Vividness as the Similarity Between Generated Imagery and an Internal Model

Sean N. Riley & Jim Davies

Department of Cognitive Science

Carleton University

Author Note

This is a pre-publication version of this article. The citation for the final version of the article is Riley, S.N. & Davies, J. (2023). Vividness as the similarity between generated imagery and an internal model. *Brain and Cognition*, 169, 105988, https://doi.org/10.1016/j.bandc.2023.105988. Correspondence concerning this article should be addressed to Jim Davies, Department of Cognitive Science, Carleton University, 2201 Dunton Tower–1125 Colonel BY Drive, Ottawa, Canada, K1S 5B6. E-mail: jim@jimdavies.org

Abstract

Vividness in visual mental imagery has been relatively under-explored compared to imagery's representational format and neural mechanisms. In this paper, we take a deeper look at vividness and suggest that in re-framing it, we can potentially reconcile disparate findings regarding visual cortex activation during imagery. Unlike traditional views of vividness that define the concept in terms of perception, we frame vividness in terms of imagery's relation to an internal model; the closer the generated imagery is to this model, the more vivid it is. This view is considered alongside existing neuroscientific, psychological, and philosophical research, as well as directions for future research.

Keywords: Mental Imagery; Vividness; Aphantasia; Hyperphantasia; Qualia

Vividness as the Similarity Between Generated Imagery and an Internal Model

Introduction

After decades of research into visual mental imagery (VMI), there appears to be three core questions that remain unanswered: (i) is the underlying representation of VMI depictive or descriptive? (ii) What is the role of the visual cortex in VMI? And (iii) how do we define vividness? There is no shortage of literature on the first two questions, but the third remains comparatively neglected. This paper will explore how a deeper investigation of vividness may help shed light the first two questions.

The debate over VMI's representational format has a long and rich history, and with the advent of more advanced imaging techniques has largely turned to understanding the role of early visual areas in VMI. If the topologically-organized areas of the visual cortex can be linked to VMI, then one would have evidence supporting the idea that VMI's underlying representation is depictive. And there have been studies that support this (e.g., Kosslyn et al., 2006), with more recent investigations suggesting that early visual areas encode low-level visual features, such as orientation and space (Naselaris et al., 2015), and that top-down connections into the V1 project to deeper layers linked to image segmentation and the filling in of features (Bergmann et al., 2019; Kok et al., 2016; Self et al., 2013). This coincides with some preliminary evidence that VMI might operate in a manner similar to predictive coding under perception, with top-down connections carrying the representation to early visual areas, and bottom-up connections carrying prediction errors (Dijkstra et al., 2020).

Unfortunately, the evidence tying VMI to these regions is correlational, and there is a growing body of causal evidence suggesting they are not explicitly required for VMI (e.g., Bartolomeo et al., 2020; Bridge et al., 2012; de Gelder et al., 2015; Zago et al., 2010). This causal evidence is further supported by correlational work on aphantasia (a lack of mental imagery experiences) that suggests that aphantasia is linked to damage along the ventral "what" pathway (Thorudottir et al., 2020), and that what neurally separates hyperphantasics from aphantasics are stronger top-down connections from prefrontal areas to occipital regions, and increased activation in the left anterior parietal region in the precentral gyrus during imagery (Milton et al., 2021). This falls in line with an alternative view of VMI, which suggests (i) that VMI is centered on a fusiform imagery node that receives semantic information from the left temporal lobe, (ii) that the medial temporal lobe works to combine past experiences into episodic simulations via VMI, (iii) that VMI activity in temporal regions is initiated and maintained by fronto-parietal networks, and (iv) that vividness stems from how well involved regions integrate information (Spagna et al., 2021). However, this view also precludes the recruitment of visual areas, which is potentially problematic given the convergence of correlational evidence supporting their selective involvement. It could be that this correlational evidence is spurious, but this conflicting pattern of results might be reconciled by a deeper investigation of vividness.

To this end, the simplest and most common approach to measuring vividness in VMI comes via self-reports, such as the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973), Betts' Questionnaire Upon Mental Imagery (Betts' QMI; Sheehan, 1967), and the Plymouth Sensory Imagery Questionnaire (Psi-Q; Andrade et al., 2014). In the case of the VVIQ, the questionnaire asks participants to rate their visual mental images on a 5-point scale where 1 = *perfectly clear and as vivid as normal vision* and 5 = *no image at all* (Marks, 1973). By contrast, Betts' QMI contains 35 scenarios that participants are instructed to imagine (e.g., "the sun sinking below the horizon"), with the resulting mental image's clearness and vividness then being rated on a 7-point scale: 1 = *perfectly clear and vivid*, 7 = *no image at all* (Sheehan, 1967). Finally, the Psi-Q takes a sampling of items from both the VVIQ and Betts' QMI, and combines them with 25 new questions to produce a 35-item questionnaire whose questions come in the form "Imagine {sensory modality} {item}," (e.g., "Image the appearance of a bonfire"; "Imagine touching fur") with each question being scored on a scale of 0 = *no image at all* to 11 = *as vivid as real life* (Andrade et al., 2014). However, measures such as the VVIQ assess vividness at the trait level, and there is some evidence to suggest it is better to consider vividness at the state level via trial-by-trial self-reports, as they provide a more accurate and reliable measurement (e.g., D'Angiulli et al., 2013; M. Runge et al., 2015; M. S. Runge et al., 2017).

Regardless of the level at which one assesses vividness, there remains the broader question of how to define it. In the aforementioned questionnaires, vividness is defined in relation to perception such that the greater the overlap between VMI and perception – particularly along the dimensions of clarity and detail – the more vivid the VMI. Sometimes this is explicit, such as "as clear and vivid as normal vision" in the VVIQ, but sometimes it is suggested, as in the Betts' QMI's "perfectly clear and vivid" option. This provides an easy-to-understand definition, but also relies on subjective, intuitive notions of detail and clarity that can vary between participants (Denis, 1995). Others have attempted to ground vividness in objective measures, such as the overlap in visual cortex activation under perception and VMI (e.g., Cui et al., 2007; Fazekas et al., 2020), but given VMI's selective recruitment of visual areas, this perceptual-overlap view of vividness appears problematic.

Moreover, there are also important philosophical considerations that any definition needs to take into account. Clarity, when used to describe visual perception, refers to blurriness/fuzziness (e.g., an image out of focus), but this sense of clarity does not seem to capture clarity in VMI; the generated image is not unclear in the same way that removing one's glasses makes the world unclear, it is unclear in some other sense of the word (Kind, 2017). One may be tempted to shift vividness away from clarity and more towards detail, but this is also problematic. A VMI of a solid white wall can be just as vivid as that of a checkerboard wall, even though it has fewer details. We could, however, think of detail in terms of determinacy and suggest that the more determinate a VMI is, the more vivid it is. Though this also falls apart as one's VMI of a tiger with an indeterminate number of stripes can be just as vivid as one's VMI of solid black panther whose colour is fully determinate (Kind, 2017). So it seems we need an objective account of vividness that (i) does not depend on visual cortex activity, and (ii) is not framed strictly in terms of perception.

