



## City Research Online

### City, University of London Institutional Repository

---

**Citation:** Epping, G., Fisher, E., Zeleznikov-Johnston, A., Pothos, E. M. & Tsuchiya, N. (2023). A quantum geometric framework for modeling color similarity judgments. *Cognitive Science*, 47(1), e13231. doi: 10.1111/cogs.13231

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/29388/>

**Link to published version:** <https://doi.org/10.1111/cogs.13231>

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

---

---

City Research Online:

<http://openaccess.city.ac.uk/>

[publications@city.ac.uk](mailto:publications@city.ac.uk)

---

# A Quantum Geometric Framework for Modeling Color Similarity Judgments

Gunnar P. Epping<sup>a,\*</sup>, Elizabeth L. Fisher<sup>b,c</sup>, Ariel M. Zeleznikow-Johnston<sup>b,d</sup>, Emmanuel M. Pothos<sup>e,1</sup>, Naotsugu Tsuchiya<sup>b,d,f,g,1</sup>

<sup>a</sup>*Department of Psychological and Brain Sciences, Indiana University, Bloomington, 47405, Indiana, USA*

<sup>b</sup>*School of Psychological Sciences, Monash University, Clayton, Victoria, Australia*

<sup>c</sup>*Cognition and Philosophy Lab, Philosophy Department, School of Philosophy, Historical and International Studies, Monash University, Clayton, Victoria, Australia*

<sup>d</sup>*Turner Institute for Brain and Mental Health, Monash University, Clayton, Victoria, Australia*

<sup>e</sup>*Department of Psychology, City, University of London, London EC1V 0HB, UK*

<sup>f</sup>*Center for Information and Neural Networks (CiNet), Osaka, Japan*

<sup>g</sup>*Advanced Telecommunications Research Computational Neuroscience Laboratories, Kyoto, Japan*

---

## Abstract

Since Tversky (1977) argued that similarity judgments violate the three metric axioms, asymmetrical similarity judgments have been particularly challenging for standard, geometric models of similarity, such as multidimensional scaling. According to Tversky (1977), asymmetrical similarity judgments are driven by differences in salience or extent of knowledge. However, the notion of salience has been difficult to operationalize, especially for perceptual stimuli for which there are no apparent differences in extent of knowledge. To investigate similarity judgments between perceptual stimuli, across three experiments we collected data where individuals would rate the similarity of a pair of temporally separated color patches. We identified several violations of symmetry in the empirical results, which the conventional multidimensional scaling model cannot readily capture. Pothos et al. (2013) proposed a quantum geometric model of similarity to account for Tversky's (1977) findings. In the present work, we extended this model to a more general framework that can be fit to similarity judgments. We fitted several variants of quantum and multidimensional scaling models to the behavioral data and concluded in favor of the quantum approach. Without further modifications of the model, the best-fit quantum model additionally predicted violations of the triangle inequality that we observed in the same data. Overall, by offering a different form of geometric representation, the quantum geometric framework of similarity provides a viable alternative to multidimensional scaling for modeling similarity judgments, while still allowing a convenient, spatial illustration of similarity.

*Keywords:* similarity, asymmetrical judgments, triangle inequality, multidimensional scaling, geometric model of similarity, quantum cognition

---

## 1. Introduction

### 1.1. Similarity

One of the most significant questions in cognitive theory concerns the way objects (e.g., ideas, concepts, stimuli) are represented. A hugely influential approach to representation concerns the assumption that e.g. objects are represented as points in a putative, internal metric space, with similarity computed as an exponentially decaying function of distance – the latter is Shepard's (1987) law of generalization. Following this approach, a matrix of confusability, or similarity ratings, can be translated to a spatial, geometric representation for a set of items. This point-wise, geometric representation is estimated through multidimensional scaling (MDS). MDS takes in a symmetric, distance matrix between pairs of objects in a group, and outputs a set of coordinates that specify a point for each object embedded in a low-dimensional space (Kruskal & Wish, 1978). This low-dimensional, spatial representation recreates the

---

\*Author for correspondence (gunnarepping@gmail.com)

<sup>1</sup>These authors contributed equally to this work

distances between the points that represent each object as closely as possible to the original distance matrix. Theoretically, Shepard's method pioneered a framework, and accompanying psychological theory, to simultaneously encode representation and similarity. Shepard's work has offered a variety of advantages including a spatially intuitive way to illustrate similarity and a simple technique to quantify relations between objects, which has been the basis for the formalization of several categorization models (Nosofsky, 1986; Reed, 1972).

As a consequence of Shepard's work, point-wise, geometric representations of objects in a metric space have become the default method to model similarity (Rosenthal et al. (2022) and Bonnardel et al. (2016) are examples of recent work employing MDS). Despite the advantages offered by this strategy, several fundamental constraints follow from using distance to encode similarity, which have been empirically challenged. The three axioms of a metric space are minimality, symmetry, and the triangle inequality. Therefore, examining the consistency of human similarity judgments with these axioms offers a general way to test the psychological accuracy (or generality) of representation approaches based on metric spaces. In one of the more famous works in the similarity literature, Tversky (1977) argued that all three of these axioms can be violated. For example, when focusing on the symmetry axiom, Tversky (1977) reported that the similarity between (North) Korea and (Red) China was often rated higher compared to the similarity between China and Korea; several other pairs of stimuli were employed and analogous results reported in the study which Tversky (1977) referred to. If one wants to employ a geometric representation to study similarity, then using a point-wise representation, such as those that are assumed in MDS, is a limited approach when at least one of these three axioms is violated. In addition to violations of the metric axioms, Tversky (1977) demonstrated a context effect on similarities, coined the diagnosticity effect. With its fixed representation, any distance-based representation fails to account for this context effect as well. Therefore, Tversky (1977) argued for an entirely new representation approach, featural models, as opposed to geometric models of similarity.

An example of a featural model is the contrast model, which Tversky (1977) devised to explain his findings that challenged the geometric models. In the contrast model, stimuli are represented as a set of features, and similarity is computed as a linear combination of common and distinctive features according to the following formula:  $Sim(A, B) = \theta \cdot (A \cap B) - \alpha \cdot f(A - B) - \beta \cdot f(B - A)$ , where  $A$  and  $B$  are the objects being compared,  $f(A \cap B)$  is the number of common features,  $f(A - B)$  is the number of features distinct to  $A$  ( $A$  has these features but  $B$  does not),  $f(B - A)$  is the number of features distinct to  $B$ , and  $\theta$ ,  $\alpha$ , and  $\beta$  are free parameters. Therefore, this model can produce asymmetries in similarity ratings whenever  $[\alpha - \beta] \cdot f(A - B) \neq [\alpha - \beta] \cdot f(B - A)$ .

### 1.2. Violations of symmetry in similarity

Although violations of the metric axioms and the diagnosticity effect all offer equivalent challenges to the distance-derived description of similarity, asymmetries in similarity judgments have attracted the most investigation. To explain this phenomenon, Tversky (1977) suggested that asymmetries are driven by prominence or extent of knowledge. For the China – Korea example, the contrast model predicts the similarities to be:

$$Sim(China, Korea) = \theta \cdot f(China \cap Korea) - \alpha \cdot f(China - Korea) - \beta \cdot f(Korea - China)$$

$$Sim(Korea, China) = \theta \cdot f(Korea \cap China) - \alpha \cdot f(Korea - China) - \beta \cdot f(China - Korea)$$

In this case, China would be regarded as more prominent compared to Korea according to the average observer. So, in the contrast model,  $f(China - Korea) > f(Korea - China)$ . Next, Tversky predicts that when a similarity task is directional, meaning that stimulus is the target while the other is a referent, the parameters for the distinctive features,  $\alpha$  and  $\beta$ , may differ. Since the similarity of  $A$  to  $B$  invokes one's knowledge of  $A$  in terms of  $B$ ,  $A$  is the subject of comparison and  $B$  is the referent. According to Tversky, since one naturally allocates more attention to the subject, the distinct features of the subject are weighted more heavily relative to the distinct features of the referent. Therefore, in the contrast model,  $\alpha > \beta$ , which Tversky (1977) called the focusing hypothesis. Given that China is more prominent than Korea and the relationship between  $\alpha > \beta$  is governed by the focusing hypothesis, the contrast model predicts that  $Sim(Korea, China) > Sim(China, Korea)$  which matches the empirical results found by Tversky (1977). Therefore, in a directional similarity task, the direction of asymmetry is determined by the relative extent of knowledge, or prominence, of the stimuli being compared. There are other similarity models which offer contrasting explanations (primarily) for asymmetries in similarity judgments and we will consider these briefly below.

Even though Tversky's (1977) ideas concerning similarity asymmetries were formulated in terms of differences in number of features, his model was expressed in terms of higher vs. lower prominence (more knowledge would

be equated with higher prominence). In this more general instantiation, the model appears to benefit from additional supporting evidence. As an example with perceptual stimuli instead of conceptual ones, [Rosch \(1975\)](#) reported asymmetries between colors such as focal red vs non-focal red, where focal red was defined as the perceptually salient prototype of the red color category. Also, [Polk et al. \(2002\)](#) found asymmetries between colors where one color was presented more frequently compared to the other color. In this study, the more frequently presented stimulus was identified as the more salient stimulus. In both cases, it appears that the similarity between a ‘less prominent’ item and a ‘more prominent’ item is higher than the converse similarity, consistent with [Tversky’s \(1977\)](#) theorizing.

[Tversky \(1977\)](#) formalized the idea of higher prominence for conceptual stimuli in terms of a higher number of known features for such stimuli – this is straightforward. However, he offered no analogous formalization for prominence concerning perceptual stimuli, for which there is no (obvious) difference in extent of knowledge. Therefore, using [Tversky’s \(1977\)](#) framework, it is difficult to understand apparent inconsistencies in the literature. For example, [Polk et al. \(2002\)](#) found that focal blue and green vs their respective, non-focal counterparts did not reproduce the asymmetries that were found with focal vs non-focal colors in [Rosch’s \(1975\)](#) experiment. In a study that investigated discriminability using faces as stimuli which varied in facial expression, [Roberson et al. \(2007\)](#) reported that good vs poor exemplar pairs were more discriminable compared to poor vs good exemplar pairs. Focal colors equate to good exemplars because they are considered the best example of that color. Since discriminability is inversely proportional to similarity ([Shepard, 1987](#)), this study appears to suggest that non-focal vs focal colors were more similar compared to focal vs non-focal colors. This finding corroborates the results of [Rosch’s \(1975\)](#) experiment, but not those with [Polk et al. \(2002\)](#). [Best and Goldstone \(2019\)](#) conducted a similar study, except that they employed faces varying in color, rather than facial expression. In two out of the three experiments conducted, Best and Goldstone reported that poor vs good exemplar pairs were more discriminable compared to good vs poor exemplar pairs. Clearly, this is the opposite of [Roberson et al.’s \(2007\)](#) results and therefore also conflicts with the findings of [Rosch’s \(1975\)](#) experiment. Note, these studies differ in that facial expression is evaluated more conceptually, whereas facial color is evaluated more perceptually. This indicates that the concept of salience or prominence, as formalized for conceptual stimuli, does not transfer well to the perceptual domain. In another study that investigated discriminability using pairs of stimuli that varied on shape and color (there were only three possible values for each stimulus dimension), [Hoggets and Hahn \(2012\)](#) predicted and reported asymmetries using transformational distance between perceptual stimuli, which does not correspond to any definition of prominence as ([Tversky, 1977](#)) theorized. The results of these various studies are summarized in [Table 1](#).

Table 1: Summary table of various study results on the direction of similarity asymmetry

Study	Stimuli	Independent Variable	Results	Results in terms of <a href="#">Rosch (1975)</a>
<a href="#">Rosch (1975)</a>	Colors	Focality	$\text{Sim}(\text{non-focal, focal}) > \text{Sim}(\text{focal, non-focal})$	$\text{Sim}(\text{non-focal, focal}) > \text{Sim}(\text{focal, non-focal})$
<a href="#">Polk et al. (2002)</a>	Colors	Presentation Frequency	$\text{Sim}(\text{low frequency, high frequency}) > \text{Sim}(\text{high frequency, low frequency})$	$\text{Sim}(\text{non-focal, focal}) > \text{Sim}(\text{focal, non-focal})$ and sometimes $\text{Sim}(\text{non-focal, focal}) < \text{Sim}(\text{focal, non-focal})$
<a href="#">Roberson et al. (2007)</a>	Faces	Facial Expression	$D(\text{good exemplar, poor exemplar}) > D(\text{poor exemplar, good exemplar})$	$\text{Sim}(\text{non-focal, focal}) > \text{Sim}(\text{focal, non-focal})$
<a href="#">Best and Goldstone (2019)</a>	Faces	Color	$D(\text{good exemplar, poor exemplar}) < D(\text{poor exemplar, good exemplar})$	$\text{Sim}(\text{non-focal, focal}) < \text{Sim}(\text{focal, non-focal})$
<a href="#">Hoggets and Hahn (2012)</a>	Pair of objects	Transformation Distance	$D(\text{long distance}) > D(\text{short distance})$	NA, there is no concept of focality as the values on each stimulus dimension are: Shape (Circle, Square, and Triangle) Color (Black, Gray, and White)

### 1.3. Geometric models of similarity

Overall, the notion of salience, as introduced by ([Tversky, 1977](#)), has proven difficult to apply to different sets of stimuli. Moreover, the relevant evidence points to several apparent inconsistencies, challenging our expectation that similarity asymmetries can be comprehensively understood using Tversky’s proposal ([Aguilar & Medin, 1999](#)). Note, when we talk about salience of perceptual stimuli, we refer to the relative salience of one stimulus relative to another, in a comparison, and not to other ways in which salience can be understood (e.g., in relation to the salience of a lone stimulus against the display background). A reasonable starting point for further examining similarity asymmetries would be using simple, perceptual stimuli, with no apparent differences in extent of knowledge. Color is a reasonable choice, because of the intrinsic interest in color processing in science, and because it has been part of the debate concerning asymmetries in similarity. This is the approach we adopt in the present work.

