor to sensory experience, creating hyperphantasic or hallucinatory imagery. Our model therefore illustrates that the apparent difference in reported vividness between veridical and imagined percepts could be explained by the degree to which imagined and veridical stimulus representations spread hierarchically, and that this difference can arise due to external sensory input competing with our internally-generated thoughts.

Our model results in some counterintuitive implications. Despite the suitability of the visual cortex for representing retinotopic, visual information, our account suggests that hierarchically low-level neural regions analogous to layer $l = 1$ (e.g. primary sensory cortices) may be less involved in representing an imagined stimulus than any other region in the corresponding perceptual hierarchy. However, this implication is consistent with existing findings. Individuals still report being able to create mental images despite near-complete bilateral lesions to primary visual cortex, suggesting that the primary visual cortex is not necessary for visual mental imagery (Bridge et al., 2012; Chatterjee & Southwood, 1995; de Gelder et al., 2015; Zago et al., 2010). Further, a systematic review of the neural correlates of visual mental imagery by Spagna et al. (2020) found that the left fusiform gyrus, a high-level region within the visual ventral stream, was the only region reliably involved during mental imagery across studies (although the authors do not exclude the involvement of more anterior regions of the temporal lobe given the known difficulty in imaging these regions). These findings are entirely in line with the predictions of our model: while primary sensory cortices can be involved in mental imagery, the only essential component of the experience is the internal generation of a non-veridical stimulus representation, which in our model occurs at the hierarchical apex, corresponding, for instance, to late in the visual ventral stream. This may be why subjective vividness ratings better correlate to how distinctly imagined information is represented in retinotopic, rather than associative, neural regions (S.-H. Lee et al., 2012). Under our model, activation at levels below the clamped imagined representation can add progressively more

sensory components to the experience of imagining, although at what stage a relatively amodal, abstract thought formally becomes a sensory mental image may be an arbitrary distinction.

Our model also predicts that, while moving from posterior to anterior regions of the visual hierarchy, the dominance of internally-generated stimuli should gradually increase. In agreement, VanRullen & Reddy (2019) used a mental imagery decoding paradigm to show that more information about imagined face images was present in temporal regions than occipital regions. Lee et al. (2012) likewise showed that less information about imagined object images could be decoded towards more posterior, retinotopic areas of the visual system. Both, studies also showed that externally- and internally-generated stimulus representations became more similar towards higher-level regions, aligning with the notion that imagined content is propagated from areas active during late-stage veridical perception (Breedlove et al., 2020).

4.2 Functional impacts of competition between imagined and veridical stimuli

Our model also has implications for the functional aspects of imagery. For instance, the relegation of imagination to relatively high-level representations suggests that thinking in a non-sensory may be more commonplace than manifesting thoughts in a sensory way (e.g. as a visual image). This makes pragmatic sense in everyday life: thinking about what you are seeing is often favourable during veridical visual perception, yet it may be quite unfavourable to hallucinate what you are thinking about given that hallucinations may supersede important survival-relevant environmental information. However, this process could involve a functional trade-off, as the semantic content of sensory input may be processed poorly if high-level regions are already occupied by unrelated thoughts. This could contribute to the environmental inattentiveness present during daydreaming (Schooler et al., 2011; Smallwood, 2011). Another side-effect of abstract representations being activated without any significant low-level representation during imagery is that a perceiver might plausibly even *feel* like they have seen their imagined thought without ever experiencing it in any substantially visual way. That is, units involved in the recognition of an object may be activated despite minimal low-level activation. Hence people could theoretically report being able to visualise objects yet fail to be able to reconstruct the details of these imagined stimuli accurately.

4.3 Hallucinations

Our model also delineates conditions for hallucinations. Visually, a hallucination entails seeing a non-veridical percept in a given region of the visual field *instead of* a percept that reflects external reality. Under our model, for any modality, the most perceptually compelling hallucinations would then be those where a non-veridical stimulus dominates most, if not all, low-level representations. In our model, this can occur from overweighting top-down information relative to bottom-up information. However, hallucinations could also occur simply if aberrant low-level activity is fed through an otherwise normal perceptual system (Hahamy et al., 2021). Note that in both cases,

our model only accounts for the perceptual experience of imagery, hallucinated or otherwise, and not whether it was induced voluntarily (as imagery tends to be), or involuntarily (as hallucinations tend to be).

