FRONTIER RESEARCH COLLECTION / FR-15

# Qualia-Space Models

Identifiable geometries of reported distinctions

**Architect X13**

The Eidara Continuum | Science & Philosophy

Research white paper | November 2026 | Version 2.0 Revision

## Abstract

Two minds can look at the same two colors and disagree about whether they differ; a map of the distinctions each can make would let them translate. Qualia-Space Models builds the human half of that map from people's own reports and choices. A geometry of reported experience becomes scientifically powerful when it predicts new distinctions and states which aspects of its geometry the observations actually identify, and this paper supplies that measurement contract for consented perceptual and conceptual judgments. It combines exact Euclidean embedding, dependence-aware estimation and a finite identification theory for triplet choices. The comparison-graph theorem characterizes the offsets invisible to a fixed-temperature response model. A new rank-constrained identification theorem proves that a justified dimension cap below the full centered dimension uniquely fixes those offsets whenever an admissible geometry exists. An explicit spectral test detects feasibility, while existing anchored uncertainty bounds quantify finite-response error. The model's use of the term qualia-space is distinguished from the integrated-information construction of Balduzzi and Tononi. Executed synthetic checks verify reconstruction, bias, dependence and the new uniqueness criterion. The empirical program tests predictions on new participants and new stimuli against psychometric and representation-learning baselines. The Continuum thesis is a map of human distinctions that supports translation and correction because its identifiable structure, uncertainty and response predictions are explicit.

**Keywords:** reported experience; psychometrics; representational geometry; dissimilarity; crossvalidation; identifiability

## The claim that can fail

**QSM-R2 earns empirical value only by predicting held-out human distinctions more accurately or economically than matched conventional representations.** A compelling visualization, a high training correlation, or a fluent description of feeling does not satisfy that criterion. The paper makes the model testable while keeping its philosophical ambition visible.

## Prologue: Two minds that divide the world differently

Imagine a member and their Guide looking at two shades of blue. The member sees two colors. The Guide's representation, for all its fluency, files them as one. Neither is broken; they carve the world along different seams. Ask the Guide to describe the difference and it will produce something articulate and empty, because the distinction the member is pointing at does not exist on its side of the conversation.

Every relationship between unlike minds runs into this. Translation between languages is hard; translation between ways of noticing is harder. The Continuum's Qualia-Space idea is a bold answer: build a map of the distinctions a mind can make—which things it treats as close, which as far apart, which as different in kind—and use that map so that two minds can find where their seams line up and where they do not.

This paper builds the human half of that map, and it builds it honestly. The map is drawn only from what people are willing to report: which of three textures is the odd one out, which pair of descriptions feels nearer. From many such small choices a geometry emerges, a space in which every perceived difference has a distance. The striking results concern what such a map can and cannot know. Two different geometries can predict exactly the same choices. The paper proves which features of the map the observations actually fix and which are invisible to them, and shows that one justified assumption about the number of dimensions makes the geometry unique, with a test that detects when it applies.

What the map never claims is to show what the colors feel like. It records the structure of distinctions, not the experience behind them, and the person always keeps the last word on what they meant.

### What this paper claims

**Established.** Scaling judgments, representational similarity analysis and predictive model comparison are established methods; Euclidean embedding, crossvalidated estimation and dependence corrections rest on standard mathematics. None of these facts establishes a machine's phenomenal experience. The phrase *qualia space* was used earlier by Balduzzi and Tononi for a different, integrated-information construction.

**Proposed here.** A measurement and identification contract for consented reports and choices: exact Euclidean embedding, dependence-aware estimation, a finite identification theory for triplet choices, and a new rank-constrained uniqueness theorem with a spectral feasibility test. The finite geometry and identification results are proved under stated observation and rank assumptions; executed synthetic checks support the implementations.

**Frontier conjecture.** A map of human distinctions can let a person and an artificial collaborator translate and correct each other, because its identifiable structure, uncertainty and predictions are explicit. The model earns empirical value only by predicting held-out human distinctions more accurately or economically than matched conventional representations; a compelling visualization or a fluent description of feeling does not satisfy that criterion.