To reconcile the metric space violations and diagnosticity effect with the powerful and intuitive framework offered by the point-wise, geometric modeling approach, several theories have been put forth ([Ashby & Perrin, 1988](#); [Krumhansl, 1978](#); [Nosofsky, 1991](#)). These theories differ from [Tversky’s \(1977\)](#) primarily in terms of not requiring

lists of discrete features. [Krumhansl \(1978\)](#) proposed a distance-density model that computes similarity according to both the distance between two points and the local density around each point. The density is defined as the number of items within a certain radius of an object's position in psychological space. [Krumhansl \(1978\)](#) suggested that prominent objects are likely to have many features and thus these objects are likely to share features with a greater number of other objects as compared to objects with fewer features. Therefore, prominent objects are more likely to exist in denser regions of psychological space. In the China-Korea example used by [Tversky \(1977\)](#), China, associated with more knowledge, would have a higher local density relative to Korea, and this difference would drive the order effect in asymmetry. This logic cannot be extended straightforwardly to perceptual stimuli, for which it is hard to motivate differences in the density of psychological space (for pairs of colors which are both either focal or non-focal).

Continuing with the point-wise approach, [Ashby and Perrin \(1988\)](#) developed general recognition theory, where each stimulus corresponds to a point in a psychological space along with an accompanying perceptual distribution. The region in psychological space spanned by the distribution, where the density of the distribution is above some recognition threshold value, is a stimulus' corresponding optimal response region. Using the notion of response regions, similarity is defined as the proportion of the first stimulus' perceptual distribution that overlaps with the second stimulus' optimal response region. Less prominent concepts, such as Korea, have a less concentrated, or flatter probability distribution in psychological space because less knowledge implies greater variability in perception. Since the perceptual distribution of Korea is more diffuse in the psychological space, fewer points exceed the optimal recognition threshold and therefore the optimal response region of Korea would be smaller compared to China. Also because of this higher variability, Korea's perceptual distribution will span a larger region of psychological space, increasing its overlap with other concepts' optimal response regions such as China (note, the implication here is the opposite to that from [Krumhansl's \(1978\)](#) model). From this, the similarity asymmetries are driven by the variability of the distributions that represent each stimulus in the metric space. Concerning perceptual stimuli, such an explanation could be adapted by assuming that less prominent e.g. colors are associated with higher variability. So, [Ashby and Perrin's \(1988\)](#) model can provide an approach to similarity asymmetries, of a form analogous to that of [Tversky \(1977\)](#) and [Rosch \(1975\)](#), which encompasses both conceptual and perceptual stimuli.

Finally, [Nosofsky \(1991\)](#) added individual-bias parameters to the original MDS approach to compute similarity. The individual-bias parameter was included as a free parameter, so it could account for asymmetries in either direction depending upon the individual and pair of objects being compared. The inclusion of the extra parameter allows for a direct way to reconcile asymmetries with metric models. Besides the three approaches outlined here, several other models can capture similarity asymmetries (for a review of these, see [Goldstone and Son \(2005\)](#)).

Similarity theory has not been confined to asymmetries, of course. Even just restricted to [Tversky's \(1977\)](#) influential work, together with asymmetries, an aspiring model would have to cover violations of minimality, violations of the triangle inequality, and the diagnosticity effect. In addition, there are several other important insights in similarity, which require adequate explanation, notably the finding that similarity judgments often reflect the structure of the compared objects ([Bowdle & Gentner, 1997](#); [Goldstone, 1994](#)). The formalisms considered above, including [Tversky's \(1977\)](#), represent some of the most influential ideas in the last few decades, but they also suffer from limitations that prevent (fairly) seamless coverage across the relevant body of evidence.

#### *1.4. Original quantum model of similarity*

When a field of study is bending over backwards to salvage a theory in light of conflicting evidence, it may be appropriate to approach the phenomena from a different theoretical perspective. For example, in physics, this line of thought has led to the development of quantum mechanics and general relativity, where novel mathematical techniques were ushered in and revolutionized the field. In behavioral sciences, the preponderance of order, interference, or context effects for some aspects of cognition has suggested the use of quantum theory as a modeling framework ([Busemeyer & Wang, 2015](#); [Pothos & Busemeyer, 2022](#)).

In traditional theories of probability, measurement outcomes are represented as subsets of a universal set, and each measurement outcome has an associated probability. In quantum theory, measurement outcomes are represented as subspaces of a vector space, specifically a Hilbert space, and the probability for a given measurement outcome is computed by projecting the state of the system onto the corresponding subspace. Using this representation, quantum probabilities are computed through projection. When a projection occurs, the state is 'laid down' onto the subspace that it is projected onto, and the resulting length of that projection can be converted to a probability by squaring its magnitude. After a projection, the state of the system is now the renormalized, projected state, rather than the initial

state before the projection. Since projection often alters the state of the system, sequential projections can yield different results, depending on the order, because the first projection establishes the context for the subsequent projection. This noncommutativity of projections allows the quantum model to naturally predict order effects in measurements, whereas traditional models typically require some augmentation to describe these effects. Generally, transitioning from subsets to subspaces offers an alternate approach to geometric modeling in similarity judgments and to probabilistic inference. In the latter case, order and context effects have been more reliably demonstrated, so most work with quantum models so far has focused on such applications. For example, recent work has demonstrated that quantum theory can sometimes provide more accurate models of cognition than those based on classical classical theory. The application includes attitudes (Wang et al., 2014; White et al., 2014), causal reasoning (Trueblood & Busemeyer, 2012), concept combination (Aerts & Gabora, 2005), decision making (Lambert-Mogiliansky et al., 2009; Pothos & Busemeyer, 2009; Busemeyer et al., 2006, 2009; Lambert-Mogiliansky & Danilov, 2010), memory (Bruza et al., 2009; Brainerd et al., 2013), perception (Atmanspacher et al., 2004; Conte et al., 2009), and probability judgments (Busemeyer et al., 2011; Trueblood & Busemeyer, 2011; Asano et al., 2012).

Since quantum theory inherently gives rise to order and context effects, this approach may provide a consistent description of the order and context effects that have been found in similarity judgments. Additionally, quantum theory offers a natural method to incorporate differences in extent of knowledge (including features) in a geometric manner, by associating them with differences in subspace dimensionality. In Pothos et al. (2013), researchers proposed a quantum geometric model of similarity that provides a parsimonious account of all three violations of metric space axioms and the diagnosticity effect under a coherent framework. This model assumes that comparison is an inherently sequential process, comparing one thing to another. The quantum similarity model represents percepts as subspaces of a vector space as opposed to points in a metric space and posits that similarity is computed using sequential projections onto subspaces as opposed to distance between points. The process of evaluating the similarity of A to B is one of projecting to the subspace for A followed by the subspace for B (i.e., similarity is sequential projection). Figure 1 illustrates the sequential projection used in the quantum similarity model and how different orders of projections can produce different results. In the example of China and Korea, the subspace that represented China was assumed to have a greater dimensionality compared to the subspace that represented Korea, and this difference drove the asymmetries found in similarity judgments. By representing features of an object using dimensions of a subspace, and taking advantage of the other properties of quantum processing, this geometric representation accounted for many of the puzzling findings of similarity judgments and bridged the gap between traditional geometric models (Shepard, 1987), featural models (Tversky, 1977), and alignment models (Goldstone, 1994; Pothos & Trueblood, 2015).

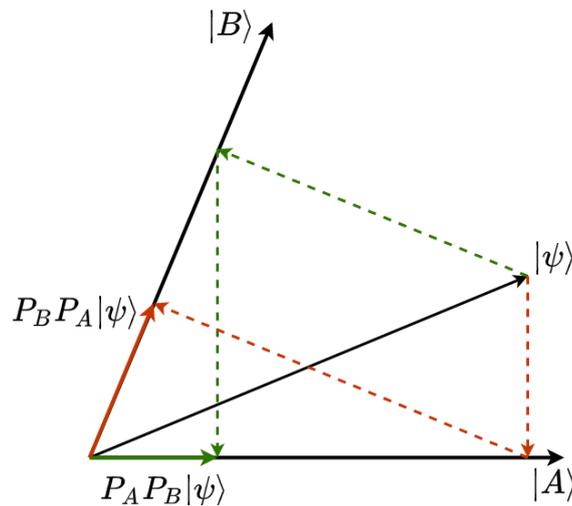


Figure 1: Example of projecting the state onto subspace A then B ( $P_B P_A |\psi\rangle$ ) and projecting the state onto subspace B then A ( $P_A P_B |\psi\rangle$ ), in the case where both subspaces are one-dimensional.

The sequential projection process, which is assumed when comparing the similarity between A and B, is essen-

tially a process of thinking about A first and then about B. So, in a sense, similarity is the ease of thinking first about B, once a thought about A has been established (Hahn et al., 2003). This exact process has been postulated for decision making too, when people estimate the probability of conjunctions (Tversky & Kahneman, 1983; Busemeyer et al., 2011). So, the quantum similarity model views similarity and probabilistic inference as based on the same fundamental process.

Although the quantum geometric model successfully reproduced asymmetries between conceptual stimuli by relating dimensionality to extent of knowledge, it has not been evaluated on similarity judgments between perceptual stimuli. As mentioned above, extent of knowledge, as a driver of asymmetries, is of arguable relevance when transitioning from conceptual stimuli to perceptual stimuli. In this work, we extend the quantum geometric model to a general framework that describes similarity judgments of perceptual stimuli using similarity ratings between colors and directly compare it to the MDS approach. Since extent of knowledge cannot be attributed to perceptual stimuli such as colors, we use the order of projection onto subspaces, rather than subspace dimensionality, to model asymmetries in similarity judgments. Additionally, the quantum framework developed here allows us to implement a variety of hypothesized mechanisms that are thought to play a role in generating similarity judgments. By comparing the fits of the quantum model variants with different combinations of the various mechanisms, we can evaluate whether there is evidence supporting each hypothesis.

Generally, geometric models are ones that represent percepts in a multidimensional psychological space. Table 2 summarizes how the quantum framework for modeling similarity judgments compares to the previously mentioned models in terms of: (1) stimulus representation, (2) similarity computation, (3) production of asymmetries, (4) encoding of extent of knowledge, and (5) encoding of attention-grabbing properties. Although Table 2 covers the predominant proposals for the source of asymmetries in the literature, these proposals do not exhaust relevant ideas in perceptual processing. As we demonstrate later (Section 4.4.1), the use of the quantum framework allows us to explore some additional possibilities for similarity asymmetries, relating to biases in serial processing and contextual influences from previous trials. We hope to show how the quantum framework offers a form of geometric representation that is different to that of MDS. The different properties between quantum and MDS representations underwrite the differing perspectives about human similarity judgments.

Table 2: Overview of how the various approaches model similarity judgments

Model/Framework	Stimulus Representation	Similarity Computation	Production of Asymmetrical Judgments	Encoding Extent of Knowledge	Encoding Attention-grabbing Properties
Quantum	Subspace in a vector space	Sequential projection	Non-commutativity of projectors	Subspace dimensionality	Positioning of initial state
Augmented MDS	Point in a multidimensional space	Inversely proportional to the distance between two points	Bias and directionality parameters	NA	NA
Contrast model	Set of features	Increases with the number of shared features and decreases with the number of features that are not shared	Features of the target stimulus are weighted more heavily than features of the referent stimulus	Number of features	Weighting of the target stimulus versus the referent stimulus
Distance-density model	Point in a multidimensional space	Inversely proportional to the distance between two points, in a way that is affected by local density	Density around the target stimulus is weighted more heavily than the density around the referent stimulus	Local density	Weighting of the target stimulus versus the referent stimulus
General recognition theory	Probability distribution in a multidimensional space	Overlap of the referent's distribution with the target's optimal response region	Either differences in the size of the stimulus optimal response regions or differences in the variability of the stimulus distributions	Variability of the stimulus distribution	Variability of the stimulus distribution

## 2. Methods

### 2.1. Participants

45 participants (15 for each experiment) were recruited with the online platform Prolific (www.prolific.co). Prolific users with a 100% past approval rate were selected to help ensure participants would be completing the survey in

a reasonably conscientious way. Participants were incentivized with a monetary reward of £5.75/hr for their participation.

## 2.2. Design and overview

The experiment was a similarity task where participants were asked to evaluate the similarity of one color to another color. The objective of the experiment was to investigate whether the order of presentation affected the resulting similarity judgment between colors, thereby producing similarity asymmetries. The experiment was created in PsychoPy (Pierce et al., 2019) and uploaded to Pavlovia (Pavlovia.org) to be run online. The structure of a typical trial is shown in Figure 2.