In the first route to hallucinations, top-down information becomes overweighted such that non-veridical thoughts can permeate low-level areas to nearly the same extent as uncontested bottom-up veridical input. In this case, hallucinations are modelled as an anomalous case of mental imagery with an extremely low representational equilibrium point, similar to hyperphantasia. If our model is correct, then individuals with hallucinatory disorders should tend to have improved mental imagery abilities. Accordingly, Parkinson's hallucinators tend to have stronger mental imagery priming effects than non-hallucinating controls, with the degree of imagery-induced priming predicting hallucination frequency (Shine et al., 2015). Parkinson's hallucinators also mind-wander more than non-hallucinators, corresponding to increased connectivity to the early visual cortex from default network regions (Walpola et al., 2020). Additionally, although individuals with schizophrenia can have impaired memory, their abilities on visuospatial mental imagery manipulation tasks are enhanced compared to neurotypical controls (Benson & Park, 2013; Matthews et al., 2014). These findings should be interpreted cautiously however given that some individuals are able to complete mental imagery manipulation tasks while reporting no visual mental imagery (Zeman et al., 2010).

Hallucinations should also occur when sensory input is disconnected altogether such that internally-generated percepts are entirely uncontested. If feedforward information is downweighted completely ($w_{l-1} = 0$) to any layer, then that layer and any layer above it are free to be commandeered by any high-level internally-generated stimulus present. Representations at such layers will then inevitably converge towards the internally-generated stimulus representation (Figure 2B, 2E, & 3D). This could contribute to hallucinations in disorders of afferent visual processing (e.g. Charles Bonnet Syndrome). Hallucinations are reasonably common following cortical blindness (Aldrich et al., 1987), and in one reported case, hallucinations could be induced directly from attempting voluntary mental imagery (Wunderlich et al., 2000). Dampening external sensory input might also contribute to hypnagogic hallucinations, which occur during sleep onset. When falling asleep, the cortex remains active several minutes after the thalamus begins silencing external sensory information (Magnin et al., 2010), potentially providing a window of time in which internally-generated cortical activity has reduced competition from the external environment.

These types of top-down influences likely aren't the only contributor to hallucinatory perception, however. Hallucinations can be visually detailed, and sometimes reportedly even clearer than veridical percepts (Ffytche et al., 1998; Manford & Andermann, 1998; Teunisse et al., 1996). This implies hallucinations cannot be a purely top-down phenomena given that retrohierarchically extrapolating high-level information into low-level regions should result in a degradation of visual information (Breedlove et al., 2020). Many hallucinations are therefore likely to also involve aberrant low-level cortical activity (Hahamy et al., 2021; Manford & Andermann, 1998) perhaps independent of, or in tandem with, amplified feedback signals.

4.4 Eyes-open versus eyes-closed imagery

We note that the model proposed in this paper does not presume any difference between an eyes-closed and eyes-open state apart from the change to the content of the sensory input each act produces. Though closing one's eyes might be ostensibly interpreted as disconnecting sensory input, if this were the case, our model would predict that closing one's eyes would result in hallucinations. While hallucinations can happen when the eyes are deprived of light, such phenomena generally require either drugs (Fisher, 1991), or multiple hours (if not days) of light deprivation to manifest (Merabet et al., 2004). This implies that an eyes-closed state alone does not equate to sensory disconnection. A formal disconnection or down-weighting of sensory input may more likely require large-scale neurochemical or neurophysiological changes, such as which occurs when shifting between different states of arousal. This may be why eyes-closed hallucinations in neurotypical individuals are generally isolated to sleep onset and offset (i.e. hypnagogic and hypnopompic hallucinations) or while asleep (i.e. dreaming), but not while awake, and why hallucinations which occur in blind individuals become most likely later in the day, regardless of light perception (Manford & Andermann, 1998).

Closing one's eyes (at least on short time-scales) might then be most aptly interpreted as changing sensory input to a diffuse, dark, and generally unremarkable stimulus image rather than disconnecting it altogether. Eye-closure tends to aid the recall of fine-grained sensory information, potentially by improving mental imagery, however not significantly more so than looking at blank space (Vredeveldt et al., 2011). However, we acknowledge that this too may be a simplification given that neural activity patterns between eyes-closed and eyes-open states can differ substantially even when both occur in darkness (Marx et al., 2003, 2004). Even so, an eyes-closed state could still improve mental imagery simply due to the signal-to-noise ratio advantage gained by pitting an imagined stimulus representation against a homogenous closed-eye scene, with the additional benefit of hiding attentional distractors.