Tooming and Miyazono (2020) provide one such account, framing vividness as a natural kind. Here, they suggest that vividness is a homeostatic property cluster– i.e., a cluster of co-occurring properties, none of which are strictly necessary (*detail, clarity, intensity, perception-likeness*), and that have an underlying mechanism explaining their co-instantiation under VMI: the availability of sensory information in long-term memory. This vividness-as-availability perspective suggests that a highly-vivid VMI is highly vivid because sensory information from long-term memory has a higher level of availability to the neurocognitive mechanisms that drive imagery, which allows for the construction and manipulation of a VMI that contains a high level of (at least some of) the properties from the cluster (e.g., detail, clarity, and so on). For example, the increased availability of sensory information in long-term memory might make it easier to fill in boundary information and/or represent details of the image with higher resolution, leading to a higher level of clarity (Tooming & Miyazono, 2020).

However, this view of vividness is largely schematic and does not lay out how this availability precisely works. For example, does availability matter only during generation, or during inspection of the VMI, or both? In the case of generation, it is not hard to imagine two people having equal access to said information, but who differ in their ability to assemble the information into a coherent VMI. If vividness is defined in terms of access during generation, then these two people would have equally vivid imagery despite objective differences in said imagery. If access only matters during generation, then vividness is something external to the image representation itself. This means that two identical VMI representations would be considered to have different vividness depending on the access to long-term memory during their initial generation. This is problematic, and it may be more apt to suggest that access to sensory information helps facilitate the formation of vivid VMI, but does not play a role in defining vividness, much in the same way that cold weather facilitates the formation of slippery ice, but does not define what ice, or slipperiness, is.

Conversely, if availability matters during inspection of a VMI already created, then perhaps vividness is better understood in terms of the relationship between what was generated and what is stored in long-term memory. This shifts vividness towards being a relational property inherent to the VMI, but it again seems that availability serves as a facilitator rather than a definition; a VMI's vividness is defined by its relationship to items in long-term memory, but that relationship is facilitated by access, be it access that facilitates the generation of the VMI by way of memory retrieval, or access that facilitates comparisons between the generated VMI and items in memory. Regardless of the implementational specifics of Tooming and Miyazono (2020)'s view, there is at least some cursory evidence that the availability of sensory information plays a role in vividness (D'Angiulli et al., 2013), though the nature of this role remains unknown.

Alternative View

Alternatively, we suggest that vividness is rooted in beliefs and expectations (B&E), and that VMI occurs in a manner akin to predictive coding. Here, B&E serve as the priors, and the brain works to generate a signal that matches the predictive model born from these B&E. In other words, we suggest VMI is a method of exploratory constraint satisfaction (Van Leeuwen, 2013) whereby our B&E serve to create a predictive model of the world that gets simulated via VMI. Vividness, then, can be viewed as the extent to which the generated VMI matches this model. Moreover, VMI generation can be viewed through multiple dimensions: (i) encoding the external world into long-term memory, (ii) retrieval of items from long-term memory, (iii) generation of images from retrieved items, and (iv) error correction. Where one falls within this ERG–Error space will dictate what pattern of activation is leveraged to generate VMI, with the connectivity between relevant regions dictating the quality of these ERG and error correction processes, which, subsequently, dictates how vivid the VMI is.

To expand on all this, when we generate a VMI, we are "seeing" what the world would be experienced as, perhaps if altered in some way. What it would be like if we painted our walls red, for example, or if cats had wings, or any number of other things. These alterations can be radical, but in all cases, the components of the generated VMI are created based on our beliefs about them (e.g., our beliefs about cats, wings, and how the two should come together). Moreover, these beliefs come with expectations, and when I generate a VMI of my friend, I expect the generated image to look like my friend. However, because the VMI generation process is noisy and imperfect, the VMI that gets generated can differ from my B&E about them. It may be missing features, or otherwise be obscured in detail and clarity in some way or another; my visualized friend *should* look one way, based off my prior experiences with them, but looks another way due to blurriness, obscurity, or whatever else. Here, my B&E of my friend forms a predictive model of what they are supposed to look like, given my prior experiences with them; and when I generate the visualization, I am using VMI to simulate what the world would be like if they were standing in front of me. The more the generated VMI aligns with my model, the more vivid it is.

When it comes to the specifics of this view, we suggest that VMI's underlying representation is descriptive, but is also depictive by virtue of directly encoding continuous space into symbols. For example, from the perspective of a vector symbolic architecture (VSA), we can encode continuous space into a unitary vector via fractional binding (referred to as spatial semantic pointers, or SSPs; Voelker et al., 2021). In a VSA, symbols are mapped to d-dimensional vectors, denoted in bold. These symbols can then be combined via a binding operator of choice, such as circular convolution, denoted by $\circledast$; or via superposition, denoted $+$. This binding can be computed as

$$\mathbf{a} \circledast \mathbf{b} = \mathcal{F}^{-1}\{\mathcal{F}\{\mathbf{a}\} \bigodot \mathcal{F}\{\mathbf{b}\}\} \qquad \text{(Eq. 1)}$$

where $\mathcal{F}$ is the discrete Fourier transform and $\odot$ element-wise multiplication. This enables the creation of more complicated cognitive structures, such as

$\mathbf{mary} = \mathbf{hair} \circledast \mathbf{brown} + \mathbf{affect} \circledast \mathbf{happy}$, which can then be used in other cognitive computations. Following Voelker et al. (2021), to encode a value $k \in \mathbb{R}$ into $\mathbf{a}$, we compute

$$\mathbf{a}^k = \mathcal{F}^{-1}\{\mathcal{F}\{\mathbf{a}\}^k\} \tag{Eq. 2}$$

where $\mathcal{F}\{\mathbf{a}\}^k$ is element-wise exponentiation of the complex vector. Moreover, we can create a vector that encodes a point in 2-d space, $\mathbf{S}(x, y)$, by defining vectors for each axis, $\mathbf{x}$ and $\mathbf{y}$, then computing

$$\mathbf{S}(x, y) = \mathbf{x}^x \circledast \mathbf{y}^y. \tag{Eq. 3}$$

But we can also encode an entire region, R, in 2-d space:

$$\mathbf{S}(R) = \int_{(x,y)\in R} \mathbf{x}^x \circledast \mathbf{y}^y \; dxdy, \tag{Eq. 4}$$

and specify what region every object $i \in I$ encompasses within a larger scene:

$$\mathbf{scene} = \sum_{i\in I} \mathbf{S}(R_i) \circledast \mathbf{obj}_i, \tag{Eq. 5}$$

or every object's (x, y) coordinate location:

$$\mathbf{scene} = \sum_{i\in I} \mathbf{S}(x_i, y_i) \circledast \mathbf{obj}_i. \tag{Eq. 6}$$

These SSPs also have the convenient property that

$$\mathbf{a}^{x_1} \circledast \mathbf{a}^{x_2} = \mathbf{a}^{x_1+x_2}, \tag{Eq. 7}$$

which allows us to move around space in a continuous way by shifting the (x, y) coordinates by some delta:

$$\mathbf{S}(x_2, y_2) = \mathbf{S}(x_1, y_1) \circledast \mathbf{S}(\Delta x, \Delta y). \tag{Eq. 8}$$

If we want to retrieve the location of object i, we can compute $\mathbf{scene} \circledast \sim \mathbf{obj}_i$ where $\sim \mathbf{obj}_i$ is the inverse of $\mathbf{obj}_i$, which gives us back an approximation of $\mathbf{S}(\cdot)$ (see Komer & Eliasmith, 2020; Voelker et al., 2021, for an in-depth discussion of SSPs). Overall, the symbolic nature of these representations fall in line with Pylyshyn's view (Pylyshyn, 2002), but the ability to encode and move through continuous space supports Kosslyn's view as well (Kosslyn et al., 2006).