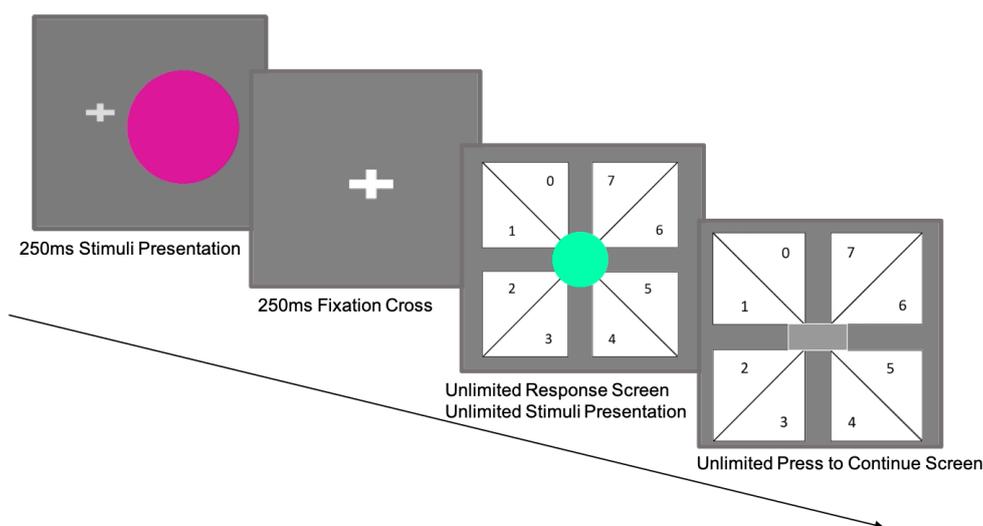


Figure 2: Time course of the stimulus presentation screens and response screens upon button press from the previous trial. First, a foveal color stimulus is displayed for 250ms together with the fixation cross. Foveal stimuli were presented on the right of the fixation cross and peripheral stimuli were presented on the left, with a random polar angle. This was followed by the fixation cross for an additional 250ms (no masking). The second stimulus is shown in the center of the response screen and is displayed for an unlimited amount of time until a response is selected in Experiments 1a and 1b, but only shown briefly (250 ms) in Experiment 1c. The final screen is shown for an unlimited amount of time until the grey rectangle is selected. Clicking the gray rectangle initiated the next trial. NB: Objects in this figure are exaggerated in size for illustrative purposes relative to what subjects were presented. A video of an example trial can be found at <https://youtu.be/1Bt20x1aUuc>.

We ran four experiments, which varied some aspects of the procedure. In Experiment 1a, we had 344 trials. There were 9 colors used, resulting in 81 ( $9 \cdot 9$ ) color sequences. For each of the 81 sequences, the first stimulus in the sequence was displayed once in the periphery and once in the fovea during one block. The fovea stimulus had a radius of 1 degree of visual angle (DVA) and was positioned 1 DVA away from the fixation cross. The periphery stimulus was larger, with a radius of 4.5 DVA and was positioned 10 DVA away from the fixation cross. The radii of the color patches for both fovea and periphery stimuli was chosen based on the size to eccentricity ratio in Freeman and Simoncelli (2011). For a separate research question (not addressed here), there was interest in whether similarity ratings would be impacted by whether the stimuli were presented in the fovea or in the periphery and the size of the color patches. This resulted in 162 ( $81 \cdot 2$ ) unique non-catch trials per block. These 162 trials were completed twice with 20 catch trials, and therefore Experiment 1a consisted of a total of 344 trials ( $162 \cdot 2 + 20$ ). Experiments 1b, 1c, and 2 were replications of Experiment 1a with minor modifications. Below, we describe Experiment 1a and subsequently present Experiments 1b, 1c, and 2.

## 2.3. Materials

The stimuli used in this experiment were color patches. The color patches varied on three dimensions: color, position, and size. In this report, position and size are treated as incidental variables; we note that preliminary analyses indicated non-interaction between these factors and similarity asymmetries. Additionally, these factors informed, in

part, some of the models we tested. This further confirmed that position together with size did not influence similarity judgments.

In our analysis we will define focal colors as colors whose RGB codes contain one non-zero value and that non-zero value is the maximum, 255. For example, red is a focal color because it has RGB code [255, 0, 0]. Therefore, red, green, and blue are defined as focal colors. Non-focal colors will be the rest of the colors used in the experiment. The 9 colors used in this experiment are displayed in Table 3. In HSV space, these 9 colors are equally spaced in hue with saturation and value kept constant. The selection of colors was essentially a random sample, but with our (intuitive) bias of selections which might be more conducive to asymmetries. That is, we sought out stimuli we thought might challenge a symmetry constraint in similarity, since such stimuli offer a more interesting test of whether the additional mechanisms in the quantum approach are needed versus the established framework of MDS. Nevertheless, for the purpose of modeling, the choice of colors used in this study is inconsequential. These stimuli were displayed on a background corresponding to [0, 0, 0.5] in HSV space.

Table 3: Nine colors that were used in the experiment presented with their name and hex code

Color	Name	Hex Code	Color	Name	Hex Code
	Red	#FF0000		Deep Sky Blue	#00A9FF
	Orange	#FFAA00		Blue	#0000FF
	Spring Bud	#AFFF00		Electric Purple	#AA00FF
	Green	#00FF00		Hollywood Cerise	#FF00AA
	Medium Spring Green	#00FFA9			

The first stimulus in the color sequence could either be presented in the fovea position or the peripheral position. The periphery and fovea positions were diametrically opposite, and the polar angle of the stimuli was randomized between -30 degrees to 30 degrees from the horizontal axis for each trial. The fovea color patch was always displayed on the right side of the screen and the peripheral color patch was always displayed on the left side of the screen.

#### 2.4. Calibration

The experiment began with an online viewing distance calibration, which was adapted from the Virtual Chinrest (Li et al., 2020). The calibration task provided us with an estimate of screen width, viewing distance, and screen resolution for each participant. Screen resolution and viewing distance were used to control the size and location of the stimuli presented to participants. Therefore, we wanted participants to complete the experiment on a desktop or laptop, rather than a mobile phone or tablet, so that the stimulus size and location would be as uniform as possible. If the screen width was less than 230 mm or the viewing distance was less than 300 mm, we terminated the experiment at this point, as these results would suggest that participants were attempting to complete the experiment on a mobile phone or tablet. For details, see Zelenikow-Johnston et al. (under review) and Fisher et al. (in prep).

#### 2.5. Procedure

##### 2.5.1. Experiment 1a

Before starting the experiment trials, participants were given instructions for the task. Participants were instructed to focus on the center cross on the screen while a circle would flash out at the periphery. The 9 colors used in the experiment were displayed simultaneously, so participants could see the range of colors and consider how they would assign similarity ratings. Participants were then informed that they “will need to decide the similarity levels of the previous circle to a circle on the next screen. Ignore any size differences.” They were told to rate the similarity using the range of ratings from 0 to 7, by clicking one of the triangles in the response screen (Figure 2). 0 indicated the most similar color (least different color) and 7 indicated the most different color (least similar color). Participants were informed that after they make a rating, “move your cursor back to the centre and click and hold the box to continue”. The instructions also included a description of the catch trials where participants would be asked to click

on a particular value. All the instructions were accompanied with a video of the task. Participants then completed 7 practice trials with one additional catch trial. For these practice trials, they received feedback, for example “You selected 7, 7 indicates you found the colors very different.”

For a non-catch trial, the first color was displayed for 250 ms in the periphery or the fovea along with a center cross. The center cross remained for 250 ms after the first color was removed. The second color was always displayed in the center of the screen and remained on the screen until the participant selected their response. After the participant had selected their similarity rating, they had to click a rectangle in the middle of the response screen to proceed to the next trial. The procedure for non-catch trials is shown in Figure 2.

The 344 trials were presented in two blocks, where each block consisted of 162 non-catch trials and 10 catch trials. We treated them as two blocks but participants performed the task as a single continuous block of 344 trials. In each block, the trials were repeated in the exact same order and location. However, in the second block, the polar angle used was randomized in each trial. Also, temporal location of the catch trials and the number participants were asked to select was randomized in each block. Participants were not informed of the repeated ordering across the two blocks. After completing the 344 trials, participants answered a questionnaire concerning any vision impairments they might have. This concludes the description of Experiment 1a.

### 2.5.2. Experiment 1b

Experiment 1b was a replication of Experiment 1a where we randomized the order of trials in the second block. Also, the temporal location of the catch trials was the same in both blocks. All other elements in Experiment 1b were the same as Experiment 1a.

### 2.5.3. Experiment 1c

In the final replication of Experiment 1a, we examined the effect of unlimited viewing time for the second stimulus. To accomplish this, we limited the duration of the presentation of the second color patch to 250 ms. Also, the temporal location of the catch trials was the same in both blocks. All other elements in Experiment 1c were the same as Experiment 1a.

### 2.5.4. Experiment 2

Experiment 2 is a brief experiment with 50 participants. Participants were recruited through CloudResearch, a virtual wrapper for Amazon’s Mechanical Turk (MTurk). Only English native speakers from Australia, Canada, India, the the United States of America were recruited. Also, participants were screened so that they had to have completed at least 1000 unrelated MTurk tasks with at least 97% approval to ensure quality data. Participants were incentivized with a monetary reward of \$ 6.00 / hour.

Experiment 2 consisted of 253 trials, with 81 trials where both color patches were located in fovea positions, 81 trials where both color patches were located in periphery positions, 81 trials where one color patch was located in the fovea position and the other color patch was located in the periphery position, and 10 catch trials. The catch trials were identical to the ones in Experiment 1.

After completing the same calibration task as in Experiment 1, participants were prompted to rate the similarity between two colors that were presented simultaneously in Experiment 2, using the same nine colors and 0-7 rating scale as Experiment 1. The procedure of each trial in Experiment 2 differs from Experiment 1 in exactly one way. Rather than the two color patches being temporally separated as in Experiment 1, the two color patches are spatially separated in Experiment 2.

Experiment 2 had 4 different conditions: FIX SMALL, FIX BIG, MAG SMALL, and MAG BIG. In FIX conditions, stimuli presented in the fovea and the periphery were presented with the same diameter. In MAG conditions, stimuli presented in the periphery were magnified relative to the stimuli presented in the fovea. In SMALL conditions, the stimuli were presented with a relatively small diameter. In BIG conditions, stimuli were presented with a relatively large diameter.

We use the results from Experiment 2 to generate the underlying color spaces as a group baseline. Upon that, we will test the prediction of two classes of models on the results of independent experiments (Experiment 1). Results from only the BIG conditions were employed in the construction of the multidimensional spaces. The other conditions were included for unrelated research.

### 3. Results

#### 3.1. Exclusions

In Experiment 1a, 15 participants completed the online study (9 males, mean age 25). Participants who either had a catch score lower than 80% or a mean response time of 700 ms or less were excluded from the study. After the exclusion criteria, 14 participants remained.

In Experiment 1b, 15 participants completed the online study (11 males, mean age 23). Experiment 1b had the same exclusion criteria as Experiment 1a. After the exclusion criteria, 14 participants remained.

In Experiment 1c, 15 participants completed the online study (8 females, mean age 25). Experiment 1c had the same exclusion criteria as Experiment 1a. After the exclusion criteria, 12 participants remained.

After all analyses were completed, a minor error was discovered in the experimental code, which meant that for some of the catch trials some participants may have been inappropriately excluded. This affects the one participant excluded in Experiment 1b and the three participants excluded in Experiment 1c. We decided to retain these exclusions because, for these participants, the coding error in the catch trials means that there is no sufficient data to verify they were being attentive.

#### 3.2. Calibration results

All participants across experiments had a successful screen width of 230 mm or greater and a successful viewing distance of 300 mm or greater.

#### 3.3. Reliability check

To see how reliable participants are in their similarity ratings, we calculated the correlation between the ratings in the first and second block. Experiment 1b served as a control experiment to test whether the order of the trials influenced the reliability of participants' ratings. Similarly, Experiment 1c served as a control experiment to test whether the presentation time of the second stimulus influenced participants' ratings. We calculated the Spearman's rank order correlation coefficient ( $\rho$ ) for the similarity ratings for the first block and the second block for each participant in each experiment. Figure 3 shows a box plot of the  $\rho$  values grouped by experiment, where the individual points correspond to participant  $\rho$  values. As shown in Figure 3, the first and second blocks in Experiment 1a (median  $\rho = 0.891$ ) and in Experiment 1c (median  $\rho = 0.911$ ) were highly correlated. The first and second blocks in Experiment 1b appear less correlated, as the median  $\rho$  was lower (0.833). Although the correlation was lower for Experiment 1b, it is still high and, overall, we conclude that participants were consistent across repetitions of similarity judgments. Since the two manipulations had no discernible behavioral impact, we pooled the participants into one group when displaying the dissimilarity matrix, asymmetry matrix, and variance matrix (Figures 4a, 4b, and 4c, respectively) and when fitting the model. But, when computing which asymmetries were significant, we corrected for multiple comparisons within each experiment (corresponding results are displayed in Figure 4d and Table 4).

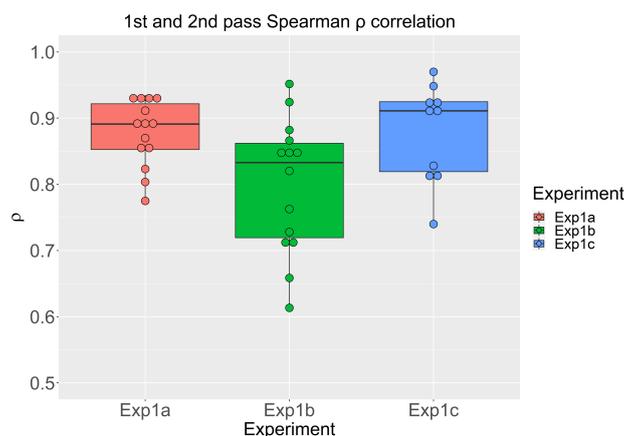


Figure 3: 1st and 2nd block Spearman's rank order correlation for each experiment. Plotted is box plots of the corresponding correlation values. Individual points correspond to participant correlation values. The bold black line corresponds to the median correlation value for that experiment group. Experiment 1a (N = 14); Experiment 1b (N = 14); Experiment 1c (N = 12).