4.5 Serotonin and acetylcholine: neurologically-plausible bottom-up and top-down weighting agents

This model implies that neural agents exist which modulate the relative contribution of top-down and bottom-up processes. Given that this model predicts that weighting schemes affect veridical and hallucinatory perception as well as mental imagery, agents which modulate these processes might also be apt candidate weighting agents in our current model.

Serotonin may be one such agent that modulates the contribution of bottom-up information during visual experience. Classic hallucinogens act on the serotonergic system via 5HT2A receptors (Nichols, 2016), and cortical serotonin levels are dependent on plasma levels of tryptophan, serotonin's blood-brain-barrier-permeable precursor (Fernstrom & Wurtman, 1971). Plasma tryptophan increases over the

course of the day, peaking in the late evening (Rao et al., 1994), approximately corresponding to the most common time of hallucination onset for individuals with Charles Bonnet Syndrome, peduncular hallucinosis, and Parkinson's Disease, including for those without any light perception ability (Manford & Andermann, 1998). Further, 5HT2A receptor expression in the ventral visual pathway of Parkinson's Disease hallucinators is elevated compared to non-hallucinators (Ballanger et al., 2010; Huot et al., 2010), while inverse agonists for this receptor are used to treat Parkinson's Disease hallucinations (Yasue et al., 2016). Outside of hallucinatory states, serotonin administered in the primary visual cortex of awake macaques acts as a gain modulator, dampening cell responses to external sensory input without affecting stimulus tuning or selectivity profiles (Seillier et al., 2017). In mice, this dampening in V1 was mediated by a 5HT2A receptor agonist (Michaiel et al., 2019).

Additionally, acetylcholine may modulate the contribution of top-down information during visual experience. The ratio of acetylcholine to serotonin is a factor in hallucination aetiology (Manford & Andermann, 1998), and both serotonin and acetylcholine target layer-five cortical pyramidal neurons, cells which have been proposed as essential for manifesting conscious sensory perception (Aru et al., 2019). High concentrations of acetylcholine can transition the activity of layer-five cortical pyramidal neurons from being driven by feedforward inputs to feedback inputs, which may mediate the hallucinatory experience of dreaming (Aru et al., 2020). The concentration of cortical acetylcholine peaks during rapid-eye-movement sleep, where dreaming is most likely, higher than during wakefulness and higher still than during non-rapid-eye-movement sleep where consciousness is greatly impaired (M. G. Lee et al., 2005). Hence, cortical acetylcholine concentrations may modulate feedback connectivity in sensory cortices.

Given their actions in the visual system, serotonin and acetylcholine levels could be plausible factors accounting for contextual and inter-individual differences in the sensory experiences of mental imagery. Accordingly, they could also be considered as factors which modulate inter-layer weightings within our proposed model. However, ascertaining the role of such neurotransmitters in mental imagery research is challenged by a reliance on human models for self-report, indistinct measurement constructs (e.g. vividness ratings), and the restricted access of psychoactive serotonergic drugs. Ultimately, there are many mechanisms by which our model can facilitate hallucinations and such mechanisms may work in tandem. Note that in all cases, our model does not explain how the initial internally-generated source stimulus might be produced, only how it may be represented, regardless of whether it was generated as part of a voluntary thought, involuntary hallucination, or otherwise.

4.6 Mental imagery in audition, other sensory-modalities, and as a dynamic process

Though we have focused on visual mental imagery, mental imagery can occur in other sensory modalities. A complete account of mental imagery should also be able to predict and explain modality-specific differences in imagery experience. Generative feedback might logically also account for imagined sensory content in non-visual

modalities, yet the quality of imagery produced may depend on the brain's ability to synthesise information specific to that modality. This is especially apparent with respect to inner speech, a subset of auditory imagery specifically relating to the imagination of oral sound content like words and onomatopoeia.

Unlike images, tastes, and smells, our body is readily capable of producing its own auditory stimuli using speech. Further, self-initiated movements, including that during speech, produce corollary discharges (or *efference copies*), signals which carry temporally-precise predictions of the sensory consequences of such movements (Curio et al., 2000). These signals have been suggested to supply or support imagined stimuli when produced in the absence of real movements (Jack et al., 2019; Scott, 2013; Whitford et al., 2017), effectively acting as sources of simulated sensory information.