Moreover, in line with Martin (2002), we suggest that in visualizing X, we are imagining the experience of seeing X. Thus, we suggest object properties in VMI's underlying representation carry semantic content pertaining to experience. That is, the representation of a visualized red brick carries information about *experiencing* red, not red itself. To model this, we draw from emerging research on quality spaces (Lee, 2021). Here, the colour property, for example, can be viewed as a multi-dimensional quality space comprised of some constituent dimensions, such as hue, saturation, and brightness. As Lee (2021) points out, this quality space can be partitioned into regions, with each region having a corresponding conceptual label, and all of the points within each region reflecting values that are permissible for that label. For example, the label 'red' will correspond to a large region of the colour quality space as there are a large number of hue-saturation-brightness combinations that would be classified with the label 'red'. The label 'dark red' would then be a subset of the region for 'red', with 'brick red' then being a subset of the region for 'dark red'. And so on. This region-style approach to understanding properties relates to something that Lee (2021) calls precision:

> Consider your color experience in foveal vision versus in peripheral vision. In foveal vision, you see an object as a specific shade of red, such as crimson. But in peripheral vision, you no longer see it as any specific shade of red, but instead just as red. It is not merely that you see the object as a different specific shade of red across the two cases. Instead, even if your peripheral color experience represents its object as having some specific shade of red or other, it

> leaves open which shade of red that might be, and it is compatible with your peripheral color experience that you are seeing any given shade of red within a certain range. Speaking somewhat metaphorically, peripheral color experience is less sharp and crisp than foveal color experience. This difference in phenomenal character is what I call precision.

Here, the colour experience is not thought of as corresponding to a particular point in the colour space, but rather a region of it: the smaller the region, the more precise the colour experience– 'brick red' is more precise than 'dark red', with 'dark red' being more precise than 'red', and so on.

Speaking more formally, a regional model of a quality space as laid out by Lee (2021) is defined by the tuple $< \mathfrak{Q},\ dist,\ \mathfrak{R},\ \mu,\ qual,\ prec >$ where $\mathfrak{Q}$ is the set of all points that define the entire space, *dist* is a point-distance metric that gives the distance between two points, $\mathfrak{R}$ a set of regions within the space, μ a measure that captures degrees of (im)precision, *qual* a metric that captures qualitative/phenomenal similarity, and *prec* a metric that captures precision similarity (i.e., how similar two regions are in terms of their precision). Overall, there is a nice connection between SSPs and Lee (2021)'s regional approach to quality spaces. Biologically, $\mathfrak{Q}$ likely resides within memory, where regions can be encoded via (Eq. 4) and the entire quality space, $\mathbf{qsm}(\cdot)$, as

$$\mathbf{qsm}(Property) = \sum_{R_i, label_i \in\ \mathfrak{R}} \mathbf{label}_i \circledast \mathbf{S}(R_i) \qquad \text{(Eq. 9)}$$

where $label_i$ is the conceptual label attached to the region R_i, if such a label exists. Overall, we view the representation for the experience of a property instance (e.g., one's experience of the colour 'crimson red') as $\mathbf{S}(R_i)$ where R_i is all the points permissible for the instance, the representation for the instance itself as $\mathbf{S}(\delta)$ where δ reflects a specific coordinate within the property's quality space (e.g., a specific (hue, saturation, brightness) coordinate within the colour space, such as $\mathbf{S}(0.2, 0.4, 0.1)$), and the conceptual label for the instance as $\mathbf{label}_i$ (e.g., **crimsonRed**). For most cognitive processing we suspect the conceptual label is sufficient, but in the case of VMI, $\mathbf{S}(R_i)$ is used as VMI is thought to involve imagining the experience of seeing X. Thus, when a VMI is to be generated, relevant $\mathbf{S}(R_i)$ can be extracted by

$$\mathbf{S}(R_i) \approx \mathbf{qsm}(\mathit{Property}) \circledast \sim \mathbf{label}_i \tag{Eq. 10}$$

and used to construct an object vector

$$\mathbf{obj} = \sum_{p \in P} \mathbf{p} \circledast \mathbf{S}(R_i) \tag{Eq. 11}$$

where P is the set of all properties for the object (colour, texture, and so on). This object vector, be it for a brick, cat, or any other object that may be placed in a VMI, can then be combined with spatial information via (Eq. 5) or (Eq. 6); however, this process of extracting parts from memory and assembling them into a VMI is likely to be noisy. Thus, we need a mechanism to compare the properties from the generated VMI against those from memory. That is, to compare how things actually look against how they should look. To do this, we can compute the dot product between a generated object's property and each region within that property's quality space. For example, if one generates a VMI of a brick, the generated colour of the brick would be compared against each region within the colour quality space. This can be done by arranging all the $\mathbf{S}(R_i)$ regions within the colour quality space into a matrix $\mathbf{Q}$, and then multiplying $\mathbf{Q}$ by the generated colour vector $\mathbf{S}(R_{colour})$:

$$\mathbf{d} = \mathbf{QS}(R_{colour}). \tag{Eq. 12}$$

This will produce dot products that tell us how similar the generated colour is to each of the colour regions within the quality space (i.e., lemon yellow, dark blue, and any other colour the agent would recognize). If all our vectors are unitary, this similarity metric will fall between –1 and 1, where 1 indicates equal vectors. If the vectors are not unitary, then the cosine similarity between the generated property and quality space regions should be calculated instead of the dot product as we want to constrain the similarity metric to –1 and 1. Regardless, with a similarity metric in hand, we can get a continuous measure of (colour) property vividness by dividing the largest similarity by the size of the region $\mathbf{S}(R_{colour})$ is most similar to, or

$$\mathcal{V}_{colour} = \frac{max\ \mathbf{d}}{[R_{max}]} \qquad \text{(Eq. 13)}$$

where $[\cdot]$ denotes area in the continuous case and set cardinally in the discrete case, and R_{max} the region that is most similar to $\mathbf{S}(R_{colour})$. If R_{max} is a small region that is highly similar to $\mathbf{S}(R_{colour})$, the colour vividness for the object in question will be high. As R_{max} becomes larger in size and/or $\mathbf{S}(R_{colour})$ becomes less similar, vividness will decrease. And the same goes for all other properties in (Eq. 11), such as texture, hardness, and so on. There will be a quality space for the property, $\mathfrak{Q}_{property}$, defined by some constituent axes and partitioned into various regions that reflect different instances of that property. The texture quality space may have a large region for *scaly*, with subset regions for *rough scales*, *slimy scales*; and there may be a large *furry* region separate from this *scaly* regions, perhaps subset with regions for *mangy fur*, *fluffy fur*, and so on (Figure 1). But regardless of how the quality space is partitioned, the vector representation of the space, $\mathbf{qsm}(Property)$, can be constructed via (Eq. 9), and regions extracted from it via (Eq. 10).