To visualize the similarity ratings for each sequence of colors, we computed mean similarity ratings for each sequence, across all participants, which allowed us to generate the dissimilarity matrix, shown in Figure 4a. See Supplementary Figures 1a-1c and 2a-2c for individual participant dissimilarity matrices and mean dissimilarity matrices for each experiment. The first color in the sequence is portrayed on the x-axis and the second color in the sequence is portrayed on the y-axis. Cell colors represent how similar each sequence was rated, so that lighter colors indicate higher similarity. The dissimilarity matrix shown in Figure 4a appears asymmetrical for many entries across the light diagonal axis. To further visualize the asymmetries, we computed an asymmetry matrix. We calculated the asymmetry matrix by subtracting the transpose of the dissimilarity matrix from the original dissimilarity matrix. Figure 4b shows the asymmetry matrix, where the color of the cells represents the direction of asymmetry. Figure 4c illustrates the between participant response variances in mean asymmetry across all participants, where darker green indicates more variance and lighter green indicates less variance. To test if asymmetries are statistically significant for different color combinations, we performed a linear mixed-effect analysis of the effect of order on similarity. We identified six color pairs that were affected by order, when uncorrected for multiple comparisons. There were two such pairs that were affected by order, when employing a correction for multiple comparisons ( $q < 0.05$  FDR).

Note, the color pairings for which we observe significant asymmetries do not have the highest between participant response variances in asymmetry, and so any asymmetries are less likely to be due to noise. To summarize the asymmetry across experiments, Table 4 lists the color pairs with significant asymmetries.

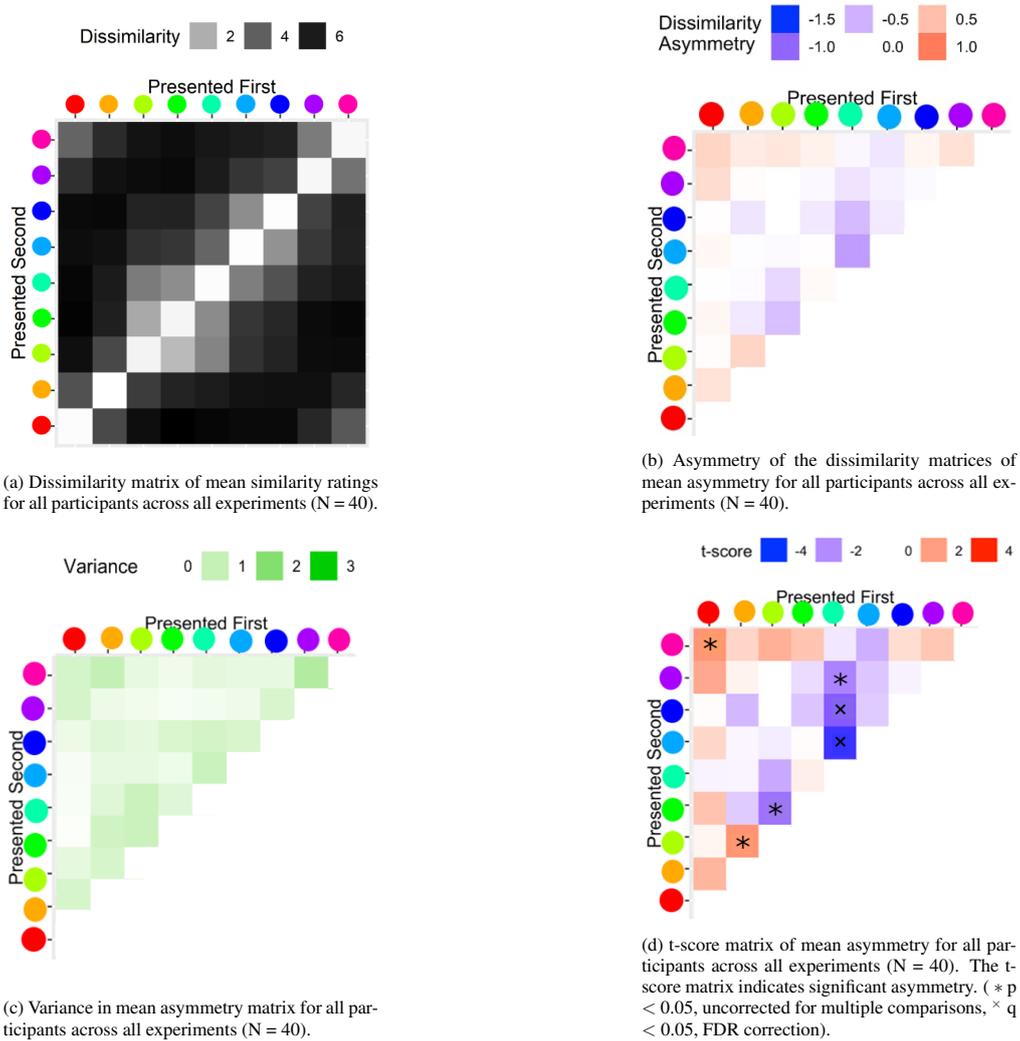


Figure 4: Mean dissimilarity, asymmetry, variance in asymmetry, and t-score matrices across all participant (N = 40).

Table 4: Color pairs with significant asymmetries (\*  $p < 0.05$ , uncorrected for multiple comparisons,  $\times$   $p < 0.05$ , FDR correction)

Color Pair Direction of Asymmetry (More similar order)	Effect Size (Asymmetry)
Red Magenta*	0.33
Orange Yellow*	0.33
Green Yellow*	0.41
Purple Cyan*	0.18
Blue Cyan $\times$	0.64
Blue Cyan $\times$	0.44

### 3.4. Driving factors of asymmetry

There are a number of well-motivated possibilities regarding what might be driving asymmetries. As a preliminary step in our attempt to predict the direction of asymmetries, we tested two different predictors of salience that align with Tversky’s concept of prominence (1977). Note, we do not declare each color’s salience as being equal to the output of these predictors, as the concept of prominence does not transfer well to perceptual stimuli. But, for the sake of modeling, we are going to treat each color’s salience as the output of these predictors. Both predictors use RGB code to estimate salience. The first predictor equates each color’s salience to the maximum value of each of the three RGB elements, divided by the sum of the elements. Therefore, a color such as red with RGB code [255,0,0], has a salience of 1 since  $\max([255, 0, 0]) / \text{sum}(255 + 0 + 0) = 255 / 255 = 1$ . The second predictor contrasts the RGB code of each color with the RGB code of the background, which was [128,128,128]. This method sums together the absolute value of the differences between each color’s RGB element and the background. After computing each color’s salience, the different salience values would be normalized onto a scale of 0 to 1 by dividing each of them by the maximum value. For example, red would have salience 1 because  $|255 - 128| + |0 - 128| + |0 - 128| = 127 + 128 + 128 = 383$  and 383 is the maximum salience value, so  $383 / 383 = 1$ . Table 5 displays the salience value for each color using each of the methods.

Table 5: Salience values for each color based on both predictors

Color	RGB Code	Predictor 1 Salience	Predictor 2 Salience
Red	[255, 0, 0]	1	1
Orange	[255, 170, 0]	0.600	0.776
Spring Bud	[170, 255, 0]	0.600	0.776
Green	[ 0, 255, 0]	1	1
Medium Spring Green	[ 0, 255, 169]	0.601	0.773
Deep Sky Blue	[ 0, 169, 255]	0.601	0.773
Blue	[ 0, 0, 255]	1	1
Electric Purple	[ 170, 0, 255]	0.600	0.776
Hollywood Cerise	[ 255, 0, 170]	0.600	0.776

It is interesting to examine whether these predictors could directly predict asymmetries, without any further elaboration (that is without employing these predictors in a more elaborate model of similarity asymmetry). To test the accuracy of these predictors, we computed the correlation between each of the 36 mean asymmetries between color pairs (excluding color pairs of the same color) and the asymmetries predicted by each of the salience predictors. The asymmetry for each color pair was computed as the difference between the ratio of color 1 to color 2 and the ratio of color 2 to color 1. The results are shown in Table 6.

Table 6:  $R^2$  and correlation coefficient ( $\rho$ ) for both predictors

Predictor	$R^2$	$\rho$	p-value from $\rho$
1	0.2500	0.5000	0.001901
2	0.2501	0.5001	0.001897

While the correlations are significant, 75% of the variance in asymmetries is unexplained by either of the predictors. Since existing ideas about the origins of asymmetries are not quite sufficient, our analysis of similarity asymmetries calls for more sophisticated modeling.

## 4. Modeling

### 4.1. Preliminary mathematics

Quantum theory models a system, e.g. a participant making similarity ratings, using an  $N$ -dimensional vector space (specifically a Hilbert space), where the state of the system is represented using an  $N$ -dimensional state vector

$|\psi\rangle$ . To compute the probability of some event (or concept)  $A$ , the state of the system is projected onto the subspace that represents event  $A$ . The projection of the state is computed with a projection operator, or simply a projector, which can be found by taking the outer product of the event's subspace with itself  $P_A = |A\rangle\langle A|$ , where  $|A\rangle$  is a column vector and  $\langle A|$  is the conjugate transpose of this column vector. According to quantum probability theory, the probability of event  $A$  is found by taking the squared length of the projected state  $p(A) = \|P_A \cdot |\psi\rangle\|^2 = \||A\rangle\langle A| \cdot |\psi\rangle\|^2 = |\langle A|\psi\rangle|^2$ , where  $\langle A|\psi\rangle$  is the inner product between the state and the subspace that represents event  $A$ . These quantum probabilities can be thought of as a measure of overlap between the state vector and a subspace. Psychologically, they are often interpreted as the probability that the participant is thinking about the concepts represented by the corresponding subspaces.

Using this formulation, the probability of two, sequential events is found by projecting the state onto the subspace that represents the first event, and then projecting the state onto the subspace that represents the second event. For example, the probability of event  $A$  and then event  $B$  is given by

$$p(A \longrightarrow B) = \|P_B \cdot P_A \cdot |\psi\rangle\|^2 = \||B\rangle\langle B| \cdot |A\rangle\langle A| \cdot |\psi\rangle\|^2 = |\langle B|A\rangle|^2 \cdot |\langle A|\psi\rangle|^2$$

Similarly, the probability of event  $B$  and then event  $A$  is given by

$$p(B \longrightarrow A) = \|P_A \cdot P_B \cdot |\psi\rangle\|^2 = \||A\rangle\langle A| \cdot |B\rangle\langle B| \cdot |\psi\rangle\|^2 = |\langle A|B\rangle|^2 \cdot |\langle B|\psi\rangle|^2$$

Since matrix algebra is not commutative,  $P_B \cdot P_A$  is not necessarily equivalent to  $P_A \cdot P_B$ . Such order effects, violating commutativity in this kind of sequential conjunction, are a hallmark of the way quantum probability differs from classical probability.

Perhaps the clearest way to investigate order effects in quantum theory is to consider how the first projection sets the context for the second projection. For example, consider the probability of event  $A$  and then event  $B$   $p(A \longrightarrow B) = p(B|A) \cdot p(A)$ , where  $p(B|A)$  is the probability of event  $B$  given event  $A$  has occurred. In quantum probability theory,  $p(B|A) = P_B \cdot |\psi_A\rangle$ , where  $|\psi_A\rangle$  is the renormalized state of the system conditioned on event  $A$ . According to Lüder's rule,  $|\psi_A\rangle = P_A \cdot |\psi\rangle / p(A)$ . By employing a conditioned state, we can compute the corresponding conditional probabilities just as in classical probability theory. But in quantum probability theory, the context of the conditioned state is set by the first event, whereas the state is unaffected in classical theory. Order effects arise in quantum theory due to the differences in how context is set by the first event.

#### 4.2. General ideas for modeling similarity

What is the explanatory level intended by quantum cognitive models? An influential framework for resolving such questions is the one from Marr's viewpoint: Marr proposed three levels of explanation. Those presently relevant are the most abstract level, the computational level, and the algorithmic level, which is intended as an intermediate between the computational level and the implementation level (the latter is about neurological/biological processes and is not relevant here). The computational level concerns the "what" and the "why" questions for the system that is studied, that is "what is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?" (Marr, 2010, p.25). By contrast, the algorithmic level concerns a process explanation of the studied system, specifically the representations that are employed by the system and the algorithms that operate on the representations to produce an output from an input. In other words, the algorithmic level is about the "how" question regarding the studied system. Concerning quantum cognitive models, we follow (Griffiths et al., 2010, p.357), who provided a well-thought generalization to Marr's 2010 analysis by noting that "probabilistic models of cognition pursue a top-down or 'function-first' strategy, beginning with abstract principles that allow agents to solve problems posed by the world - the functions that minds perform - and then attempting to reduce these principles to psychological and neural processes." Griffiths et al. (2010) further explain the intention of a probabilistic cognitive model with a range of questions, such as in relation to the information needed, the necessary representations, and the constraints on the computation. While some of these questions are computational-level explanations, it is also clear that some of them are algorithmic-level ones. Therefore, concerning the present similarity models, we would say that they are top-down (in the sense of Griffiths et al., 2010), with parts across both the computational and algorithmic levels (the latter because quantum theory is constrained in terms of how computations are carried out, e.g., involving order effects). This is because the present similarity models aim to describe human representations of stimuli and the

process by which similarity judgments are produced. We think these points are analogous to how the MDS similarity models should be interpreted.

To model similarity judgments, we employed two different frameworks: MDS and quantum theory. As mentioned in the introduction, the MDS approach represents colors as points in a metric space. The distance between these points is converted to a similarity rating using an exponentially decreasing function (Shepard, 1987). After converting the distances to similarities, the similarity values are linearly mapped onto the 0-7 ratings scale used in Experiments 1 and 2. Since distance is symmetric between any two points, the MDS model does not predict asymmetries in similarity ratings.

The quantum approach represents colors as subspaces of a vector space. In this application, the subspaces that represent each color are restricted to one-dimensional subspaces, meaning that they are rays. We chose to restrict the subspace dimensionality to one because in previous work subspace dimensionality has been associated with different features of a stimulus (Pothos & Trueblood, 2015) and also extent of knowledge (Pothos et al., 2013). Since the color patches used in our experiments are simple, perceptual stimuli that have no differences in extent of knowledge and similarity is only based on one aspect of the stimuli (color), we believe that each color should be represented using a one-dimensional subspace.