We expect that the production of such somatosensory simulations could affect the temporal dynamics of simulations in our model. While the representational equilibrium point for a given hierarchy is determined by the ratio of ascending and descending weightings, it takes time for the system to stabilise towards equilibrium. Yet, if the internally-generated stimulus source representation is unstable while the system is evolving (i.e. clamping fluctuates in and out, such that sometimes $w_{l-1} \neq 0$), the system may never reach its predetermined equilibrium point. Hence, imagined stimuli may spend less time on average being represented by low-level regions than predicted by the representational equilibrium point alone, resulting in a less sensory mental imagery experience.

Anchoring an internally-generated source representation to a steady flow of somatosensory predictions, such as those provided by efference copies, could therefore stabilise an internally-generated source representation. Sensory modalities which can utilise such time-resolved sensorimotor simulations should then have improved mental imagery for the aspects simulated. For audition, this would include a boon to the content of speech imagery or other auditory imagery that could be mimicked through onomatopoeia, perhaps even encouraging auditory imagery to be habitually imagined as inner speech by default for some individuals. Yet, for visual mental imagery, eye movement simulations could provide information to solidify the spatial location of imagined images but would be uninformative regarding their content. Consequently, we expect auditory mental imagery to be a more compelling sensory experience than visual mental imagery, which may be related to the increased incidence of auditory hallucinations relative to visual hallucinations in schizophrenia (2.2x) and bipolar disorder (1.9x) (Waters et al., 2014).

4.7 Neural plausibility and opposing explanations

The primary intention of this article is to propose that competition from externally-generated sensory input could be a major factor which explains the unique phenomenology of mental imagery. To provide a falsifiable, explicit example of how this interaction could occur, we chose to operationalise our conceptual proposal using a simple hierarchical model framework grounded in tangible concepts (e.g.

dimensionality). Weighted averaging was adopted as a layer evolution rule in our model because it is an interpretable first-approximation of how competition may be carried out neurologically. Averaging could be implemented in the brain merely by transmitting bottom-up stimulus representations through an excitatory pathway while generative, top-down representations are transmitted through an opposing inhibitory pathway, or vice versa. However, the exact manner in which perceptual competition may be operationalised within sensory systems in the brain is unlikely to be encapsulated sufficiently by a simple weighted-average model. We nevertheless anticipate that our model can act as a springboard for neurophysiologically-precise models of mental imagery in the future. Note also that the model does not comment on any of the nonlinear transformations necessary to encode a representation within a layer, such as which occurs within a trained artificial neural network, but only makes assumptions about how such representations, once formed, may be transmitted and combined. Future studies may therefore seek to expand the approach here to alternate models of sensory processing, although our model is still capable at emulating findings on how imagined and veridical information is distributed in real human brains (e.g. Lee et al., 2012; Spagna et al., 2020; VanRullen & Reddy, 2019).

We also note that our model assumes that feedforward and feedback processes are computationally equivalent. However, Koenig-Robert & Pearson (2021) argue that while feedforward signals initiated from visual input may drive action potentials in the visual cortex, feedback signals tend to be more modulatory, guiding but not overriding activity induced by sensory input. If mental images are constructed using feedback processes, then this asymmetry could prevent mental images from ever outcompeting bottom-up veridical visual perception. This is one alternative explanation for the apparent difference in vividness between veridical and mental imagery. However, feedback connections within the visual cortex likely can drive action potentials in neurotypical individuals (Aru et al., 2020), including in neurons which are arguably crucial for manifesting conscious sensory perception (Aru et al., 2019), depending on local neurotransmitter concentrations. Whether such feedback connections are modulatory or driving may then more so depend on moment-to-moment fluctuations in cortical neurochemistry than hard neurophysiological limits (Aru et al., 2020). Asymmetries in feedforward and feedback processing can also still be operationalised in our current model by manipulating bottom-up and top-down weightings. However, our model distinctly predicts that imagined percepts should be experienced in a less sensory manner than veridical percepts, even if feedforward and feedback processes are equivalent, due to the interference caused by competing sensory input.

5 Conclusion

Ultimately, this work provides a formal explanatory framework for explaining the unique experience of mental imagery. Our model of hierarchical representational competition provides conditions for the different sensory experiences associated with veridical perception, hallucination, and mental imagery, as well as individual variation within these phenomena. We suggest that internally-generated thoughts are experienced along an axis from highly-abstract to highly-sensory in nature,

corresponding to the degree of dimensionality that the thought is represented in. We show that competition between external sensory input and imagination could potentially account for the degraded perceptual quality of mental images, supplementing the Breedlove et al. (2020) hierarchical generative inference account. Our approach is also consistent with the proposition that the degree to which mental images contain sensory information depends on the degree to which early sensory cortices are recruited during mental imagery, although our model predicts that low-level activity may not be a necessary component of imagination. The approach presented here also sufficiently differentiates the distinct visual experiences of veridical, imagined, and hallucinatory perception in terms detached from the ambiguity intrinsic to traditional vividness terminology. Finally, our model has the potential to integrate explanations of mental imagery in different sensory modalities, supporting its plausibility as a modality-general imagery model.