[**Insert Figure 1 Here**]

However, (i) the vividness of a single object within a scene likely depends on more than one property (e.g., the vividness of an visualized brick being dependant on its visualized colour and texture), and (ii) the vividness of a scene likely depends on more than one object within said scene. To address the first point, when calculating the vividness of an object within the scene, we can take a weighted average of how vivid each property for an object is, where the weight is determined by by how important that property is to vividness– e.g., one person may place a great deal of importance on colour and little on texture, whereas another may do the opposite.[1] We can then repeat this process to address the second point, and get a continuous measure of vividness for the entire scene by taking the weighted average of all objects in the scene, where each weight is determined by the object's salience. More formally, with the set of all properties denoted by P and the set of all objects in a scene denoted by I, object i's vividness is then computed by way of

$$\mathcal{V}_i = \frac{\sum_{p \in P} w_p \mathcal{V}_p}{\sum_{p \in P} w_p}, \quad \text{(Eq. 14)}$$

and the vividness for an entire scene as

$$\mathcal{V}_{scene} = \frac{\sum_{i \in I} w_i \mathcal{V}_i}{\sum_{i \in I} w_i}. \quad \text{(Eq. 15)}$$

This leaves the task of transitioning from $\mathcal{V}_{scene}$ to a vividness judgment, as one might in an experiment (e.g., a rating of 2 on a Likert scale self-report measure). Here, we view vividness as a property inherent to a generated property/object/scene, only this vividness property is defined in terms of the generated property/object/scene's relation to the predicted model. This relationship is captured by $\mathcal{V}$, making the quality space for vividness a 1d line comprised of all possible values for $\mathcal{V}$, and which can be partitioned into regions according to task instructions (e.g., one region for each value on a Likert scale), or the individual's beliefs about vividness if no instructions are given. We can then encode the

[1] We are assuming there will be some individual differences here, but there is no evidence yet for or against this.

generated property/object/scene's $\mathcal{V}$ into the vector **vividness** via $\mathbf{vividness}^{\mathcal{V}}$ and compute the dot product between $\mathbf{vividness}^{\mathcal{V}}$ and every region in $\mathfrak{Q}_{vivid}$ to see which vividness region it is most similar to (i.e., $\mathbf{S}(R_{max})$). The conceptual label attached to the most similar region can then be extracted via

$$\mathbf{label} \approx \mathbf{qsm}(Vividness) \circledast \sim \mathbf{S}(R_{max}) \qquad \text{(Eq. 16)}$$

and passed to brain regions responsible for providing responses.

To recap, we suggest that VMI occurs in a manner akin to predictive coding whereby the brain works to generate a signal that matches a predicted model derived from B&E. This process serves as a form of exploratory constraint satisfaction whereby we seek to see what the world would be like if altered in some way. The vividness of the generated VMI is dependent on how closely it matches the model. This VMI is both descriptive and depictive, and carries information about the experience of the X that is being visualized. That is, VMI is simulating the experience of seeing X. Experiential content is viewed in terms of quality spaces that are partitioned into regions, with these regions being encoded into symbols (e.g., via fractional binding). Spatial information is likewise encoded, allowing for continuous space to be represented (and transitioned through) in a descriptive manner. Vividness, then, contains two components: a continuous value, $\mathcal{V}$, that can be computed at the property, object, or scene level; and a discrete judgment that corresponds to the types of state-level responses made on questionnaires and self-reports. The properties used to determine vividness are likely to vary from person to person, as should their relative importance to vividness. We have no formal methods to account for trait-level vividness judgments, but suggest it may come from an averaging of multiple state-level judgments.

Circling back to Kind (2017), we suggest that concepts such as detail and clarity are important to vividness, but not in the perception-based senses of the words. Instead, we suggest they are important in the sense of convergence. To give an analogy, consider the five instructions:

1. Get the thing from the place
2. Get the scissors from the place
3. Get the scissors from the kitchen
4. Get the red scissors from the kitchen drawer
5. Get the red scissors from the kitchen drawer beside the sink

These examples become increasingly clear and detailed as *thing* and *place* are replaced with more specific instructions. However, if we replace *kitchen*, *drawer*, and *sink* with *cloud*, *pocket*, and *rainbow*, the instructions lose their detail and clarity, largely owing to violations of our beliefs about scissors and where they tend to be located. We can generate vivid VMI of bizarre things, such as a cat with wings or scissors being stored in a cloud's pocket; but our B&E about cats, wings, and how the two might interact, guides our imagery thereof (Van Leeuwen, 2013). In other words, even though we know cats do not actually have wings, if they did, these winged cats should look like X because of our B&E about cats, wings, and their interaction. X would contain some set of expected features/properties, and the closer the generated VMI is to these expected features/properties, the more detailed and clear the VMI is. We suggest it is this sense of detail and clarity that is important to vividness, not the prototypical perception-based sense often used in the literature and shown to be problematic by Kind. Moreover, these senses of detail and clarity are captured nicely by our model. B&E are implicitly encoded via quality space regions, and violations of the predictive model are penalized when we compute $\mathcal{V}$; the more generated properties converge towards small regions within their respective quality space, the more vivid the VMI becomes.

This then leaves the question of how VMI manifests itself, neurally. To this end, we suggest that during perception, property information is encoded into long-term memory as a symbolic representation. When a VMI is to be formed, fronto-parietal regions retrieve the relevant property information from long-term memory, then use it to extract the corresponding region from the quality space (Eq. 10). Once the relevant property regions and spatial information are extracted, this information is passed to temporal regions for VMI generation, as per Spagna et al. (2021). However, the generation of a VMI will be monitored by fronto-parietal areas, which can in turn trigger the recruitment of visual areas. This recruitment should be triggered when the noise inherent to VMI generation causes the generated VMI to cross a degradation threshold. This could occur for any number of reasons, such as background noise bleeding into the signal or tuning curves that do not sufficiently cover the representation space; but more generally, this degradation is liable to occur when there are issues with memory retrieval, assembling retrieved items into a VMI, or dysfunction with the quality space itself. When this degradation occurs, fronto-parietal regions can disinhibit connections to visual areas so as to initiate a feedback-feedforward sweep through the visual hierarchy, generating a prediction error. How far down the hierarchy it goes depends on (i) what aspects need to be altered, with more fine-grain aspects travelling further down the hierarchy (e.g., fine-grained details of a leaf), and (ii) how important determinability is to the visualization. (E.g., The number of black dots on dice faces plays a larger role in defining the stimuli than the number of black spots on a cheetah, thus determinability can be said to be more important to visualizing dice than cheetahs). Thus, if one needs to increase the determinability of low-level visual features, V1 is likely to be invoked; otherwise, representations are likely to be constrained to later regions. But once at target region(s), an error term can be calculated and passed back to tempo-parietal regions where the VMI is then adjusted accordingly. If an individual is a weaker visualizer, then either a poorer error term will be calculated via visual areas and/or a poorer honing process will occur via tempo-parietal ones. Importantly, though, we would be remiss if we did not specify that this form of error correction is unconscious and automatic, and should not be considered the same as intentional changes being made to the VMI by the visualizer.