In the quantum framework, the state of a system is represented by a vector  $|\psi\rangle$ . The state vector represents the cognitive state of an individual, so in the present case this would be the cognitive state prior to engaging with a similarity judgment. To set the initial state at the beginning of each task, we assume that, after a person perceives the two colors on each trial, but, prior to the similarity judgment, the mental state reflects a combination of the presented stimuli in the current trial and the final state in the previous trial. Such a state is constructed as a superposition of the vectors corresponding to the two colors in the current trial and the final state in the previous trial (also a vector). Note, there is no analog to this state in the MDS approach. Rather than using distances to produce similarity ratings, the quantum framework uses projection, where the state is sequentially projected onto the subspace that represents the first color in the sequence and then onto the subspace that represents the second color in the sequence. After undergoing the sequential projections, the quantum framework generates a probability, which corresponds to the squared length of the projected state vector.

Psychologically, this is how the technical details of quantum theory map onto a proposal for similarity judgments: if we project the state  $|\psi\rangle$  onto a subspace, the corresponding probability can be interpreted as the degree to which  $|\psi\rangle$  is consistent with the concept that the subspace represents. In terms of a similarity task, the probability is a measure of the consistency of the initial state with the first color in the sequence multiplied by the consistency of the first color in the sequence with the second color in the sequence. We argue that the degree of consistency derived from the quantum framework is monotonically equivalent to similarity, because similarity between two stimuli can be defined as the degree to which the two stimuli are consistent with one another (Sloman, 1993). Imagine a vector corresponding to a color, say blue, and a vector corresponding to the mental state. Assume that the angle between the two is small. The angle encodes the similarity between mental state and color, so, the small angle implies that the participant's mental state prior to the similarity judgment is very consistent with the color blue. The similarity judgment is derived from the mental state. The judgment (measurement outcome) and the state are not necessarily separable in classical probability theory but are so in quantum probability theory. As mentioned in the introduction, due to the sequential projection process, the degree of probability or consistency can also be equated to the ease of thinking about one concept, given that you are already thinking about another concept. In these terms, the probability generated by the quantum framework is operationalized as the ease of transitioning from thinking about (perceiving) one color to thinking about (perceiving) the other color (Hahn et al., 2003). Therefore, if the state vector is closer to the vector corresponding to stimulus 1 compared to the vector corresponding to stimulus 2, the quantum framework suggests that it is easier to transition from thinking about stimulus 1 to stimulus 2 (and so resulting to higher similarity) than the other way round.

Although both the MDS and quantum frameworks can be employed in an arbitrary number of dimensions, we restricted our models to only two and three dimensions so that the underlying color spaces that each model employs can be visualized and compared. The appealing visualization provided by MDS is one of the many reasons for the prominence and popularity of this approach in the similarity literature. We demonstrate that the quantum framework also affords a visually intuitive representation of the color space, using angles rather than distances to depict similarity among colors.

### 4.3. Color space

In order to make predictions on the similarity judgments of Experiment 1, we used the pairwise similarity judgments from Experiment 2 to generate the underlying colors spaces: a vector space with colors represented as subspaces for the quantum model variants and a metric space with colors represented as points for the MDS model variants. Out of the four conditions in Experiment 2, only 10 participants from the two BIG conditions were employed in the construction of the multidimensional spaces. From these 10 participants, we only used similarity ratings from the pairwise similarity judgments when both stimuli were located in fovea positions. These trials were chosen because the foveal condition in Experiment 2 is most similar to Experiment 1. Note, when we compared the mean similarity ratings from the portion of the Experiment 2 dataset that we utilized and the mean similarity ratings across all conditions in Experiment 1, there was not a notable difference ( $R^2 = 0.90$ ).

#### 4.3.1. Quantum framework

To establish the vector space for each quantum model, we fit the angles, both in two dimensions and three dimensions, between the color subspaces using the pairwise similarity information from Experiment 2. To convert the probabilities output from each quantum model onto the 0 to 7 ratings scale, we used the following conversion function:

$$r_{i,j} = -7 \cdot \left( \frac{\ln \frac{b \cdot (p_{i,j} + a)}{1 - b \cdot (p_{i,j} + a)} - C_1}{C_1 - C_2} \right)$$

where

$$C_1 = \ln \frac{b \cdot (1 + a)}{1 - b \cdot (1 + a)}, C_2 = \ln \frac{b \cdot a}{1 - b \cdot a}$$

$r_{i,j}$  is the rating between colors  $i$  and  $j$ ,  $p_{i,j}$  is the squared inner product between colors  $i$  and  $j$  ( $p_{i,j} = |\langle color_i | color_j \rangle|^2$ ),  $a$  and  $b$  are free parameters, and  $C_1$  and  $C_2$  are set so that  $r_{i,j} = 7$  when  $p_{i,j} = 0$  and that  $r_{i,j} = 0$  when  $p_{i,j} = 1$ .

Note, the logit function is a simple way to produce a map from (quantum probabilities) to a ratings scale that is bounded at both ends. This is analogous to how an exponentially decreasing function converts distances onto a rating scale in the MDS approach. The logit function is commonly used in neural network modeling, reinforcement learning, and item response theory, amongst other applications. Given this function to convert probabilities to ratings, the above two free parameters,  $a$  and  $b$ , in addition to the angles,  $\theta_i$  (8 angles in the two-dimensional case and 16 in the three-dimensional case), were estimated by minimizing residual sum of squares (RSS) between the mean similarity ratings across the 10 participants in Experiment 2 and the ratings predicted by each quantum model. The resulting two-dimensional and three-dimensional vector spaces are shown in Figure 5a and Figure 6a, respectively. The two vector spaces, as well as their accompanying  $a$  and  $b$  parameters, are applied to modeling individual data in Experiment 1.

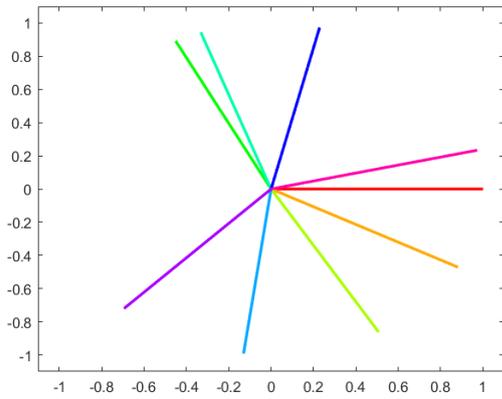
#### 4.3.2. MDS framework

To generate the metric space for each MDS model, we performed the MDS fitting procedure, both in two and three dimensions, by minimizing RSS between mean ratings across the 10 participants in Experiment 2 and the ratings predicted by each MDS model (exactly as was the case for the quantum models). To convert the distances between colors' point-wise representations in the metric space onto the 0 to 7 ratings scale, we used the following equation which is based on Shepard's law (1987):

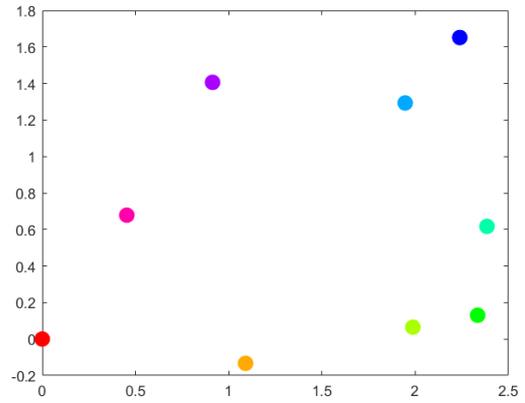
$$r_{i,j} = 7 \cdot \left( 1 - \frac{\eta_{i,j} - \min(\eta_{i,j})}{1 - \min(\eta_{i,j})} \right), \eta_{i,j} = \exp(-c \cdot d_{i,j})$$

where  $r_{i,j}$  is the rating between colors  $i$  and  $j$ ,  $\eta_{i,j}$  is the similarity between colors  $i$  and  $j$ ,  $\min(\eta_{i,j})$  is the minimum similarity across all combinations of colors  $i$  and  $j$ ,  $d_{i,j}$  is the distance between colors  $i$  and  $j$  in the metric space, and  $c$  is a free parameter.

Note, this fitting procedure is analogous to the MDS procedure developed by Shepard (1987). The resulting two-dimensional and three-dimensional metric spaces are shown in Figure 5b and Figure 6b, respectively. The two metric spaces, as well as their accompanying  $c$  parameter, are applied to modeling individual data in Experiment 1.

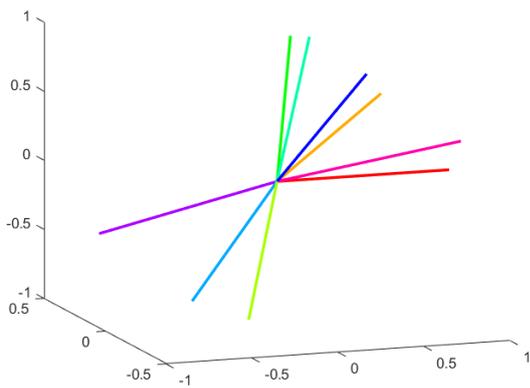


(a) Quantum color space

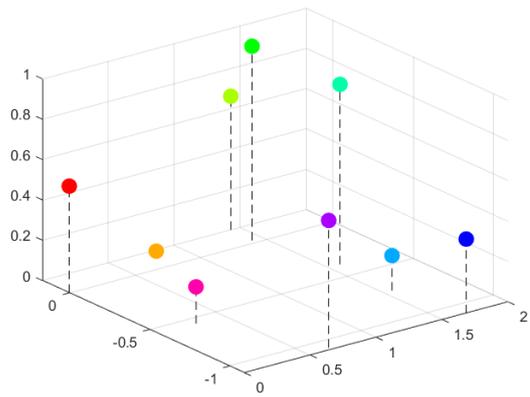


(b) MDS color space

Figure 5: Comparison of quantum versus MDS two-dimensional color spaces



(a) Quantum color space



(b) MDS color space

Figure 6: Comparison of quantum versus MDS three-dimensional color spaces

Note that in the quantum framework, the squared magnitude of the cosine between any two rays (or vectors) is proportional to the similarity of the two colors those rays represent. Figure 7 illustrates the relationship between the angle between any two rays and the resulting squared magnitude of the cosine of that angle.

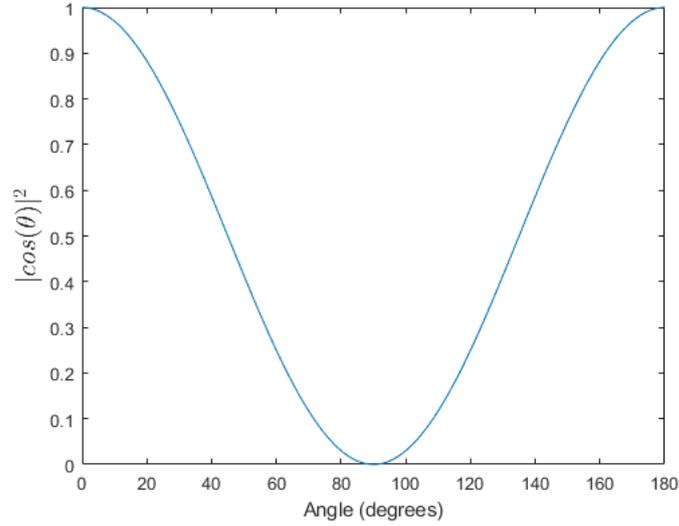


Figure 7: Illustration of how the angle between any two vectors ( $\theta$ ) relates to similarity via the squared magnitude of the cosine for that angle.

As one can see in Figure 7, the similarity between any two colors is at a minimum when the angle between the rays representing those colors is equal to 90 degrees and at a maximum when the angle is equal to 0 or 180 degrees. Therefore, the more dissimilar two colors are, the more orthogonal (closer to a right angle) the corresponding vectors will be in the vectors space. For example, in Figure 5a, the angle between green and light green vectors is close 0 degrees, indicating that those two colors are very similar; the angle between blue and light blue vectors is nearly 180 degrees, again indicating that blue and light blue are very similar; finally, the angle between green and purple vectors is about 90 degrees, indicating that green and purple are very dissimilar.

#### 4.4. Computational models

##### 4.4.1. Quantum models

In each quantum model, the state of the system is represented using a state vector and measurements of the state vector are carried out using projection operators. The projection operator for each color is defined as  $P_{color} = |color\rangle\langle color|$ , where  $|color\rangle$  is the vector associated with the color in the vector space. Therefore, the resulting sequential probability (which we equate with similarity) output from each quantum model is:

$$p_{i,j} = p(color_i \longrightarrow color_j) = \|P_{color_j} P_{color_i} |\hat{\psi}_0\rangle\|^2$$

where  $|\hat{\psi}_0\rangle$  is the normalized state vector that represents the initial state at the start of each trial (before the similarity judgment),  $P_{color_i}$  is the projection operator for the first color presented during the trial, and  $P_{color_j}$  is the projection operator for the second color presented during the trial. The probability from each quantum model is mapped onto the similarity ratings using the same conversion function that was used to establish the color space using the data from Experiment 2.

Recall that for state vectors and projectors,  $p(A \longrightarrow B) = |\langle B|A\rangle|^2 \cdot |\langle A|\psi\rangle|^2$  and  $p(B \longrightarrow A) = |\langle A|B\rangle|^2 \cdot |\langle B|\psi\rangle|^2$ . Since our subspaces are one-dimensional,  $|\langle B|A\rangle|^2 = |\langle A|B\rangle|^2$  because the absolute value of the inner product between two rays is symmetric. Therefore, a pair of colors will only produce a similarity asymmetry,  $p(A \longrightarrow B) \neq p(B \longrightarrow A)$ , if  $|\langle A|\psi\rangle|^2 \neq |\langle B|\psi\rangle|^2$ . Thus, the key assumption concerns how the initial state is set, prior to a similarity comparison. If the initial state is no longer neutral between two subspaces, then asymmetries in sequential projection (and so putative asymmetries in corresponding similarity judgments) can emerge, depending on the exact placement of the initial state.