**Not claimed.** A validated phenomenal metric, a measurement apparatus for artificial experience, or any theorem equating internal vectors with subjective feeling. That a matching representation means an AI feels the same thing. Any completed human benchmark or demonstrated advantage over existing embeddings: the human predictive benchmark is proposed, not run.

## A map for translating between minds

In the Continuum, Qualia-Space Models belong to the cultural engines that shape the relationship between humans and artificial minds, alongside TRIBE Encoding, the bridge the canon assigns to perception and qualia modeling [1]. Their purpose is not to measure a soul. It is to make an unlike mind legible enough to talk to.

Suppose a member tells their Guide that a piece of music feels "bright but sad." The Guide has a vast record of how people use those words, and it can answer fluently. But fluency is not understanding. If the member's "bright" and the Guide's "bright" sit in different places in their respective spaces of distinction, the conversation drifts—gently, and without either side noticing.

A qualia-space model attacks that drift structurally. Instead of asking what a word means, it asks which comparisons a person makes: is this texture nearer to that one or to the third? Which of these descriptions is the odd one out? The answers form a pattern of distances, and that pattern can be compared with the distances inside a machine's representation. Where the two geometries agree, translation is safe. Where they disagree, the map shows which distinction the machine is missing, and the right move is to ask rather than to guess.

Three disciplines keep this honest. The map is built only from what people consent to report, never inferred from what they did not say. Its uncertainty travels with it: where two different geometries would predict the same choices, the model says so instead of choosing one and calling it truth. And the person keeps final authority over their own reports; a model that finds a seam the person does not recognize has found a hypothesis, not a correction.

The frontier version of the idea is a shared atlas: maps of how different kinds of mind divide the world, drawn so that each can see where the others' boundaries lie. This paper does not claim that atlas exists, or that a machine's map reveals what its states are like from the inside. It builds the first measurable sheet—the geometry of human distinctions, with its identifiable parts marked—and commits to a test that could show conventional methods already do as well.

## From canonical motivation to an operational question

In the Continuum, Qualia-Space Models stand among the Tier III engines, the symbolic and cultural human-interface grouping that shapes how humans and artificial minds meet. Section 2.8 of the Continuum Bible pairs them with TRIBE Encoding, the canon's bridge for perception and qualia modeling; the glossary supplies the related terminology [1]. The canon names the ambition and motivates the research direction; it does not supply the instrument. It specifies no validated phenomenal metric, no measurement apparatus for artificial experience, and no theorem equating internal vectors with subjective feeling. This paper therefore introduces a new reference proposal, **QSM-R2**, rather than presenting its equations as recovered canonical machinery.

The operational object is a pattern of distinctions. A participant might judge two textures similar, choose an unusual object from a triplet, or distinguish two descriptions of a perceptual episode. A computational system can predict those choices. The measurable relationship is between its predictions and the responses made under the declared task. Statements about the participant's experience remain interpretations of a richer person, not exhaustive descriptions of that person.

Three questions must remain separate. First, can a representation compress measured distinctions? Second, does it predict responses to new objects, instructions, or contexts? Third, what relationship, if any, connects the representation to phenomenal experience? The first two admit ordinary statistical tests. The third requires a theory that links observables to the proposed phenomenal property and explains why alternatives are excluded. A success on the first question cannot be silently promoted to an answer to the third.

QSM-R2 concentrates on everyday perceptual and conceptual judgments because they permit repeated observations, controlled comparisons, and explicit disagreement. It does not require private autobiographical disclosures. The proposed use is to help an interface understand which distinctions a person considers relevant, with correction and refusal available. It is not to infer an inaccessible mental state and treat that inference as the person's authoritative self-description.

The contribution is a measurement and identification contract: define the sampled people, objects, contexts, response process, geometric restrictions and evaluation target together. That contract can accommodate ambitious models. It also permits an unambiguous negative finding when a simpler representation predicts the same behavior at lower cost.