To generate this prediction error, the objects/features of the generated VMI are first attended to (consciously or unconsciously). This is thought to involve retrieving the label(s) in the quality space(s) that the generated feature(s) are most similar too (i.e., the label for R_{max}) and the label for the target region (i.e., what was supposed to be generated). Upon this inspection, if a mismatch is detected (e.g., the VMI is blood red when it's supposed to be brick red) then fronto-parietal regions can disinhibit connections into visual regions, allowing the two labels into the visual hierarchy. The top-down sweep would then produce sensory information for both labels, with this sensory information then being used to compute an error term that can be passed back to tempo-parietal areas and used to hone the VMI. Importantly, even though the visualizer has a model about what the visualization should look like, this model exists as a set of labels stored in long-term memory, with said labels being used to extract quality space regions that are then assembled into a VMI. Thus, there is a type difference between the model and the generated VMI. One is a comprised solely of labels (the model) and the other of quality space regions (the VMI). When the generated VMI is inspected, the labels for quality space regions are retrieved and compared against the labels for the model; does the VMI have the right colour(s)? Texture(s)? And so forth. The debate over strong versus weak perceptualism is beyond the purview of this text, but by our account VMI falls under the weak category, with sensory information only coming into consideration when error correction is needed. And even then, this sensory information is not part of the VMI itself, instead serving as a guide for honing said VMI.

As a proof of concept regarding the generation of **scene** vectors, we used (Eq. 6) to encode 8x8 MNIST images of handwritten digits from the UCI machine learning repository (Dua & Graff, 2017). First, we scaled down all pixel values in the image by a factor of 10. We then defined quality space regions for the colour concepts *black* (pixel values 0 and 0.1), *grey* (pixel values from 0.2 to 1.1), and *white* (pixel values from 1.2 to 2.5). With this colour quality space defined, we then used (Eq. 6) to create a scene vector. Here, each pixel was treated as a separate object in the scene, where $\mathbf{S}(x, y)$ encoded the pixel's position within the 8x8 image, and $\mathbf{obj} = \mathbf{colour} \circledast \mathbf{S}(R)$ where R is the colour region the pixel's value falls within (*black*, *grey*, or *white*).

With **scene** in hand, we then computed $\mathbf{scene} \circledast \sim \mathbf{S}(x, y) \circledast \sim \mathbf{colour}$ for every pixel in the image to extract the $\mathbf{S}(R)$ for that pixel (i.e., what region from the colour quality space was represented). Then, for each pixel, we used (Eq. 12) to get the similarity between the pixel's $\mathbf{S}(R)$ and each region within the colour quality space. As we can see in Figure 2, we were able to closely mirror the original image; the various shades of grey were encoded as being part of the *grey* region of the quality space, with the various shades of black and white being correctly encoded as well. So it seems the described computational approach is capable of generating and interacting with **scene** vectors, which gives some preliminary credence to our model of VMI. However, further research is needed to explore the plausibility of the suggested vividness computations, error correction, and the model's neural instantiation.

[**Insert Figure 2 Here**]

Discussion

Broadly construed, our theory holds that the vividness of a VMI consists of how similar it is to the model that was used to generate it, and this vividness is influenced by (i) the quality of memory encoding, (ii) the quality of memory retrieval (e.g., retrieving regions from encoded quality spaces), (iii) the quality of image generation, and (iv) the quality of error correction. This means there are multiple forms of neural processing capable of generating vivid VMI, which can help explain the conflicting pattern of results seen in the literature. One may have poor encoding–retrieval–generation (ERG) processes but strong error correction, or strong ERG processes with weaker error correction, or fall somewhere between those two extremes. Thus, where one falls within the ERG–Error space will dictate what pattern of neural activity is used to elicit vivid VMI, which can help explain the interindividual differences found in the literature. Those that make heavier use of error correction should rely more heavily on the visual cortex relative to those who make less use of error correction; however, this does not necessarily mean that greater visual cortex activation should correlate with less vivid VMI, as highly vivid VMI can be generated via strong error correction. Nor does it mean that greater activation should correlate with more vivid VMI, as there may be dysfunction in tempo-parietal areas that hinder the VMI generation and honing process. In fact, the proposed view lends itself to the generation of vivid imagery without any visual cortex recruitment. The aforementioned lesion studies have demonstrated that VMI can be generated in the face of bilateral damage to early visual areas, which makes sense under the proposed view as the visual cortex is not explicitly needed for VMI generation. Instead, the model would predict that bilateral cortical blindness in V1 would result in a reduced capacity to generate images where low-level feature determinability is more important, but not images where this determinability is less important. This model prediction matches Bridge et al. (2012) findings; patient SRB, who suffered from near-complete cortical blindness, struggled to generate imagery of checkerboards but was able to generate imagery of faces and houses.

In a related vein, that VMI's top-down connections into V1 project to deeper layers (Bergmann et al., 2019), themselves tied to image segmentation and the filling in of features (Kok et al., 2016; Self et al., 2013), provides at least some support for the idea that (i) the visual cortex, broadly construed, is leveraged for error correction, and (ii) how far representations travel down the visual hierarchy depends on the need for determinability, as these are precisely the computations one would need for increasing determinability of low-level features. Interestingly, there is also evidence that prior expectations evoke activity in deep layers of V1, at least in the case of perception (Aitken et al., 2020), so the notion that, under imagery, deep layers of V1 would receive information regarding expectations for the purpose of error correction seems plausible, though it should again be stressed that the proposed view suggests that the entirety of the visual cortex is responsible for error correction, and that posterior regions are only recruited in select cases. Moreover, the proposed view also coincides with evidence suggesting that vividness is associated with decreased inhibition from the intraparietal sulcus (IPS) to visual areas, and increased inhibition from the fusiform gyrus (FG) to visual areas (Dijkstra et al., 2017). Here, the increased inhibition of FG could help keep visual cortex noise from bleeding into the imagery signal, with the decreased inhibition of IPS perhaps being part of a broader gating mechanism that allows representations into the visual hierarchy for as-needed error correction. Whether FG and IPS perform these specific computations remains to be seen, but the pattern of results is consistent with the proposed theory. A related but tangential study by Bergmann et al. (2016) found that vividness was positively associated with a smaller primary visual cortex, and it could be that this reduced size leads to easier inhibition by FG, thereby allowing less noise into the generated VMI, which ultimately helps facilitate vividness. This idea that inhibition of visual areas improves vividness by way of a stronger signal-to-noise ratio is further supported by the work of Keogh et al. (2020), who found that reduced excitability of the visual cortex increased vividness, as did increased excitability of frontal regions. More generally, there is evidence suggesting that stronger connectivity between prefrontal areas and posterior visual areas is one of the factors that separates hyperphantasics from aphantasics (Milton et al., 2021), perhaps due to increased executive functioning and a better signal-to-noise ratio.