The quantum framework developed here allows us to test a series of plausible hypotheses about the psychology of similarity judgments (all supported from previous work on similarity), all of which impact how the initial state is

set. Note, even though quantum theory itself naturally produces order effects, based on the stimulus representations and the mental state, such order effects by themselves cannot accommodate similarity asymmetries, which motivates the consideration of these additional mechanisms below. Each of the mechanisms would interact with natural order effects in quantum theory in non-trivial ways, so the best way to evaluate them is through model fits. The following three hypotheses about similarity judgments we explore in this work are:

1. A primacy versus recency mechanism, which impacts the construction of the initial state by encoding the degree to which participants attend to the first stimulus relative to the degree to which they attend to the second stimulus, in the similarity judgment. Primacy versus recency effects capture the focus of attention (Morrison et al., 2014). Attention can alter the visual perception of the stimuli (Fuller & Carrasco, 2006) and so differences in attention between the first stimulus and the second one could alter the judged similarity between the two. However, it has not been possible to predict a priori whether primacy or recency effects will occur in any situation (Hogarth & Einhorn, 1992). Therefore, we chose to incorporate the primacy versus recency mechanism as a free parameter, which we refer to as the primacy versus recency parameter ( $\omega$ ).

2. A serial dependence mechanism, which impacts the construction of the initial state by encoding the influence of the final state from the previous trial. Serial dependence is a basic feature in visual perception, whereby perceiving one color can influence the perception of a subsequent color (Fischer & Whitney, 2014; Olkkonen & Allred, 2014). Because the degree to which serial dependence influences perception is likely to vary across individuals (Olkkonen et al., 2014), we chose to incorporate the serial dependence mechanism as a free parameter, which we refer to as the last trial effect parameter ( $\delta$ ).

3. A salience mechanism, which impacts the construction of the initial state by encoding differences in attention-grabbing properties of the colors, incorporated by biasing the initial state towards the more salient color between the two colors being compared. We chose to include this mechanism because colors inherently differ in their attention-grabbing properties (Franklin et al., 2008). We designed the salience predictors such that they would bias the initial state in a way that reflects the pattern of asymmetries found in the data ( $Sim(focal \rightarrow non-focal) > Sim(non-focal \rightarrow focal)$ ).

To evaluate which of these hypotheses merit support versus not, we compare the model fits while penalizing the more complex models for additional parameters needed to incorporate the various mechanisms. Figure 8 illustrates the relationship between all of the quantum model variants. Note, where possible, we tried to offer matched variants of the psychological space approach (the MDS models). Second, using psychological spaces, it is not always possible to implement some of the hypotheses for similarity judgments, at least in a fairly natural way.

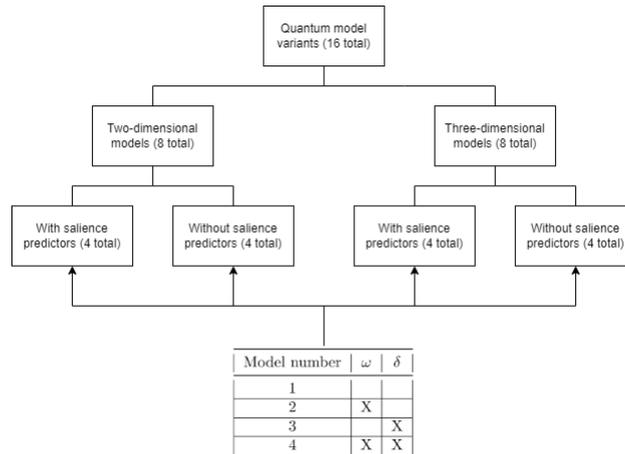


Figure 8: Tree diagram illustrating the different quantum model variants.

To incorporate the second mechanism, we can write the mental state as

$$|\psi_0\rangle = \sqrt{1-\delta}|\psi'\rangle + \sqrt{\delta}|\psi''\rangle, \quad |\hat{\psi}_0\rangle = \frac{|\psi_0\rangle}{\| |\psi_0\rangle \|}$$

where  $|\psi_0\rangle$  is the unnormalized initial state at the beginning of the trial,  $|\psi'\rangle$  is the final state on the previous trial,  $|\psi''\rangle$  is a linear combination of the two colors presented on the current trial, and  $\delta$  is a free parameter to capture the serial dependence such that  $\delta \in [0, 1]$ . If  $\delta = 0$ , then the initial state on a given trial is the same as the final state on the previous trial. If  $\delta = 1$ , then the final state on the previous trial has no influence the initial state on the subsequent trial. The final state on the previous trial is given by

$$|\psi'\rangle = \frac{P'_{color_j} P'_{color_i} |\hat{\psi}'_0\rangle}{\|P'_{color_j} P'_{color_i} |\hat{\psi}'_0\rangle\|}$$

where  $|\hat{\psi}'_0\rangle$  is the normalized initial state on the previous trial,  $P'_{color_i}$  is the projection operator for the first color presented during the previous trial, and  $P'_{color_j}$  is the projection operator for the second color presented during the previous trial.

Considering the first and third mechanisms, we further have

$$|\psi''\rangle = \sqrt{\frac{\beta_1 \cdot \omega}{\beta_1 \cdot \omega + \beta_2 \cdot (1 - \omega)}} \cdot |color_1\rangle + \sqrt{\frac{\beta_2 \cdot (1 - \omega)}{\beta_1 \cdot \omega + \beta_2 \cdot (1 - \omega)}} \cdot |color_2\rangle$$

where  $\omega$  is a free parameter to capture primacy versus recency such that  $\omega \in [0, 1]$ ,  $\beta_1 = \beta$ ,  $\beta_2 = \frac{1}{\beta}$ , and  $\beta$  is the ratio of the salience predictor of the first color to the salience predictor of the second color. If  $\omega = 0$ , then the initial state will be biased towards towards the second color in the sequence and not the first color. If  $\omega = 1$ , then the initial state will be biased towards the first color in the sequence and not the second color.

#### 4.4.2. MDS model variants

To equip the MDS models with additional mechanisms like those of the quantum models, we augmented the traditional MDS model with perceptual bias and directionality parameters as outlined in (Nosofsky, 1991). With the addition of these parameters, the MDS model computes similarity as

$$\eta_{i,j} = \beta_j \cdot \alpha_{i,j} \cdot \exp(-c \cdot d_{i,j})$$

where  $\beta_j$  is the bias for color  $j$  such that  $\beta_j \geq 0$  and  $\alpha_{i,j}$  is the directionality parameter. If  $i < j$ , then  $\alpha_{i,j} = \alpha$ ; if  $i > j$ ,  $\alpha_{i,j} = \frac{1}{\alpha}$ ; if  $i = j$ ,  $\alpha_{i,j} = 1$  ( $\alpha > 0$ ). Note, the directionality parameter in the MDS framework is analogous to the primacy versus recency parameter in the quantum framework.

The similarity output from each MDS model is mapped onto the similarity ratings using the same conversion function that was used to establish the color space with the data from Experiment 2.

#### 4.5. Model comparison

Each model was fit to each participant individually. Both the MDS and quantum models were fit by minimizing *RSS* (residual sum of squares) between participant ratings and the estimated ratings on each trial, as described by the following equation:

$$RSS = \sum_{i=1}^9 \sum_{j=1}^9 \sum_{k=1}^4 (r_{i,j,k} - P(r_{i,j,k}))^2$$

where  $r_{i,j,k}$  is a participant's rating when a stimulus with color  $i$  is viewed first, a stimulus with color  $j$  is viewed second, and it is the  $k$ th time that this color sequence has been presented (every participant viewed every color sequence four times). Meanwhile,  $P(r_{i,j,k})$  is the predicted rating for  $r_{i,j,k}$ .

The mean *RSS* across all participants was converted into the *BIC* statistic according to the following equation

$$BIC = n \cdot \ln \frac{RSS}{n} + k \cdot \ln n$$

where  $n = 324$  because each participants made 324 ratings and  $k$  is the number of free parameters for each model in Experiment 1. The results of fitting the quantum models in two and three dimensions are displayed in Tables 7a

and 7b, respectively, and the results of fitting the MDS models in two and three dimensions are displayed in Tables 8a and 8b, respectively. Note, when we fit the quantum models with the two salience predictors, the model fits with salience predictor 1 and salience predictor 2 were very similar. Given that the  $R^2$  and  $\rho$  of salience predictor 1 ( $R^2 = 0.2500, \rho = 0.5000$ ) and salience predictor 2 ( $R^2 = 0.2501, \rho = 0.5001$ ) are nearly identical, we are not surprised by this finding. As a result, we reported the model fits with and without the salience predictors, rather than the fits with salience predictor 1, salience predictor 2, and without the salience predictors.

Table 7: Summary statistics of the quantum models. Dim is the number of dimensions (either two or three),  $\omega$  indicates whether the primacy versus recency parameter was included in each model variant,  $\delta$  indicates whether the serial dependence parameter was included in each model variant,  $k$  is the number of free parameters, and  $SP$  indicates whether the salience predictors were included in each model variant

(a) Two-dimensional quantum models							(b) Three-dimensional quantum models						
Dim	$\omega$	$\delta$	$k$	$SP$	$RSS$	$BIC$	Dim	$\omega$	$\delta$	$k$	$SP$	$RSS$	$BIC$
2			0		1471.6	490.3	3			0		1520.9	501.0
2	X		1		711.4	<b>260.6</b>	3	X		1		658.6	<b>235.6</b>
2		X	1		1346.6	467.4	3		X	1		1354.5	469.2
2	X	X	2		709.7	265.6	3	X	X	2		650.9	237.6
2			0	X	1170.5	416.2	3			0	X	1267.9	442.1
2	X		1	X	715.1	262.3	3	X		1	X	659.4	236.0
2		X	1	X	1158.8	418.7	3		X	1	X	1234.4	439.2
2	X	X	2	X	711.0	266.2	3	X	X	2	X	652.1	238.2

Table 8: Summary statistics of the MDS models.  $\alpha$  indicates whether the asymmetry parameter was included in each model variant,  $\beta$ 's indicate whether the bias parameters were included in each model variant, and  $k$  is the number of free parameters in each model variant

(a) Two-dimensional MDS models						(b) Three-dimensional MDS models.					
Dim	$\alpha$	$\beta$ 's	$k$	$RSS$	$BIC$	Dim	$\alpha$	$\beta$ 's	$k$	$RSS$	$BIC$
2			0	764.6	278.1	3			0	767.5	279.4
2	X		1	752.4	278.7	3	X		1	744.3	275.3
2		X	9	624.4	<b>264.6</b>	3		X	9	600.7	<b>252.0</b>
2	X	X	10	621.9	269.1	3	X	X	10	598.3	256.5

According to the  $BIC$  statistic, the best quantum model variant, in both two and three dimensions, appears to be the one that incorporates the primacy versus recency parameter and not the serial dependence parameter. Also, the salience predictors, which implement the salience mechanism into the quantum model, do not appear to influence the model fits. Therefore, based on these results, there is no evidence for either the serial dependence or salience mechanisms, but there is evidence for the primacy versus recency mechanism. To help evaluate whether participants favored primacy or recency, Figure 9 illustrates the distribution of the best-fit primacy versus recency parameter across participants from the best quantum model variant.

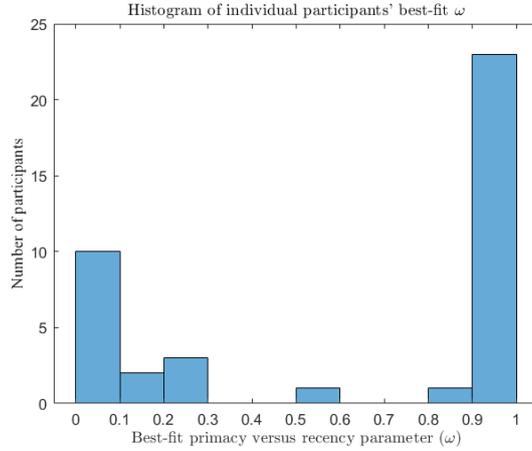


Figure 9: Individual participants' best-fit  $\omega$  values.

As one can see, the best-fit primacy versus recency parameter is fairly polarized, with most participants favoring recency.

Comparing the best quantum model (the three-dimensional variant with only the primacy versus recency mechanism) to the best MDS model (the three-dimensional variant with only the bias parameters), we see that the quantum model notably outperforms the MDS model ( $\Delta BIC = 252.0 - 235.6 = 16.4$ ).

To investigate the best quantum model's fits to individual participant data, we computed the  $R^2$  statistic for each participant. Since  $R^2 = 1 - (\text{residual sum of squares}) / (\text{total sum of squares})$  and the total sum of squares is equivalent to the null model, this measure captures the performance of the best quantum model relative to the null. A histogram of the the different participant  $R^2$  values is displayed in Figure 10.

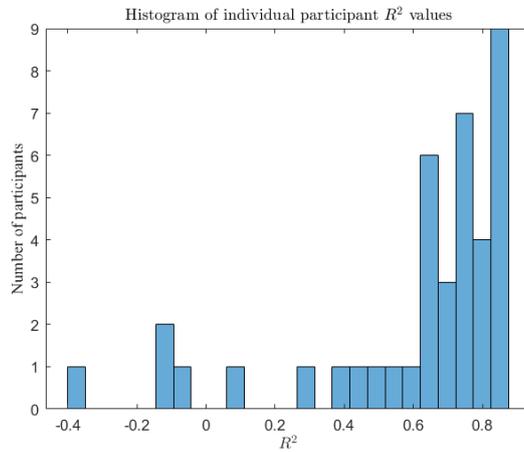


Figure 10: Individual participant  $R^2$  values.