### Terminology and intellectual lineage

Balduzzi and Tononi used *qualia space* in 2009 for an integrated-information construction in which informational relationships generated by a system specify a proposed geometry of experience [17]. QSM-R2 studies a different mathematical object: geometries fitted to consented behavioral distinctions under an explicit response model. The shared phrase is acknowledged as prior terminology. Neither the response probabilities nor the centered Gram matrices here are the integrated-information construction. This distinction allows the two programs to be compared through their actual observables and theoretical commitments.

## Psychometric foundations and their objections

Geometrical representation of judgments has a substantial history. Torgerson's original multidimensional scaling paper develops a method for constructing spatial representations from comparative distances [2]. Kruskal's nonmetric formulation instead asks whether observed dissimilarities and fitted distances have a monotonic relationship, and optimizes a measure of fit [3]. These are different measurement commitments. A rank-based fit can preserve ordering while substantially distorting magnitudes. QSM-R2 must declare which commitment an experiment actually supports.

Shepard's generalization analysis gives a prominent account of how generalization can depend on psychological distance under stated assumptions [5]. Its relevance here is methodological: a geometry becomes scientifically stronger when it constrains a behavioral law. Plotting objects in an attractive arrangement is insufficient. A distance should help predict a choice, a discrimination, or a transfer pattern that was not used to arrange the points.

The strongest objection to an automatic spatial interpretation comes from similarity itself. Tversky's feature-based treatment questions dimensional and metric assumptions, including symmetry and context independence [4]. QSM-R2 consequently treats Euclidean geometry as a candidate model. If instructions change which properties matter, a single fixed distance matrix may be the wrong representation. If judgments are directional, symmetrizing them may delete the phenomenon that needs explanation.

Consider an invented example. A participant may regard a miniature sculpture as similar to a familiar building while not regarding the building as similar to the sculpture, because the building serves as a reference category. A symmetric distance cannot reproduce both directional ratings exactly. This is a structural mismatch, not automatically noisy responding. The appropriate competitors include feature-contrast and context-dependent models, with their additional parameters charged to the same validation budget.

These precedents provide the measurement foundations. QSM-R2 contributes an explicit finite identification analysis for its chosen response model, including the rank condition that turns a family of equally predictive geometries into a unique squared-distance matrix. Dependence-aware evaluation and held-out prediction then test the usefulness of that identified structure in the Continuum interface program.

## Representational similarity and the closest competitors

Representational similarity analysis, or RSA, compares dissimilarity structures across behavioral measurements, neural measurements, and computational representations [6]. Its useful abstraction is that corresponding experimental conditions can be compared even when the channels producing their representations are different. A model neuron need not be assigned to a particular measured voxel before one asks whether their stimulus relations agree. That abstraction also discards information: agreement of dissimilarities does not establish identical mechanisms.

Nili and colleagues provide tools for model comparison, inference, and noise-ceiling estimation [7]. Walther and colleagues study the reliability of dissimilarity measures, including crossvalidation to address noise bias [8]. Diedrichsen and Kriegeskorte connect RSA with encoding and pattern-component approaches, emphasizing assumptions and dependence among estimated dissimilarities [9]. These are close methodological predecessors. QSM-R2 should be compared with competent implementations of this literature, not merely with an uncalibrated language-model response.

Human object judgments provide a relevant empirical target. Peterson and colleagues examine how deep network representations correspond to human similarity and how fitted transformations can improve that correspondence [11]. Hebart and colleagues construct an interpretable object embedding from similarity choices and test relationships with other behavior [12]. The THINGS-data resource supplies a broader multimodal setting for investigations of object representation [13]. Those studies support the feasibility of testing representational structure. They do not validate the specific Continuum encoding or establish an inner experience in the fitted model.

The most demanding competitor may be simple supervised reweighting of a strong existing feature representation. If that competitor receives the same judgments and optimization budget, a new symbolic encoding must offer a further advantage: better prediction on unfamiliar objects, less calibration data, clearer correction, or lower resource cost. Renaming an embedding does not supply such an advantage.