Overall, though, it seems to us that the disparate patterns of activation found in imagery research is likely a byproduct of (i) there being multiple paths towards imagery generation, depending on where one falls in the ERG–Error space, and (ii) there being slightly different activation patterns for maintenance and manipulation (Schlegel et al., 2013). It also seems to us that, regardless of which path one takes towards imagery generation, the overall quality of the generated VMI depends on the quality of their ERG and error correction processes, which themselves depend on the strength of neural connections between VMI-associated brain regions. For example, a person may rely heavily on error correction to generate imagery, thus relying more heavily on the visual cortex; however, if they lack strong top-down control, for instance, the quality of this error correction will be poor, resulting in a low-vividness VMI. All of this theorizing aside, the proposed view does not paint a complete picture of VMI as it does not take into account a variety of other facets, such as working memory, and the neural nuances that differentiate generation, maintenance, and manipulation, just to name a few. Thus, the proposed view of VMI only serves as a potential foundation to build upon, though testing its viability may prove challenging.

To this end, regardless of where one falls within the ERG–Error space, stronger visualizers should not only have better connectivity between imagery-related areas, but also greater top-down control over imagery and increased tempo-parietal activation relative to weaker visualizers. There is some evidence for this in the literature (e.g., Milton et al., 2021), but a more rigorous test may be to look at determinability and familiarity. Here, our model suggests that determinability should interact with familiarity such that increased familiarly reduces determinability's correlation with early visual areas. In fact, familiarity should elicit less recruitment of visual areas in general, as more familiar stimuli are liable to have stronger memory traces, reducing the needs for error correction.

Then there are observations surrounding quality spaces. Specifically, our theory suggests that quality space dysfunction should have negative downstream effects on imagery and its vividness. This dysfunction may manifest in terms of availability, such as the case with poor access to quality spaces or if said quality spaces only have a small number of large partitions, which broadly ties in with the work of Tooming and Miyazono (2020). Or it may manifest in terms of encoding/retrieval, with some sort of neural dysfunction causing signal degradation, such as poor connectivity between regions involved in VMI. Regardless, VMI's vividness is intimately linked to the size of the quality space regions used to construct it, so if this dysfunction prevents the encoding/retrieval of small(er) regions, the generated imagery will have reduced vividness. Interestingly, a study by Maxwell et al. (2017) found that psychopathy and borderline personality disorder (BPD) were negatively related to VMI abilities, and that empathy was positively related to VMI abilities. This makes sense within the context of our proposed view, as highly empathetic people tend to be more "in-tune" with their, and other's, personal experiences; whereas those who score high in psychopathy and BPD tend to score lower on empathy measures (Ali et al., 2009; Salgado et al., 2020), and have difficulties with insight and introspection (Haviland et al., 2004; New et al., 2012). Moreover, there is also evidence that alexithymia is negatively related to VMI (Campos et al., 2000; Mantani et al., 2005), further suggesting a potential link between quality space dysfunction and reduced VMI abilities. Overall, the specifics of the potential link between quality space dysfunction and reduced vividness in VMI needs to be elucidated, but provides an interesting avenue for testing aspects of the proposed theory.

Turning more towards the philosophical side of things, there is the interesting question of why VMI is typically less vivid than perception, and whether the proposed definition of vividness extends beyond VMI. In some cases of hyperphantasia, VMI can be comparably vivid to perception, as can hallucinations and dreams, so it seems the neural machinery is there to support this level of vividness, yet for most, there is a marked reduction in vividness with VMI (Davies, 2019). One factor at play may be the top-down nature of imagery, as generating *X* is a nosier and more complicated process than perceiving *X*, leading to less vivid representations; however, it could also be that experiential content simply does not lend itself as well to complex representation construction as perceptual content does. Putting these considerations aside, it may be a mistake to assume that there exists a single, universal definition of vividness. Consider HOT, for example. It is a subjective concept whose definition depends on context: red dwarf stars are considered cool and in the context of climate change the arctic is alarmingly hot. Something similar may play out with vividness whereby all forms of vividness involve some sort of comparison between the representation being perceived and a predicted model of some sort, but the computations underlying this comparison may vary depending on the nature of the perceived representation and its contents, as would the nature of the predicted model. For example, under foveal perception it would make sense that vividness is rooted in traditional, perception-based definitions of detail and clarity as the underlying representations carry perceptual content. Thus, determinations of vividness under perception should be rooted in traditional notions of contrast, saturation, etc., and how they relate to the predicted model. How this model is derived under perception, and how comparisons are made against it, is beyond the purview of this text, but we suspect both have deep ties to predictive coding.

To recap briefly, there is conflicting evidence that VMI involves topologically organized areas of the visual cortex. Because of this conflict, we are left to wonder if VMI is, in fact, depictive. However, we may be able to explain this conflicting pattern of results via a deeper investigation of vividness. Existing definitions of vividness are experimentally and philosophically problematic, but if we re-orient ourselves to view vividness in terms of the generated imagery's relation to a predictive model derived from B&E, this conflicting pattern of results takes a clearer focus. VMI does not explicitly require the visual cortex, but the visual cortex can be used for error correction. We find mixed correlational evidence for early visual areas' involvement in both vividness and, more generally, VMI, because there are multiple regions within the ERG–Error space that can elicit (vivid) VMI: Weaker encoding, retrieval, and generation processes can be supplemented with strong error correction, and vice versa. And by shifting VMI away from its more traditional, perception-based conception to the notion that in visualizing *X*, we are imagining the experience of seeing *X*, we are able to move vividness away from the perceptual-overlap view. Vividness no longer depends on imagery's overlap with perception – which would, presumably, necessitate visual cortex activation – but instead on representations that encoded information pertaining to experience (i.e., quality space regions). This may provide a more philosophically tenable view of vividness as VMI and its vividness is no longer tethered to perception-based senses of detail, clarity, and so forth. This can also help explain, in part, why imagery is phenomenally distinct from perception: VMI's representation of properties only carry phenomenal information, and when we inspect the generated VMI, we are inspecting said phenomenal information; we are inspecting our visualization of the experience of seeing *X*. When we report on the visualization's vividness, we are reporting on our experience of inspecting this visualization (i.e., what region of our vividness quality space the generated VMI activates). If upon inspection the generated VMI closely matches the predictive model, then the VMI will be reported as being highly-vivid.

This then leads to the question of whether VMI is depictive or descriptive. Here, we use spatial semantic pointers to encode and represent continuous space within a propositional system (a VSA, more specifically). If we go by Kosslyn's definition, then VMI as we suggest it would be descriptive. However, if we view "depictive" as the ability to encode and transition through continuous space, then VMI as we suggest it is, indeed, depictive, and we can get this depiction without requiring activation of topologically organized areas (or even the visual cortex, for that matter). Having said all this, one of the questions remaining is the role that spatial information plays in vividness. Because spatial information and quality space regions are encoded in the same manner (i.e., SSPs), we suspect that a similar comparison between the generated VMI and a spatial model can be made, but we also expect vividness to depend on complex interactions between spatial information and property information. We have purposefully eschewed spatial considerations for simplicity, but the role of spatial information on vividness is an interesting avenue for future research.