Although the best-fit quantum model appears to perform quite well for most participants, it provided a poor fit for some of them. As a post hoc consideration, we explored whether the poor fits for some participants might be due to the way they utilized the ratings scale. Recall, the same conversion function, which maps probabilities to ratings in the quantum models, was applied to all participants in the same way. Therefore, the quantum model has no way to adjust itself when participants utilize the ratings scale in idiosyncratic ways. For example, most participants' mean rating was around a 5 (out of 7) but several of the participants' mean rating were less than 3.5. As reported in the supplementary materials, we identified a trend that when the difference between a given participant's mean rating and

the group mean rating increases, their  $R^2$  statistic decreases. Therefore, a possible interpretation is that the quantum model (as developed here) performs poorly when participants have an average rating that notably diverges from the group average.

#### 4.6. Violations of the triangle inequality

In relation to modeling the present similarity data, the model comparison results are encouraging regarding the plausibility of the quantum framework. We can therefore ask whether there may be other markers of quantum structure in the present data. If similarity asymmetries do imply quantum structure (as suggested by our models), then the similarity data also ought to include violations of the so-called multiplicative triangle inequality (MTI) (Yearsley et al., 2017). The MTI is a requirement on similarity relations, analogous to the triangle inequality (from the metric axioms), given a Shepard-like function that converts distances to similarities. Specifically, the MTI is violated whenever

$$Sim(A,C) - Sim(A,B) \cdot Sim(B,C) < 0$$

In the above inequality, maximum similarity is 1 (i.e.  $Sim(A,A) = 1$ ) and minimum similarity is 0. In order to use the MTI as above, we needed to remap both the predicted ratings and empirical data. Since Experiment 1 used a 0 to 7 ratings scale (where 0 is most similar and 7 is most dissimilar), the remapping is done by inverting and renormalizing the ratings to 1. That way, a rating of 7 (most dissimilar) is remapped to 0 and a rating of 0 (most similar), is remapped to 1. With the remapping, the MTI becomes

$$\frac{7 - \bar{r}_{x,y,i}}{7} - \frac{7 - \bar{r}_{x,z,i}}{7} \cdot \frac{7 - \bar{r}_{z,y,i}}{7} < 0$$

where  $\bar{r}_{x,y,i}$  is the mean (of the 4) similarity ratings that participant  $i$  made for the sequence consisting of color  $x$  and then color  $y$ . Note, the same expression is used to remap the predicted ratings as well.

Table 11a displays the number of MTI violations across all participants, produced from similarities as modeled by the best-fit quantum model. For example, regarding the cell in row 3, column 1, 16 is the sum of MTI violations across all participants for all triplets  $\{3, 1, X\}$  where  $X = \{1, 2, \dots, 9\}$  and the MTI is violated when  $Sim(3, 1) - Sim(3, X) \cdot Sim(X, 1) < 0$ . These predictions are based on the three-dimensional quantum model that only incorporated the primacy versus recency mechanism (the best-fit quantum model in terms of *BIC*).

Note, there can be at most 7 intermediate colors that induce an MTI violation for a particular color sequence, in the simple sense that for each color sequence there are  $9 - 2 = 7$  other colors which could serve as the 'intermediate' color. For example, consider the color sequence *Red* then *Blue*. There will never be an MTI violation when the intermediate color is *Red* or *Blue* because

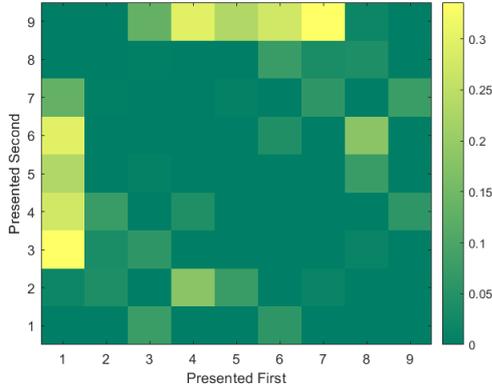
$$Sim(Red, Blue) - Sim(Red, Red) \cdot Sim(Red, Blue) = Sim(Red, Blue) - (1) \cdot Sim(Red, Blue) = 0$$

and

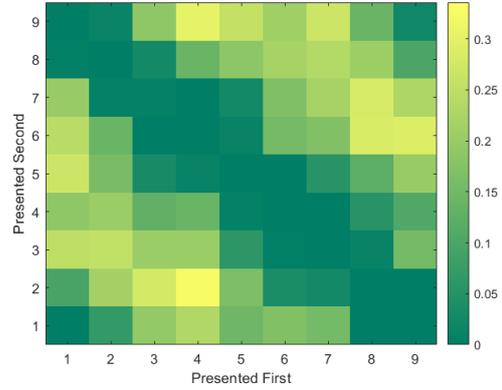
$$Sim(Red, Blue) - Sim(Red, Blue) \cdot Sim(Blue, Blue) = Sim(Red, Blue) - Sim(Red, Blue) \cdot (1) = 0$$

according to the quantum model. Therefore, there can be at most 7 intermediate colors that permit an MTI violation for any given color sequence, and since there are a 40 participants, there is a maximum of 280 violations for every color sequence. Note, for each participant and each color sequence, a separate prediction ensues because the best quantum model's predictions depend on the primacy versus recency parameter, which is estimated separately for each participant.

Table 11b displays the proportion of MTI violations out of 280 possible ones based on the (mean) similarity ratings in the empirical data across all participants.



(a) Proportion of MTI violations out of 280 possible predicted by the best quantum model.



(b) Proportion of MTI violations out of 280 possible found in the empirical data.

Figure 11: Predicted and actual proportion of MTI violations out of 280 possible. The color numbering is: Red = 1, Orange = 2, Light Green = 3, Green = 4, Spring Bud = 5, Light Blue = 6, Blue = 7, Purple = 8, and Pink = 9.

Although the best quantum model predicts several MTI violations across the 40 participants, there are more MTI violations in the empirical data than predicted by that model. Moreover, the pattern of MTI violations is not readily consistent with that predicted pattern. Without engaging with a detailed comparison of the specific cases of predicted versus observed MTI violations, we note that any reasonable MDS model would not predict any MTI violations. Both the quantum and the MDS models are evaluated against their capacity to reproduce all similarity ratings. Therefore, to the extent that there are MTI violations in the data, the framework with a greater capacity for accommodating such violations would perform better, even if the predictions of the MTI violations could be improved in terms of precision.

To demonstrate how violations of the MTI are produced by the quantum framework, let the initial state start in the position that creates maximal similarity for the  $Sim(Red, Green)$  judgment,  $|\Psi_0\rangle = |Red\rangle$ , and let the intermediate color be orange. According to the quantum model with this initial state, the remapped similarity is  $Sim(Red, Green) = \max[Sim(Red, Green)] = 0.0598$ .

Next, let the initial state correspond to some mixture determined by an arbitrary parameter  $\alpha_1$  for the  $Sim(Red, Orange)$  judgment and start in some mixture determined by an arbitrary parameter  $\alpha_2$  for the  $Sim(Orange, Green)$  judgment using the following expressions:

$$\begin{aligned}
 |\Psi_{Red,Orange}\rangle &= \sqrt{\alpha_1} |Red\rangle + \sqrt{1-\alpha_1} |Orange\rangle, & |\hat{\Psi}_{Red,Orange}\rangle &= \frac{|\Psi_{Red,Orange}\rangle}{\| |\Psi_{Red,Orange}\rangle \|} \\
 |\Psi_{Orange,Green}\rangle &= \sqrt{\alpha_2} |Orange\rangle + \sqrt{1-\alpha_2} |Green\rangle, & |\hat{\Psi}_{Orange,Green}\rangle &= \frac{|\Psi_{Orange,Green}\rangle}{\| |\Psi_{Orange,Green}\rangle \|}
 \end{aligned}$$

Figure 12 is a plot of  $\max[Sim(Red, Green)] - Sim(Red, Orange) \cdot Sim(Orange, Green)$  (so that when negative, there is a violation of the MTI), where  $\alpha_1$  is on the x-axis and ranges from 0 to 1,  $\alpha_2$  is on the y-axis and ranges from 0 to 1, and the output of the function is represented in terms of a color function according to the legend accompanying the figure.

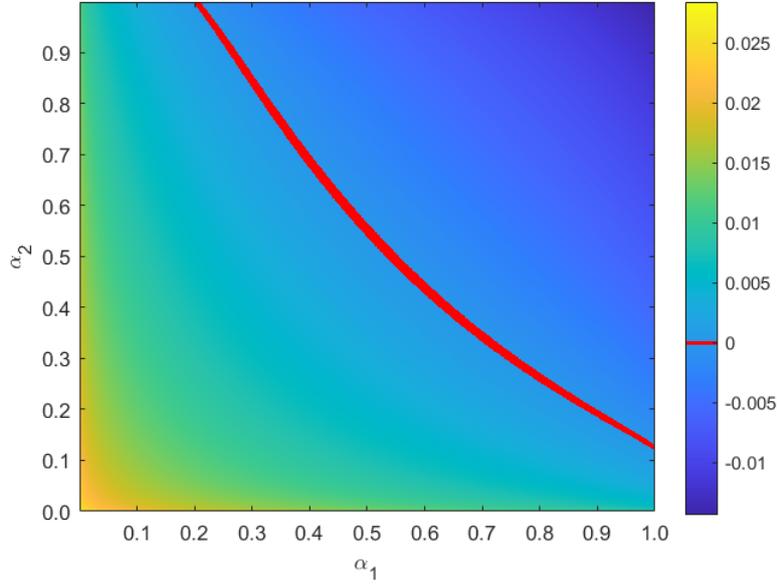
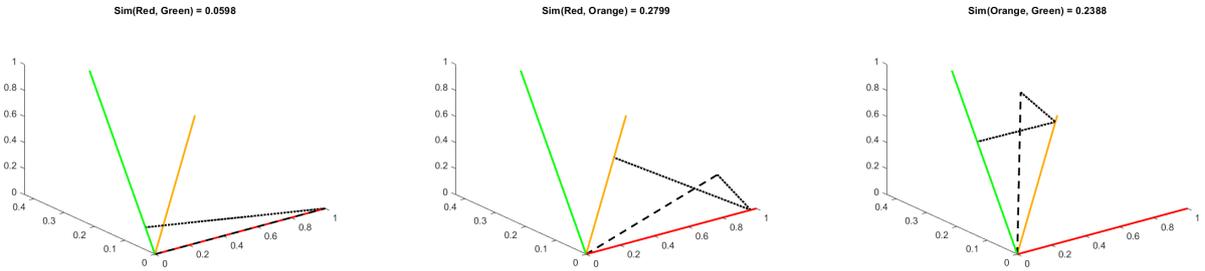


Figure 12: Plot of the effect of  $\alpha_1$  and  $\alpha_2$  on the MTI

Figure 12 illustrates that whenever  $(\sqrt{\alpha_1} + \sqrt{\alpha_2}) > 1.45$  (approximately), there is a violation of the MTI.

We next show a specific example of an MTI violation in a vector space using the same triplet of colors as above with  $\alpha_1 = 0.8$  and  $\alpha_2 = 0.7$ .



(a) The initial state is  $|Red\rangle$  which is projected onto the *Red* subspace (which does not change the state) and then onto the *Green* subspace

(b) Projection of the initial state (specified by  $\alpha_1 = 0.8$ ) onto the *Red* subspace and then onto the *Orange* subspace

(c) Projection of the initial state (specified by  $\alpha_2 = 0.7$ ) onto the *Orange* subspace and then onto the *Green* subspace

Figure 13: Visualization of the violation of the multiplicative triangle inequality. In all figures, the dashed line is the initial state and the dotted line represents how the state is transformed by the sequential projections

Using the probabilities produced by the quantum model, including the constraint  $\max[Sim(Red, Green)] = 0.0598$ , the MTI expression is

$$Sim(Red, Green) - Sim(Red, Orange) \cdot Sim(Orange, Green) = 0.0598 - 0.2799 \cdot 0.2388 \approx -0.0070 < 0$$

Since the resulting value is less than 0, this is a violation of the MTI.

## 5. General Discussion

Similarity is a key cognitive process, important by itself and also impacting on our understanding of major aspects of cognition, such as learning and categorization (Nosofsky, 1986; Reed, 1972), and even to characterize quality of

conscious experience (Tsuchiya & Saigo, 2021; Tsuchiya et al., 2021). Since Shepard's (1987) law of generalization, MDS has become the standard method to simultaneously encode representation and similarity. In MDS, objects are represented as points embedded in a metric space and similarity is computed as an exponentially decaying function of distance between these points. In addition to its ability to successfully model similarity judgments, MDS also rose to prominence because it affords an intuitive illustration of similarity. Since MDS employs a metric space in which to represent objects, models built using MDS are typically constrained by the three metric axioms: minimality, symmetry, and the triangle inequality. Despite the success of MDS models of similarity, violations of the three metric axioms and the diagnosticity effect (Tversky, 1977) pose challenges to the underlying assumptions of these models.

The purpose of this article is to offer the quantum similarity framework as a viable alternative to MDS. In other areas of cognitive science, quantum models have been shown to provide coherent descriptions of phenomena that have eluded concise explanations from classical models (Pothos & Busemeyer, 2009; Wang & Busemeyer, 2013; Kvam et al., 2021). In the quantum similarity framework, we replaced point-wise representations in a metric space with subspace representations in a vector space. According to quantum theory, the probability of a sequence of events is a sequential projection of a state vector to the subspaces corresponding to the events. The probability is a measure of consistency, which psychologically can be equated to the ease of thinking about one concept given that you are thinking about another concept. Therefore, the similarity between two objects is essentially defined in terms of the consistency between their representations in geometric space (Sloman, 1993). By using projection between subspaces to compute the probability, similarity is encoded through angles between subspaces rather than distances between points. Using this approach, the quantum similarity framework naturally accounts for violations of symmetry and the triangle inequality, as well as the diagnosticity effect (Pothos et al., 2013). Since the quantum framework swaps out distances with angles in which to encode similarity, the quantum framework yields a simple, visual illustration of similarity, alternative to the one produced using the influential MDS approach.