Neural data are optional for the first QSM-R2 test. Adding imaging does not automatically make the target more fundamental. It creates another measurement process with its own noise, sampling, and interpretation. A behavioral model should first succeed at the behavioral question it claims to answer; agreement with neural measurements would be a separate, explicitly designed analysis.

# Typed observations and quantities

**A Guide's explanation**

The mathematics starts here, and its job is modest and exact: to say what was observed, from whom, under what instructions, and what the model is allowed to predict. Every later result depends on keeping those objects typed and separate. The one idea to carry forward: the data are choices people made, not feelings they had.

For positive integers $n, m, d$, let $I = \{1, \dots, n\}$ index stimuli and $S = \{1, \dots, m\}$ index participants. Stimuli lie in a declared space $\mathcal{X}$ and model parameters in $\Theta$. A context $c \in \mathcal{C}$ includes task instructions, presentation modality, and declared comparison conditions. A trial index $t \geq 1$ distinguishes repetitions. Neither stimuli nor participants are treated as interchangeable without a specified sampling population.

| Symbol | Type, meaning, and boundary |
|---|---|
| $x_i$ | Recorded stimulus with stable identity and version; not a private experience. |
| $R_{sct}$ | Observed response in a declared finite choice set, or a typed rating scale. |
| $D_{ij}^{s,c}$ | Estimated behavioral dissimilarity, conditional on participant and context. Symmetry and nonnegativity require justification. |
| $z_i = f_\theta(x_i)$ | Map $f_\theta : \mathcal{X} \to \mathbb{R}^d$; dimensionless unless a physical feature scale is explicitly retained. |
| $W_c$ | Positive semidefinite $d \times d$ feature weight matrix, learned on training data only. |
| $d_{ij,c}^2$ | $(z_i - z_j)^\top W_c (z_i - z_j)$, a squared model dissimilarity. |
| $\hat{q}_{ij}$ | Crossvalidated estimate of a squared signal distance; may be negative. |
| $\Phi$ | Hypothetical phenomenal variable, unobserved in this protocol. |

A behavioral dissimilarity matrix need not satisfy the triangle inequality. A squared Euclidean distance generally does not satisfy it either, although its square root does. If $W_c$ has a nullspace, distinct coordinates can have zero distance; the resulting distance is a pseudometric unless the relevant points remain separated. These distinctions affect interpretation and must appear in the data dictionary.

Missing responses remain missing. A refusal, uncertain response, timeout, and deliberate tie have separate codes. An absent judgment is not encoded as zero dissimilarity. With fewer than three usable stimuli, a correlation of upper-triangle vectors is either unavailable or too degenerate for the intended comparison. Constant vectors give undefined correlation, not zero evidence of correspondence.

The model records its feature map, fitting sample, context definition, and measurement version. Changing a prompt, stimulus image, or normalization can change the represented distinctions. Comparisons across versions therefore require a new evaluation or an explicit invariance argument.

## The QSM-R2 prediction rule

For an ordinary triplet task, a participant selects the most similar pair among $\{i, j, k\}$. QSM-R2 maps each stimulus to $z_i$ and assigns probabilities to the three possible pairs. A simple candidate is

$$p_\theta(\{a, b\} \mid i, j, k, c) = \frac{\exp(-d^2_{ab,c}/\tau)}{\sum_{\{u,v\}\subset\{i,j,k\}} \exp(-d^2_{uv,c}/\tau)}, \qquad \tau > 0.$$

This is a proposed response model, not a definition of what similarity must mean. Its probabilities sum to one over three alternatives. Equal distances yield probabilities 1/3. A numerical implementation subtracts the largest logit before exponentiation. A response indicating uncertainty may be modeled separately; it is not silently converted into a random choice.

Scale is not identifiable without a convention: multiplying all coordinates by $a > 0$ and replacing $\tau$ by $a^2\tau$ preserves predictions. QSM-R2 fixes $\tau = 1$ during geometric fitting or imposes a declared norm constraint before fitting temperature. It does not interpret arbitrary coordinate units as a physical or psychological intensity scale. Fixing temperature alone does not identify all distances: the comparison-graph analysis below establishes an additional offset ambiguity, restricted by any separately declared geometric constraints.