6 References

Aldrich, M. S., Alessi, A. G., Beck, R. W., & Gilman, S. (1987). Cortical blindness: Etiology, diagnosis, and prognosis. *Annals of Neurology*, *21*(2), 149–158. https://doi.org/10.1002/ana.410210207

Aru, J., Siclari, F., Phillips, W. A., & Storm, J. F. (2020). Apical drive—A cellular mechanism of dreaming? *Neuroscience & Biobehavioral Reviews*, *119*, 440–455. https://doi.org/10.1016/j.neubiorev.2020.09.018

Aru, J., Suzuki, M., Rutiku, R., Larkum, M. E., & Bachmann, T. (2019). Coupling the State and Contents of Consciousness. *Frontiers in Systems Neuroscience*, *13*. https://doi.org/10.3389/fnsys.2019.00043

Ballanger, B., Strafella, A. P., van Eimeren, T., Zurowski, M., Rusjan, P. M., Houle, S., & Fox, S. H. (2010). Serotonin 2A Receptors and Visual Hallucinations in Parkinson Disease. *Archives of Neurology*, *67*(4), 416–421. https://doi.org/10.1001/archneurol.2010.35

Benson, T., & Park, S. (2013). Exceptional visuospatial imagery in schizophrenia; implications for madness and creativity. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00756

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review*, *23*(4), 1096–1108. https://doi.org/10.3758/s13423-015-0909-1

Breedlove, J. L., St-Yves, G., Olman, C. A., & Naselaris, T. (2020). Generative Feedback Explains Distinct Brain Activity Codes for Seen and Mental Images. *Current Biology*. https://doi.org/10.1016/j.cub.2020.04.014

Bridge, H., Harrold, S., Holmes, E. A., Stokes, M., & Kennard, C. (2012). Vivid visual mental imagery in the absence of the primary visual cortex. *Journal of Neurology*, *259*(6), 1062–1070. https://doi.org/10.1007/s00415-011-6299-z

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLOS ONE*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

Chatterjee, A., & Southwood, M. H. (1995). Cortical blindness and visual imagery. *Neurology*, *45*(12), 2189–2195. https://doi.org/10.1212/WNL.45.12.2189

Chiu, L. P. W. (1989). Differential Diagnosis and Management of Hallucinations. *Journal of Hong Kong Medical Association*, *41*(3), 292–297.

Cichy, R. M., Heinzle, J., & Haynes, J.-D. (2012). Imagery and Perception Share Cortical Representations of Content and Location. *Cerebral Cortex*, *22*(2), 372–380. https://doi.org/10.1093/cercor/bhr106

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*, *9*(4), 183–191. https://doi.org/10.1002/(SICI)1097-0193(200004)9:4<183::AID-HBM1>3.0.CO;2-Z

de Gelder, B., Tamietto, M., Pegna, A. J., & Van den Stock, J. (2015). Visual imagery influences brain responses to visual stimulation in bilateral cortical blindness. *Cortex*, *72*, 15–26. https://doi.org/10.1016/j.cortex.2014.11.009

Dentico, D., Cheung, B. L., Chang, J.-Y., Guokas, J., Boly, M., Tononi, G., & Van Veen, B. (2014). Reversal of cortical information flow during visual imagery as compared to visual perception. *NeuroImage*, *100*, 237–243. https://doi.org/10.1016/j.neuroimage.2014.05.081

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341. https://doi.org/10.1016/j.tics.2007.06.010

Dijkstra, N., Ambrogioni, L., Vidaurre, D., & van Gerven, M. (2020). Neural dynamics of perceptual inference and its reversal during imagery. *ELife*, *9*, e53588. https://doi.org/10.7554/eLife.53588

Dijkstra, N., Bosch, S. E., & Gerven, M. A. J. van. (2017). Vividness of Visual Imagery Depends on the Neural Overlap with Perception in Visual Areas. *Journal of Neuroscience*, *37*(5), 1367–1373. https://doi.org/10.1523/JNEUROSCI.3022-16.2016