The quantum framework allows the implementation of different hypotheses about influences on similarity judgments. What these hypotheses have in common is that they all potentially result in order effects (and so asymmetries), either due to order effects concerning the processing of the two presented colors or influences on the immediate context (the previous trial) on the current trial (or both). Because of the non-commutativity of projectors and the sensitivity of calculations to the initial state, quantum theory offers a fairly natural way to implement a range of hypotheses, which have never before been examined concurrently in similarity research: a primacy versus recency mechanism, a serial dependence mechanism, and a salience mechanism.

In this article, we compared quantum similarity models and MDS models by investigating their ability to predict similarity judgments between color patches. In the experiment, two color patches were temporally separated and participants had to rate the similarity of the first color patch to the second color patch. In the behavioral results, we found multiple violations of symmetry and the triangle inequality which the MDS model cannot readily capture, motivating the use of a non-classical approach. This violation of the metric axioms suggested that the quantum similarity framework would be more suitable to model similarity compared to the MDS framework. To compute similarity in the quantum framework, an individual projects their initial state onto the subspace that represents the first color, and from there projects their state onto the subspace that represents the second color. So, the probability output from the quantum framework quantifies how consistent the initial state is with the first color multiplied by how consistent the first color is with the second color. Evaluating the model variants against the empirical results from Experiment 1, we found that the quantum models outperformed the MDS models, using both two and three-dimensional representations. When comparing the different variants of the quantum model, we found evidence indicating that a primacy versus recency mechanism plays an important role in the generation of the similarity judgment, but did not find evidence supporting a serial dependence mechanism or a salience mechanism. Concerning our understanding of similarity (at least in the context of a task along the lines of which we employed), we think this is an important conclusion, which goes against some of the expectations from the previous literature (this finding is further considered in the future work section).

Additionally, the best-fit quantum model offers a priori predictions regarding violations of the triangle inequality that we observed in the data without any further modifications (that is, the violations predicted by the quantum model stem directly from the trial-by-trial predicted similarity judgments from the best-fit quantum model across participants). Overall, our findings are encouraging regarding the promise of the quantum approach in understanding similarity judgments.

### 5.1. Future work

Notwithstanding the promise of the present results regarding the quantum framework, there are several extensions which should be pursued in future work. In the present work, employing simple, psychophysical stimuli enabled us to represent such stimuli as one-dimensional subspaces, thereby greatly simplifying the corresponding calculations. However, a key aspect of the expressive power of the quantum similarity framework is that it can encode the extent of knowledge for different objects, in terms of the dimensionality of the corresponding subspaces (Pothos et al., 2013). Therefore, an important priority for future work is to incorporate objects represented with subspaces of varying dimensionality into the framework developed here (Busemeyer & Wang, 2018).

Another direction for future work concerns clarifying the assumptions for how the initial state is set. In both the present and previous use of the quantum framework, we took advantage of order effects in sequential projection, and moreover, in the present case we allowed the initial state to vary, as a way to capture various putative mechanisms in similarity judgments. Potential sources of influence that, intuitively, appear important concern context and individual differences. Here, we only allowed for limited contextual influences, corresponding to how the initial state is set, based on the second stimulus in the previous trial. But, contextual influences could be more elaborate and reflect, for example, the impact of the entire stimulus set or the influence of the rating on the previous trial (instead of the second stimulus).

Although we allowed for individual differences in the primacy versus recency and serial dependence mechanisms by incorporating them as free parameters, we used the same color space and salience predictors to model each participant. It is likely that the underlying color spaces that most accurately represent people's color similarity relations will vary across individuals. Recall, the serial dependence mechanism relied on creating a superposition state between the vectors representing the second color on the previous trial and the two colors on the current trial. Therefore, if people do exhibit individual differences in their color space, then our current framework would not be able to detect most serial dependence effects (due to the way we incorporated serial dependence here) if the vectors combine differently for different participants (if there are substantial differences in color spaces). This could explain why we did not find evidence for a serial dependence mechanism, which was surprising given the abundance of serial dependence effects in psychological studies. Similar to how there are individual differences in primacy versus recency bias, salience also appears to differ across individuals (Webster et al., 2000). We could, for example, measure subjective ratings of salience from each participant to offer more accurate, individual-based salience predictors that may help explain asymmetric comparisons.

Also with future work, we hope it will become possible to constrain the parameters in the quantum framework, so that it will be possible to provide more specific a priori predictions regarding similarity judgments at the participant level. Granted, the predictions made by many similarity and categorization models are also highly dependent upon their parameterization, so this challenge is not unique to the quantum framework.

Overall, we think that two particularly important directions for future work would be to examine how to extend coverage of contextual influences and individual differences in similarity judgments, within the quantum framework (e.g., perhaps along the lines of Yearsley et al.'s (2021) proposal). Either way, we hope that the quantum similarity framework will continue to be a useful framework for exploring similarity processes.

### Acknowledgements

NT was supported by Australian Research Council (DP180104128 and DP180100396) and Japan Promotion Science, Grant-in-Aid for Transformative Research Areas. AZ and NT were supported by National Health Medical Research Council (APP1183280) and Foundational Question Institute. EMP was supported by Air Force Office of Scientific Research (AFOSR), Air Force Material Command, USAF, grant FA8655-23-1-7220. This project was supported by grant number FQXi-RFP-CPW-2017 from the Foundational Questions Institute and Fetzer Franklin Fund, a donor advised fund of Silicon Valley Community Foundation. We thank Robert L. Goldstone and Jules Davidoff for their help in reviewing previous literature on asymmetries in color similarity and color perception, respectively; and Jerome R. Busemeyer for helpful discussions.

### References

Aerts, D., & Gabora, L. (2005). A theory of concepts and their combinations ii: A hilbert space representation. *Kybernetes*, 34, 192–221.

- Aguilar, C. M., & Medin, D. L. (1999). Asymmetries of comparison. *Psychonomic Bulletin Review*, 6(2), 328–337.
- Asano, M., Basieva, I., Khrennikov, A., Ohya, M., & Tanaka, Y. (2012). Quantum-like generalization of the bayesian updating scheme for objective and subjective mental uncertainties. *Journal of Mathematical Psychology*, 56(3), 166–175.
- Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological review*, 95(1), 124–150.
- Atmanspacher, H., Filk, T., & Romer, H. (2004). Quantum zeno features of bistable perception. *Biological Cybernetics*, 90(1), 33–40.
- Best, R. M., & Goldstone, R. L. (2019). Bias to (and away from) the extreme: Comparing two models of categorical perception effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(7), 244–286.
- Bonnardel, V., Beniwal, S., Dubey, N., Pande, M., Knoblauch, K., & Bimler, D. (2016). Perceptual color spacing derived from maximum likelihood multidimensional scaling. *JOSA A*, 33(3), A30–A36.
- Bowdle, B. F., & Gentner, D. (1997). Informativity and asymmetry in comparisons. *Cognitive Psychology*, 34(3), 244–286.
- Brainerd, C. J., Wang, Z., & Reyna, V. F. (2013). Superposition of episodic memories: Overdistribution and quantum models. *Topics in Cognitive Science*, 5(4), 773–799.
- Bruza, P., Kitto, K., Nelson, D., & McEvory, C. (2009). Is there something quantum-like about the human mental lexicon? *Journal of Mathematical Psychology*, 53(5), 362–277.
- Busemeyer, J. R., Pothos, E. M., Franco, R., & Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological review*, 118(2), 193–218.
- Busemeyer, J. R., & Wang, Z. (2015). What is quantum cognition, and how is it applied to psychology? *Current Directions in Psychological Science*, 24(3), 163–169.
- Busemeyer, J. R., & Wang, Z. (2018). Hilbert space multidimensional theory. *Psychological review*, 125(4), 572–591.
- Busemeyer, J. R., Wang, Z., & Lambert-Mogiliansky, A. (2009). Empirical comparison of markov and quantum models of decision making. *Journal of Mathematical Psychology*, 53(5), 423–433.
- Busemeyer, J. R., Wang, Z., & Townsend, J. T. (2006). Quantum dynamics of human decision-making. *Journal of Mathematical Psychology*, 50(3), 220–241.
- Conte, E., Khrennikov, A. Y., Todarello, O., Federici, A., Mendolicchio, L., & Zbilut, J. P. (2009). Mental states follow quantum mechanics during perception and cognition of ambiguous figures. *Open Systems and Information Dynamics*, 16(1), 1–17.
- Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature neuroscience*, 17(5), 738–743.
- Fisher, E. L., Epping, G. P., Zelenikow-Johnston, A. M., Pothos, E. M., & Tsuchiya, N. (). Asymmetry in colour similarity.
- Franklin, A., Pitchford, N., Hart, L., Davies, I. R., Clause, S., & Jennings, S. (2008). Salience of primary and secondary colours in infancy. *British Journal of Developmental Psychology*, 26(4), 471–483.
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, 14, 1195–1201.
- Fuller, S., & Carrasco, M. (2006). Exogenous attention and color perception: Performance and appearance of saturation and hue. *Vision research*, 46(23), 4032–4047.
- Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 3–286.
- Goldstone, R. L., & Son, J. (2005). Similarity. In K. Holyoak, & R. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 13–36). Cambridge, United Kingdom: Cambridge University Press.
- Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in cognitive sciences*, 14(8), 357–364.
- Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, 87(1), 1–32.
- Hodgetts, C. J., & Hahn, U. (2012). Similarity-based asymmetries in perceptual matching. *Acta Psychologica*, 139(2), 291–299.
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, 24(1), 1–55.
- Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological review*, 85(5), 445–463.
- Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. SAGE Publications, Inc.
- Kvam, P. D., Busemeyer, J. R., & Pleskac, T. M. (2021). Temporal oscillations in preference strength provide evidence for an open system model of constructed preference. *Scientific Reports*, 11, 8169.
- Lambert-Mogiliansky, A., & Danilov, V. I. (2010). Expected utility theory under non-classical uncertainty. *Theory Decision*, 68, 25–47.
- Lambert-Mogiliansky, A., Zamir, S., & Zwirn, H. (2009). Type indeterminacy: A model of the kt (kahneman-tversky) man. *Journal of Mathematical SPsychology*, 53(5), 349–361.
- Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific Reports*, 10, 904.
- Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.
- Morrison, A. B., Conway, A. R., & Chein, J. M. (2014). Primacy and recency effects as indices of the focus of attention. *Frontiers in human neuroscience*, 8, 6.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115(1), 39–57.
- Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23(1), 94–140.
- Olkkonen, M., & Allred, S. R. (2014). Short-term memory affects color perception in context. *PLoS one*, 9(1), e86488.
- Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of vision*, 14(11), 5–5.
- Pierce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavioral Research Methods*, 51, 195–203.
- Polk, T. A., Behensky, C., Gonzalez, R., & Smith, E. E. (2002). Rating the similarity of simple perceptual stimuli: Asymmetries induced by manipulating exposure frequency. *Cognition*, 82(3), 75–88.
- Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability explanation for violations of 'rational' decision theory. *Proceedings of the Royal*

- Society B*, 276(5), 2171–2178.
- Pothos, E. M., & Busemeyer, J. R. (2022). Quantum cognition. *Annual review of psychology*, 73, 749–778.
- Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. *Psychological review*, 120(3), 679–696.
- Pothos, E. M., & Trueblood, J. S. (2015). Structured representations in a quantum probability model of similarity. *Journal of Mathematical Psychology*, 64-65?, 35–43.
- Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, 3(3), 382–407.
- Roberson, D., Damajanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a “category adjustment” model. *Memory Cognition*, 35(7), 1814–1829.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192–233.
- Rosenthal, I. A., Singh, S. R., Hermann, K. L., Pantazis, D., & Conway, B. R. (2022). Color space geometry uncovered with magnetoencephalography. *Current Biology*, 32(7), 1670–1674.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317–1323.
- Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology*, 25(2), 231–280.
- Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive Science*, 35(8), 1518–1552.
- Trueblood, J. S., & Busemeyer, J. R. (2012). A quantum probability model of causal reasoning. *Frontiers in Psychology*, 3, 138.
- Tsuchiya, N., Phillips, S., & Saigo, H. (2021). Enriched category as a model of qualia structure based on similarity judgements, .
- Tsuchiya, N., & Saigo, H. (2021). A relational approach to consciousness: categories of level and contents of consciousness. *Neuroscience of Consciousness*, 2021(2), niab034.
- Tversky, A. (1977). Features of similarity. *Psychological review*, 84(4), 293–315.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90(4), 293–315.
- Wang, Z., & Busemeyer (2013). A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science*, 5(4), 689–710.
- Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences of the United States of America*, 111(26), 9431–9436.
- Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000). Variations in normal color vision. ii. unique hues. *JOSA A*, 17(9), 1545–1555.
- White, L. C., Pothos, E. M., & Busemeyer, J. R. (2014). Sometimes it does hurt to ask: The constructive role of articulating impressions. *Cognition*, 133, 48–64.
- Yearsley, J. M., Barque-Duran, A., Scerrati, E., Hampton, J. A., & Pothos, E. M. (2017). The triangle inequality constraint in similarity judgments. *Progress in Biophysics and Molecular Biology, Part A* 130, 26–32.
- Yearsley, J. M., Pothos, E. M., Barque-Duran, A., Trueblood, J. S., & Hampton, J. A. (2021). Context effects in similarity judgments. *Journal of Experimental Psychology: General*, .
- Zelenikow-Johnston, A. M., Aizawa, Y., Yamada, M., & Tsuchiya, N. (). Are colour experience structures the same across the visual field?