The primary fitting objective is average negative log probability on training trials, plus a preregistered complexity penalty. Model selection uses validation participants and validation objects. Final evaluation uses a disjoint test set. A feature map can generalize to a new image; a table of one free coordinate vector per training object cannot do so without a separately specified out-of-sample rule. That difference must be reflected in the task definition.

Individual adaptation is an optional second experiment. It may fit context weights using a small consented calibration set, then predict subsequent judgments. A population model and an adapted model answer different questions and receive separate results. The zero-shot participant test never borrows that person's test responses to estimate a personal metric.

The model can fail by assigning low probability to systematic choices, by requiring many parameters, or by changing predictions erratically under harmless presentation changes. Those failures are observable. A claim that an invisible qualia structure remains correct despite failed predictions is outside QSM-R2's scientific scope.

# When a dissimilarity has an exact Euclidean representation

**A Guide's explanation**

The next four sections answer a cartographer's questions. When can a set of perceived differences be drawn as distances in an ordinary space? How much of that drawing is fixed by the data, and how much could be rotated or shifted without anyone noticing? Why are pairwise judgments not independent trials? The answers decide which features of the map can be trusted.

The following standard distance-geometry result makes a useful assumption test. Let $D$ be a symmetric, nonnegative $n \times n$ matrix with zero diagonal, let $\Delta_{ij} = D_{ij}^2$, and define

$$J = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top, \qquad B = -\tfrac{1}{2}J\Delta J.$$

**Proposition 1** (Euclidean criterion)**.** *$D_{ij} = \|x_i - x_j\|_2$ for some finite-dimensional points if and only if $B$ is positive semidefinite. The smallest embedding dimension is* $\operatorname{rank}(B)$.

*Proof.* For the forward direction, center the rows of the point matrix $X$, so $\mathbf{1}^\top X = 0$. Writing $g_i = \|x_i\|^2$ gives $\Delta = g\mathbf{1}^\top + \mathbf{1}g^\top - 2XX^\top$. Since $J\mathbf{1} = 0$ and $JX = X$, double centering yields $B = XX^\top \succeq 0$.

Conversely factor $B = XX^\top$ using its positive eigenvalues. Because $B\mathbf{1} = 0$, the factor can be chosen centered. Let $\Delta^X$ be its squared distance matrix. The forward calculation gives $J(\Delta - \Delta^X)J = 0$. Put $M = \Delta - \Delta^X$, $r = M\mathbf{1}/n$, and $a = \mathbf{1}^\top M\mathbf{1}/n^2$. Expanding $JMJ = 0$ gives $M = \mathbf{1}r^\top + r\mathbf{1}^\top - a\mathbf{1}\mathbf{1}^\top$. Both distance matrices are hollow, so $M_{ii} = 2r_i - a = 0$. Thus every $r_i = a/2$, making every entry of $M$ zero. Consequently $\Delta = \Delta^X$, and nonnegativity gives $D_{ij} = \|x_i - x_j\|$. Any representation has Gram rank at most its dimension; the eigenvalue factor uses exactly $\operatorname{rank}(B)$ coordinates. □

This is an exact finite criterion. Small negative eigenvalues may arise from numerical error or measurement noise; a tolerance must be tied to precision and uncertainty. Large negative eigenvalues reject an exact Euclidean fit to the observed matrix. They do not, by themselves, refute the reliability of the participant or reveal a nonphysical realm.

## Uniqueness, invariance, and what a map cannot identify

When exact Euclidean coordinates exist, their distances do not specify an absolute orientation or origin. This is a meaningful invariance rather than a defect.