Overall, the proposed view of VMI and its vividness is an intriguing avenue to pursue. It presents a potential path to reconciling some conflicting findings within the literature, but the framework it builds runs counter some of the conventional wisdom on VMI. The debate over whether in visualizing X, we are imagining the experience of seeing X is beyond the purview of this text, but has the potential to untether us from some of the issues inherent to the perception-based view of VMI and vividness.

References

Aitken, F., Menelaou, G., Warrington, O., Koolschijn, R. S., Corbin, N., Callaghan, M. F., & Kok, P. (2020). Prior expectations evoke stimulus-specific activity in the deep layers of the primary visual cortex. *PLoS Biology*, *18*(12), e3001023.

Ali, F., Amorim, I. S., & Chamorro-Premuzic, T. (2009). Empathy deficits and trait emotional intelligence in psychopathy and machiavellianism. *Personality and Individual Differences*, *47*(7), 758–762.

Andrade, J., May, J., Deeprose, C., Baugh, S.-J., & Ganis, G. (2014). Assessing vividness of mental imagery: The plymouth sensory imagery questionnaire. *British Journal of Psychology*, *105*(4), 547–563.

Bartolomeo, P., Hajhajate, D., Liu, J., & Spagna, A. (2020). Assessing the causal role of early visual areas in visual mental imagery. *Nature Reviews Neuroscience*, *21*(9), 517–517.

Bergmann, J., Genç, E., Kohler, A., Singer, W., & Pearson, J. (2016). Smaller primary visual cortex is associated with stronger, but less precise mental imagery. *Cerebral Cortex*, *26*(9), 3838–3850.

Bergmann, J., Morgan, A. T., & Muckli, L. (2019). Two distinct feedback codes in v1 for ‘real’and ‘imaginary’internal experiences. *BioRxiv*, 664870.

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*, 1062–1070.

Campos, A., Chiva, M., & Moreau, M. (2000). Alexithymia and mental imagery. *Personality and Individual Differences*, *29*(5), 787–791.

Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Research*, *47*(4), 474–478.

D'Angiulli, A., Runge, M., Faulkner, A., Zakizadeh, J., Chan, A., & Morcos, S. (2013). Vividness of visual imagery and incidental recall of verbal cues, when phenomenological availability reflects long-term memory accessibility. *Frontiers in Psychology*, *4*, 1.

Davies, J. (2019). *Imagination*. Simon; Schuster.

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.

Denis, M. (1995). Vividness of visual imagery and the evaluation of its effects on cognitive performance. *Journal of Mental Imagery*, *19*(3-4), 136–138.

Dijkstra, N., Ambrogioni, L., Vidaurre, D., & van Gerven, M. (2020). Neural dynamics of perceptual inference and its reversal during imagery. *Elife*, *9*, e53588.

Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A., & Friston, K. (2017). Distinct top-down and bottom-up brain connectivity during visual perception and imagery. *Scientific Reports*, *7*(1), 1–9.

Dua, D., & Graff, C. (2017). UCI machine learning repository. http://archive.ics.uci.edu/ml

Fazekas, P., Nemeth, G., & Overgaard, M. (2020). Perceptual representations and the vividness of stimulus-triggered and stimulus-independent experiences. *Perspectives on Psychological Science*, *15*(5), 1200–1213.

Haviland, M. G., Sonne, J. L., & Kowert, P. A. (2004). Alexithymia and psychopathy: Comparison and application of california q-set prototypes. *Journal of Personality Assessment*, *82*(3), 306–316.

Keogh, R., Bergmann, J., & Pearson, J. (2020). Cortical excitability controls the strength of mental imagery. *eLife*, *9*, e50232.

Kind, A. (2017). Imaginative vividness. *Journal of the American Philosophical Association*, *3*(1), 32–50.

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology*, *26*(3), 371–376.

Komer, B., & Eliasmith, C. (2020). Efficient navigation using a scalable, biologically inspired spatial representation. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society.*

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery.* Oxford University Press.

Lee, A. Y. (2021). Modeling mental qualities. *Philosophical Review*, *130*(2), 263–298.

Mantani, T., Okamoto, Y., Shirao, N., Okada, G., & Yamawaki, S. (2005). Reduced activation of posterior cingulate cortex during imagery in subjects with high degrees of alexithymia: A functional magnetic resonance imaging study. *Biological Psychiatry*, *57*(9), 982–990.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*(1), 17–24.

Martin, M. G. (2002). The transparency of experience. *Mind & Language*, *17*(4), 376–425.

Maxwell, R., Lynn, S. J., & Lilienfeld, S. (2017). Failures to imagine: Mental imagery in psychopathy and emotional regulation difficulties. *Imagination, Cognition and Personality*, *36*(3), 270–300.

Milton, F., Fulford, J., Dance, C., Gaddum, J., Heuerman-Williamson, B., Jones, K., Knight, K. F., MacKisack, M., Winlove, C., & Zeman, A. (2021). Behavioral and neural signatures of visual imagery vividness extremes: Aphantasia versus hyperphantasia. *Cerebral Cortex Communications*, *2*(2), tgab035.

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, *105*, 215–228.

New, A. S., Rot, M. a. h., Ripoll, L. H., Perez-Rodriguez, M. M., Lazarus, S., Zipursky, E., Weinstein, S. R., Koenigsberg, H. W., Hazlett, E. A., Goodman, M., et al. (2012). Empathy and alexithymia in borderline personality disorder: Clinical and laboratory measures. *Journal of Personality Disorders*, *26*(5), 660–675.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, *25*(2), 157–182.

Runge, M., Bakhilau, V., Omer, F., & D'Angiulli, A. (2015). Trial-by-trial vividness self-reports versus vviq: A meta-analytic comparison of behavioral, cognitive and neurological correlations. *Imagination, Cognition and Personality*, *35*(2), 137–165.

Runge, M. S., Cheung, M. W., & D'Angiulli, A. (2017). Meta-analytic comparison of trial-versus questionnaire-based vividness reportability across behavioral, cognitive and neural measurements of imagery. *Neuroscience of Consciousness*, *2017*(1), nix006.

Salgado, R. M., Pedrosa, R., & Bastos-Leite, A. J. (2020). Dysfunction of empathy and related processes in borderline personality disorder: A systematic review. *Harvard Review of Psychiatry*, *28*(4), 238.

Schlegel, A., Kohler, P. J., Fogelson, S. V., Alexander, P., Konuthula, D., & Tse, P. U. (2013). Network structure and dynamics of the mental workspace. *Proceedings of the National Academy of Sciences*, *110*(40), 16277–16282.

Self, M. W., van Kerkoerle, T., Supèr, H., & Roelfsema, P. R. (2013). Distinct roles of the cortical layers of area v1 in figure-ground segregation. *Current Biology*, *23*(21), 2121–2129.

Sheehan, P. W. (1967). A shortened form of betts' questionnaire upon mental imagery. *Journal of Clinical Psychology*, *23*(3), 386–389.

Spagna, A., Hajhajate, D., Liu, J., & 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*.