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared Neural Mechanisms of Visual Perception and Imagery. *Trends in Cognitive Sciences*, *23*(5), 423–434. https://doi.org/10.1016/j.tics.2019.02.004

Dijkstra, N., Hinne, M., Bosch, S. E., & van Gerven, M. A. J. (2019). Between-subject variability in the influence of mental imagery on conscious perception. *Scientific Reports*, *9*(1), 15658. https://doi.org/10.1038/s41598-019-52072-1

Dijkstra, N., Mazor, M., Kok, P., & Fleming, S. (2021). Mistaking imagination for reality: Congruent mental imagery leads to more liberal perceptual detection. *Cognition*, *212*, 104719. https://doi.org/10.1016/j.cognition.2021.104719

Dijkstra, N., Mostert, P., Lange, F. P. de, Bosch, S., & van Gerven, M. A. (2018). Differential temporal dynamics during visual imagery and perception. *ELife*, *7*, e33904. https://doi.org/10.7554/eLife.33904

Fernstrom, J. D., & Wurtman, R. J. (1971). Brain serotonin content: Physiological dependence on plasma tryptophan levels. *Science*, *173*(3992), 149–152. https://doi.org/10.1126/science.173.3992.149

Ffytche, D. H., Howard, R. J., Brammer, M. J., David, A., Woodruff, P., & Williams, S. (1998). The anatomy of conscious vision: An fMRI study of visual hallucinations. *Nature Neuroscience*, *1*(8), 738–742. https://doi.org/10.1038/3738

Fisher, C. M. (1991). Visual Hallucinations on Eye Closure Associated with Atropine Toxicity. A Neurological Analysis and Comparison with other Visual Hallucinations. *Canadian Journal of Neurological Sciences*, *18*(1), 18–27. https://doi.org/10.1017/S0317167100031255

Hahamy, A., Wilf, M., Rosin, B., Behrmann, M., & Malach, R. (2021). How do the blind ‘see’? The role of spontaneous brain activity in self-generated perception. *Brain*, *144*(1), 340–353. https://doi.org/10.1093/brain/awaa384

Huot, P., Johnston, T. H., Darr, T., Hazrati, L.-N., Visanji, N. P., Pires, D., Brotchie, J. M., & Fox, S. H. (2010). Increased 5-HT2A receptors in the temporal cortex of parkinsonian patients with visual hallucinations. *Movement Disorders*, *25*(10), 1399–1408. https://doi.org/10.1002/mds.23083

Ishai, A. (2010). Seeing faces and objects with the “mind’s eye.” *Archives Italiennes de Biologie*, *148*(1), 1–9. https://doi.org/info:doi/10.5167/uzh-46897

Jack, B. N., Le Pelley, M. E., Han, N., Harris, A. W. F., Spencer, K. M., & Whitford, T. J. (2019). Inner speech is accompanied by a temporally-precise and content-specific corollary discharge. *NeuroImage*, *198*, 170–180. https://doi.org/10.1016/j.neuroimage.2019.04.038

Keogh, R., & Pearson, J. (2017). The perceptual and phenomenal capacity of mental imagery. *Cognition*, *162*, 124–132. https://doi.org/10.1016/j.cognition.2017.02.004

Koenig-Robert, R., & Pearson, J. (2021). Why do imagery and perception look and feel so different? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1817), 20190703. https://doi.org/10.1098/rstb.2019.0703

Kosslyn, S. M. (2005). Mental images and the Brain. *Cognitive Neuropsychology*, *22*(3–4), 333–347. https://doi.org/10.1080/02643290442000130

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., Hamilton, S. E., Rauch, S. L., & Buonanno, F. S. (1993). Visual Mental Imagery Activates Topographically Organized Visual Cortex: PET Investigations. *Journal of Cognitive Neuroscience*, *5*(3), 263–287. https://doi.org/10.1162/jocn.1993.5.3.263

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, *129*(5), 723–746. https://doi.org/10.1037/0033-2909.129.5.723

Lee, M. G., Hassani, O. K., Alonso, A., & Jones, B. E. (2005). Cholinergic Basal Forebrain Neurons Burst with Theta during Waking and Paradoxical Sleep. *Journal of Neuroscience*, *25*(17), 4365–4369. https://doi.org/10.1523/JNEUROSCI.0178-05.2005

Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, *59*(4), 4064–4073. https://doi.org/10.1016/j.neuroimage.2011.10.055

Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications*, *10*(1), 179. https://doi.org/10.1038/s41467-018-08080-2