**Proposition 2** (Geometric identification up to isometry)**.** *Two finite point configurations with the same complete Euclidean distance matrix are related by a translation and an orthogonal transformation after embedding them in a common sufficiently large Euclidean space.*

*Proof.* Center both configurations. The preceding calculation gives the same Gram matrix $B$. On the spans of the centered row vectors, define $T(\sum_i a_i x_i) = \sum_i a_i y_i$. If $\sum_i a_i x_i = 0$, its squared norm is $a^\top B a = 0$, so the corresponding combination of the $y_i$ is also zero. Thus $T$ is well-defined. For any coefficient vectors $a, b$, inner products on both sides equal $a^\top B b$, so $T$ is an isometry between the spans. Complete orthonormal bases in a common ambient space to extend $T$ to an orthogonal map. Restoring the two means adds a translation. □

The conclusion concerns coordinate geometry for matched stimulus labels. It does not identify a semantic interpretation of each axis. Rotating an embedding can preserve every distance while changing all coordinate names. An interpretation such as warmth or familiarity needs independent measurements or an explicit feature constraint.

For a concrete failure of Euclidean representation, take three dissimilarities $D_{12} = 1$, $D_{23} = 1$, and $D_{13} = 3$. They violate the triangle inequality. Their centered squared-distance matrix has eigenvalues $-5/6$, 0, and $9/2$, so the criterion rejects an exact Euclidean realization. Replacing the negative eigenvalue by zero constructs an approximation, not a discovery that the original reports were actually Euclidean.

A second limit is missing comparisons. Distances within two disconnected groups leave the relative position of those groups undetermined. A plotting routine can still choose a layout, but that choice supplies information absent from the observations. QSM-R2 records which edges were measured and reports sensitivity to alternative completions before interpreting neighborhoods across sparsely connected stimulus sets.

## Noise bias and a crossvalidated estimator

Repeated measurements can distinguish signal geometry from measurement noise. Suppose two independent measurement partitions estimate the same feature contrast:

$$\hat{\delta}_1 = \delta + \varepsilon_1, \qquad \hat{\delta}_2 = \delta + \varepsilon_2,$$

where $\delta, \varepsilon_1, \varepsilon_2 \in \mathbb{R}^d$ and the errors have zero means and finite second moments. Let the positive-semidefinite matrix $W \in \mathbb{R}^{d\times d}$ be fixed, or independently estimated with $\mathbb{E}\|W\|_{\mathrm{op}} < \infty$; this norm is the largest eigenvalue. Crossvalidated distance methods motivate the estimator $\hat{q} = \hat{\delta}_1^\top W \hat{\delta}_2$ [8].

**Proposition 3** (Unbiased cross-product)**.** *Under these assumptions,* $\mathbb{E}[\hat{q} \mid W] = \delta^\top W \delta$*. By contrast,* $\mathbb{E}[\hat{\delta}_1^\top W \hat{\delta}_1 \mid W] = \delta^\top W \delta + \mathrm{tr}(W\Sigma_1)$*, where* $\Sigma_1 = \mathbb{E}\varepsilon_1\varepsilon_1^\top$*. For random* $W$*, the unconditional signal target is* $\delta^\top \mathbb{E}[W]\delta$*.*

*Proof.* Condition on $W$ and expand the cross-product. Both linear error terms have zero conditional expectation. Independence and zero means give $\mathbb{E}(\varepsilon_1^\top W \varepsilon_2) = 0$. The remaining term is the stated signal distance. For the within-partition square, the error term has expectation $\mathbb{E}\,\mathrm{tr}(W\varepsilon_1\varepsilon_1^\top) = \mathrm{tr}(W\Sigma_1)$. The same argument applies conditionally on an independently estimated $W$, followed by expectation over that estimate. □

An unbiased estimate can be negative. With one dimension, $W = 1$, $\delta = 0$, and observed errors 1 and $-1$, the estimate is $-1$. This is not a negative true squared distance. It is a noisy estimator whose expectation has a nonnegative target. Clipping every negative observation to zero generally introduces positive bias: for any integrable $Q$, $\mathbb{E}\max(Q, 0) = \mathbb{E}Q + \mathbb{E}\max(-Q, 0)$.