Thorudottir, S., Sigurdardottir, H. M., Rice, G. E., Kerry, S. J., Robotham, R. J., Leff, A. P., & Starrfelt, R. (2020). The architect who lost the ability to imagine: The cerebral basis of visual imagery. *Brain Sciences*, *10*(2), 59.

Tooming, U., & Miyazono, K. (2020). Vividness as a natural kind. *Synthese*, 1–21.

Van Leeuwen, N. (2013). The meanings of "imagine" part i: Constructive imagination. *Philosophy Compass*, *8*(3), 220–230.

Voelker, A. R., Blouw, P., Choo, X., Dumont, N. S.-Y., Stewart, T. C., & Eliasmith, C. (2021). Simulating and predicting dynamical systems with spatial semantic pointers. *Neural Computation*, *33*(8), 2033–2067.

Zago, S., Corti, S., Bersano, A., Baron, P., Conti, G., Ballabio, E., Lanfranconi, S., Cinnante, C., Costa, A., Cappellari, A., et al. (2010). A cortically blind patient with preserved visual imagery. *Cognitive and Behavioral Neurology*, *23*(1), 44–48.

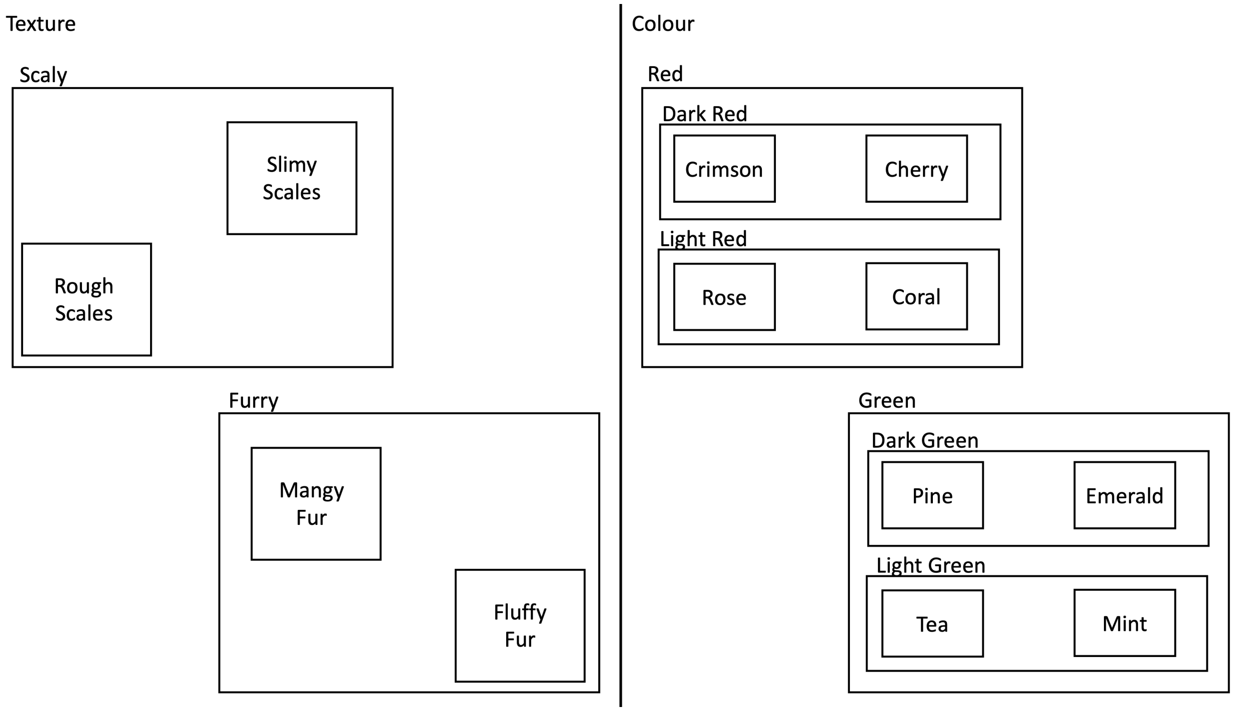

Figure 1

Schematic diagram of how a hypothetical texture (left panel) and colour (right panel) quality space may be partitioned. Here, each quality space contains a collection of large regions (i.e., red, green, scaly, furry) that are subdivided into more specific regions (e.g., mangy fur, slimy scales, etc.). Each regions reflects the points within the quality space that are permissible for the region's label (e.g., all the points that are permissible for dark red). The axes that define each quality space will vary between properties, and potentially between people.

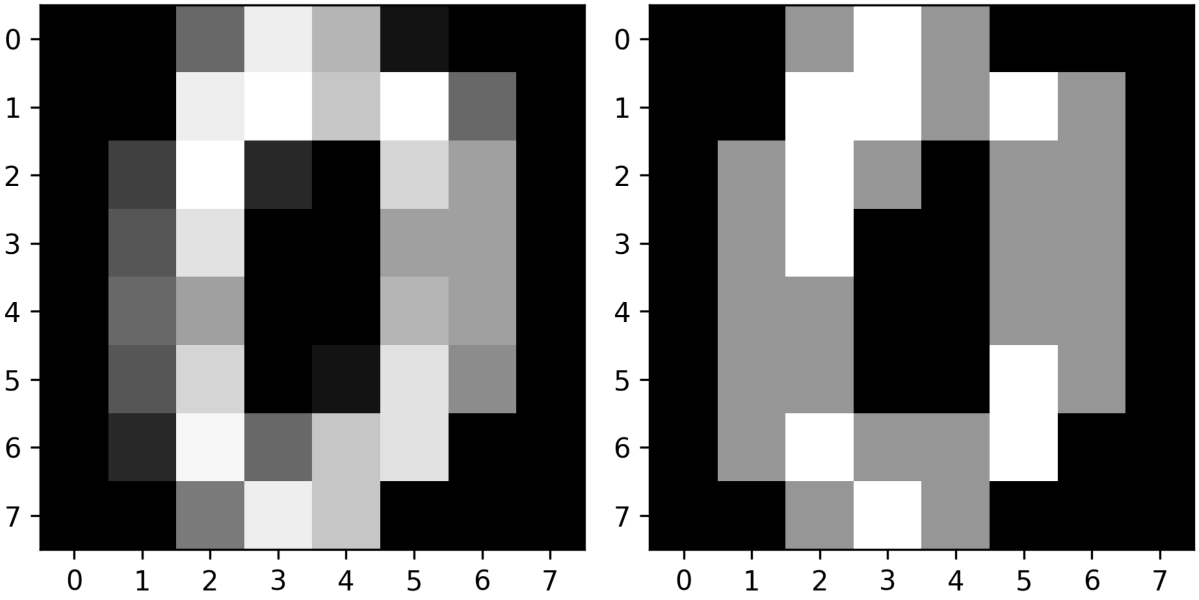

Figure 2

Left panel is the original image. Right panel is the representation encoded into the ***scene*** *vector (i.e., the generated VMI). In the right panel, the colour of each pixel reflects what region within the colour quality space was encoded/extracted from the VMI (white, black, or grey). That is, The black region was encoded at, and retrieved from, the pixel* $(0, 0)$*, the white region from the pixel* $(3, 0)$*, the grey region from the pixel* $(2, 0)$*, and so on.*