Magnin, M., Rey, M., Bastuji, H., Guillemant, P., Mauguière, F., & Garcia-Larrea, L. (2010). Thalamic deactivation at sleep onset precedes that of the cerebral cortex in humans. *Proceedings of the National Academy of Sciences*, *107*(8), 3829–3833. https://doi.org/10.1073/pnas.0909710107

Manford, M., & Andermann, F. (1998). Complex visual hallucinations. Clinical and neurobiological insights. *Brain*, *121*(10), 1819–1840. https://doi.org/10.1093/brain/121.10.1819

Marks, D. F. (1973). Visual Imagery Differences in the Recall of Pictures. *British Journal of Psychology*, *64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x

Marx, E., Deutschländer, A., Stephan, T., Dieterich, M., Wiesmann, M., & Brandt, T. (2004). Eyes open and eyes closed as rest conditions: Impact on brain activation patterns. *NeuroImage*, *21*(4), 1818–1824. https://doi.org/10.1016/j.neuroimage.2003.12.026

Marx, E., Stephan, T., Nolte, A., Deutschländer, A., Seelos, K. C., Dieterich, M., & Brandt, T. (2003). Eye closure in darkness animates sensory systems. *NeuroImage*, *19*(3), 924–934. https://doi.org/10.1016/S1053-8119(03)00150-2

Matthews, N. L., Collins, K. P., Thakkar, K. N., & Park, S. (2014). Visuospatial imagery and working memory in schizophrenia. *Cognitive Neuropsychiatry*, *19*(1), 17–35. https://doi.org/10.1080/13546805.2013.779577

Merabet, L. B., Maguire, D., Warde, A., Alterescu, K., Stickgold, R., & Pascual-Leone, A. (2004). Visual hallucinations during prolonged blindfolding in sighted subjects. *Journal of Neuro-Ophthalmology: The Official Journal of the North American Neuro-Ophthalmology Society*, *24*(2), 109–113. https://doi.org/10.1097/00041327-200406000-00003

Michaiel, A. M., Parker, P. R. L., & Niell, C. M. (2019). A Hallucinogenic Serotonin-2A Receptor Agonist Reduces Visual Response Gain and Alters Temporal Dynamics in Mouse V1. *Cell Reports*, *26*(13), 3475-3483.e4. https://doi.org/10.1016/j.celrep.2019.02.104

Nichols, D. E. (2016). Psychedelics. *Pharmacological Reviews*, *68*(2), 264–355. https://doi.org/10.1124/pr.115.011478

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, *20*(10), 624–634. https://doi.org/10.1038/s41583-019-0202-9

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences*, *112*(33), 10089–10092. https://doi.org/10.1073/pnas.1504933112

Popham, S. F., Huth, A. G., Bilenko, N. Y., Deniz, F., Gao, J. S., Nunez-Elizalde, A. O., & Gallant, J. L. (2021). Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nature Neuroscience*, *24*(11), 1628–1636. https://doi.org/10.1038/s41593-021-00921-6

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, *15*(7), 319–326. https://doi.org/10.1016/j.tics.2011.05.006

Scott, M. (2013). Corollary Discharge Provides the Sensory Content of Inner Speech. *Psychological Science*, *24*(9), 1824–1830. https://doi.org/10.1177/0956797613478614

Seillier, L., Lorenz, C., Kawaguchi, K., Ott, T., Nieder, A., Pourriahi, P., & Nienborg, H. (2017). Serotonin Decreases the Gain of Visual Responses in Awake Macaque V1. *Journal of Neuroscience*, *37*(47), 11390–11405. https://doi.org/10.1523/JNEUROSCI.1339-17.2017

Shine, J. M., Keogh, R., O'Callaghan, C., Muller, A. J., Lewis, S. J. G., & Pearson, J. (2015). Imagine that: Elevated sensory strength of mental imagery in individuals with Parkinson's disease and visual hallucinations. *Proceedings. Biological Sciences*, *282*(1798), 20142047. https://doi.org/10.1098/rspb.2014.2047

Smallwood, J. (2011). Mind-wandering While Reading: Attentional Decoupling, Mindless Reading and the Cascade Model of Inattention. *Language and Linguistics Compass*, *5*(2), 63–77. https://doi.org/10.1111/j.1749-818X.2010.00263.x

Spagna, A., Hajhajate, D., Liu, J., & Bartolomeo, P. (2020). Visual mental imagery engages the left fusiform gyrus, but not the early visual cortex: A meta-analysis of neuroimaging evidence. *BioRxiv*, 2020.02.06.937151. https://doi.org/10.1101/2020.02.06.937151