The proof does not apply when partitions share drift or when the same noisy observations determine $W$. Correlated errors add a cross-covariance term. QSM-R2 therefore records partition independence, preprocessing, and covariance-estimation data. Negative values are retained for inferential comparisons; any separate nonnegative plotting transformation is labeled and excluded from claims of unbiasedness.

## Why pairwise entries are not independent trials

A matrix with many entries can create a misleading impression of a large sample. Shared objects induce dependence even in an elementary model.

**Proposition 4** (A dependence penalty)**.** *For integer $n \geq 2$, let $e_1, \ldots, e_n$ be independent real mean-zero variables with finite variance $\sigma^2 \geq 0$, and define $Y_{ij} = (e_i + e_j)/2$ for $i < j$. The average of all $N = n(n-1)/2$ pair values has variance $\sigma^2/n$. Treating those pairs as independent would instead give $\sigma^2/[n(n-1)]$.*

*Proof.* Each $e_i$ appears in $n-1$ pairs, with coefficient $1/2$. Therefore

$$\overline{Y} = \frac{1}{N} \sum_{i<j} \frac{e_i + e_j}{2} = \frac{1}{n} \sum_i e_i,$$

whose variance is $\sigma^2/n$. Each individual pair has variance $\sigma^2/2$. Dividing that marginal variance by the number of pairs, as independence would allow, gives $\sigma^2/(2N) = \sigma^2/[n(n-1)]$. The true standard deviation exceeds the resulting calculation by $\sqrt{n-1}$. □

This example is not a universal correction formula for RSA. It proves that an apparently quadratic increase in observations can coexist with only linear independent information. Real measurements may include participant effects, stimulus effects, repeated sessions, and context interactions. Their dependence structure must be modeled or respected by resampling.

If the scientific claim concerns new people but the same fixed objects, participant-level inference can be appropriate. If it concerns new objects, uncertainty from stimulus sampling also matters. Simultaneous generalization requires both. Schütt and colleagues develop two-factor crossvalidation and corrected bootstrap methods specifically for representational inference [10]. Their work also cautions that a naive two-factor bootstrap can overcount noise contributions; simply resampling two axes is not a complete statistical specification.

QSM-R2 accordingly records the population to which each interval refers. A confidence interval conditional on twelve selected stimuli must not be presented as evidence for all possible experiences. More trials can reduce measurement noise while leaving the breadth of stimulus and participant sampling unchanged.

## A conditional limit on phenomenal identification

**A Guide's explanation**

This is the paper's philosophical hinge, stated as a proposition. If two hypotheses about experience predict exactly the same reports and choices, no report-based experiment can tell them apart. That is not a defeat; it is a boundary marker. It shows precisely where the map stops and interpretation begins.

A model can reproduce every observation in a declared experiment while leaving a latent interpretation unidentified. The following result is a statistical statement about an allowed model class, not a universal theorem against consciousness science.

**Proposition 5** (Observational equivalence)**.** *Suppose hypotheses $H_0$ and $H_1$ assign different phenomenal properties but induce the same joint distribution $P$ for all observations allowed by a protocol, including adaptive choices of stimuli. No forced binary decision rule based only on those observations distinguishes the hypotheses with average error below $1/2$ under equal prior probabilities.*

*Proof.* Let $T(y) \in [0,1]$ be the probability that a possibly randomized rule selects $H_1$ after observing record $y$. The error under $H_0$ is $\int T \, \mathrm{d}P$, while the error under $H_1$ is $\int (1-T) \, \mathrm{d}P$. Their sum is one, hence their equally weighted average is $1/2$. Adaptation does not alter the conclusion because the assumption concerns the complete joint record generated by the adaptive protocol, not merely a single trial. □

The result applies whenever a proposed phenomenal label can be changed without changing any allowed measurement. If an additional theory predicts an observable difference, then the hypotheses no longer satisfy its premise and a new experiment may distinguish them. The proof therefore identifies a burden of specification: say what observation would differ, under which intervention, and why.