Teunisse, R. J., Zitman, F. G., Cruysberg, J. R. M., Hoefnagels, W. H. L., & Verbeek, A. L. M. (1996). Visual hallucinations in psychologically normal people: Charles Bonnet's syndrome. *The Lancet*, *347*(9004), 794–797. https://doi.org/10.5555/uri:pii:S0140673696908697

VanRullen, R., & Reddy, L. (2019). Reconstructing faces from fMRI patterns using deep generative neural networks. *Communications Biology*, *2*(1), 1–10. https://doi.org/10.1038/s42003-019-0438-y

Vredeveldt, A., Hitch, G. J., & Baddeley, A. D. (2011). Eyeclosure helps memory by reducing cognitive load and enhancing visualisation. *Memory & Cognition*, *39*(7), 1253–1263. https://doi.org/10.3758/s13421-011-0098-8

Wais, P. E., Rubens, M. T., Boccanfuso, J., & Gazzaley, A. (2010). Neural Mechanisms Underlying the Impact of Visual Distraction on Retrieval of Long-Term Memory. *Journal of Neuroscience*, *30*(25), 8541–8550. https://doi.org/10.1523/JNEUROSCI.1478-10.2010

Walpola, I. C., Muller, A. J., Hall, J. M., Andrews-Hanna, J. R., Irish, M., Lewis, S. J. G., Shine, J. M., & O'Callaghan, C. (2020). Mind-wandering in Parkinson's disease hallucinations reflects primary visual and default network coupling. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *125*, 233–245. https://doi.org/10.1016/j.cortex.2019.12.023

Waters, F., Collerton, D., Ffytche, D. H., Jardri, R., Pins, D., Dudley, R., Blom, J. D., Mosimann, U. P., Eperjesi, F., Ford, S., & Larøi, F. (2014). Visual Hallucinations in the Psychosis Spectrum and Comparative Information From Neurodegenerative Disorders and Eye Disease. *Schizophrenia Bulletin*, *40*(Suppl_4), S233–S245. https://doi.org/10.1093/schbul/sbu036

Whitford, T. J., Jack, B. N., Pearson, D., Griffiths, O., Luque, D., Harris, A. W., Spencer, K. M., & Le Pelley, M. E. (2017). Neurophysiological evidence of efference copies to inner speech. *ELife*, *6*, e28197. https://doi.org/10.7554/eLife.28197

Wilson, H. R., & Wilkinson, F. (2015). From orientations to objects: Configural processing in the ventral stream. *Journal of Vision*, *15*(7), 4. https://doi.org/10.1167/15.7.4

Wunderlich, G., Suchan, B., Volkmann, J., Herzog, H., Hömberg, V., & Seitz, R. J. (2000). Visual Hallucinations in Recovery From Cortical Blindness: Imaging Correlates. *Archives of Neurology*, *57*(4), 561–565. https://doi.org/10.1001/archneur.57.4.561

Xie, S., Kaiser, D., & Cichy, R. M. (2020). Visual Imagery and Perception Share Neural Representations in the Alpha Frequency Band. *Current Biology*, *30*(13), 2621-2627.e5. https://doi.org/10.1016/j.cub.2020.04.074

Yasue, I., Matsunaga, S., Kishi, T., Fujita, K., & Iwata, N. (2016). Serotonin 2A Receptor Inverse Agonist as a Treatment for Parkinson's Disease Psychosis: A Systematic Review and Meta-analysis of Serotonin 2A Receptor Negative Modulators. *Journal of Alzheimer's Disease: JAD*, *50*(3), 733–740. https://doi.org/10.3233/JAD-150818

Zago, S., Corti, S., Bersano, A., Baron, P., Conti, G., Ballabio, E., Lanfranconi, S., Cinnante, C., Costa, A., Cappellari, A., & Bresolin, N. (2010). A Cortically Blind Patient With Preserved Visual Imagery. *Cognitive and Behavioral Neurology*, *23*(1), 44–48. https://doi.org/10.1097/WNN.0b013e3181bf2e6e

Zeman, A. Z. J., Della Sala, S., Torrens, L. A., Gountouna, V.-E., McGonigle, D. J., & Logie, R. H. (2010). Loss of imagery phenomenology with intact visuo-spatial task performance: A case of 'blind imagination.' *Neuropsychologia*, *48*(1), 145–155. https://doi.org/10.1016/j.neuropsychologia.2009.08.024