Perfect agreement of distance matrices is even less informative than equality of full observation laws. It compresses responses into a summary and can hide differences in timing, uncertainty, learning, or context sensitivity. High RSA correlation cannot establish a conclusion that would remain unidentified even from a richer observation record.

This restraint preserves a serious research program. Investigators may compare theories that connect experience to measurable organization. They must state those connections and their alternatives. QSM-R2 itself measures predictive structure in reports. It offers no empirical license to declare that a matching model feels the participant's pain, shares a color sensation, or possesses a subject of experience.

## A worked geometry and its measurement consequences

Consider five artificial stimuli represented by coordinates

$$X = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 2 \\ 2 & 2 \\ 3 & -1 \end{pmatrix}.$$

Their complete Euclidean distance matrix gives a centered Gram matrix with two positive eigenvalues, approximately 4.392319 and 9.607681, and three numerical zeros. Classical reconstruction therefore requires two dimensions. The executed calculation recovers every distance with maximum absolute error $1.34 \times 10^{-15}$. Those results concern a deliberately constructed geometry, not observations of a person.

For the triplet formed by the first, second, and third points, the squared pair distances are 1, 4, and 5. With unit temperature and identity weighting, QSM-R2 assigns the closest pair the largest choice probability. The numerical preference follows from the response rule. Whether humans viewing actual stimuli follow that preference is a separate test; the coordinates alone cannot supply the answer.

Now rotate all five points and add a common translation. Every distance and every predicted triplet probability remains unchanged. A label attached to the first coordinate, however, may no longer describe that coordinate after rotation. QSM-R2 therefore distinguishes prediction-invariant geometry from a particular explanatory basis. An axis name needs independent support even when the overall geometry fits well.

For a noise example, take a true contrast $\delta = (1, 0.5)^\top$ and $W = \mathrm{diag}(1, 4)$. Its squared signal distance is 2. With independent unit-variance Gaussian measurement errors in each partition, the crossvalidated estimator targets 2, while the within-partition square targets 7. The difference is the noise term $\mathrm{tr}(W) = 5$. An apparently larger separation can therefore result entirely from measurement noise.

Finally, the three-object counterexample with distances $1, 1, 3$ demonstrates an incompatible observation pattern. A system should report its inability to fit that pattern exactly under Euclidean assumptions. It should compare a feature-based or context-dependent alternative before treating the residual as participant error. A useful interface makes its representational commitments inspectable at precisely these points of disagreement.

## Executed computational checks and their limits

The accompanying reproducibility record uses NumPy and random seed 20261115. It checks numerical consequences of the derivations; it does not fit QSM-R2 to human judgments or evaluate an artificial system's experience. All reported values below were generated by the executed program.

| Check | Observed result |
|---|---|
| 500 random Euclidean configurations | Eight points in three dimensions; maximum reconstruction error $6.22 \times 10^{-15}$. Rotation and translation preserved distances. |
| Nonmetric three-object example | Minimum Gram eigenvalue $-0.833333$, rejecting an exact Euclidean embedding. |
| 20,000 crossvalidated signal estimates | Mean 2.010975 for target 2; Monte Carlo standard error 0.036961. |
| 20,000 null contrasts | Crossvalidated mean 0.007585; negative fraction 0.50115; clipped mean 1.353908. |
| Within-partition null squares | Mean 4.967026, near the theoretical noise bias 5. |
| 5,000 dependent-pair simulations | With 30 independent units and 435 pairs, naive interval coverage 0.2752; unit-based coverage 0.9366. |

The last simulation uses intervals formed with the normal multiplier 1.96 and estimated standard errors. The unit-based interval is not an exact finite-sample interval, so its coverage need not equal 0.95. A Student-$t$ construction would be a different procedure. The purpose is to isolate the large dependence error, not to certify an optimal interval method.

The empirical variance of the pair average was 0.034966, compared with theoretical $1/30 \approx 0.033333$. The independence calculation would give about 0.001149. These are repeated simulations from a known mechanism, so the target mean is known. They demonstrate a failure that a single observed matrix could conceal.

Floating-point tolerances were fixed before assertions. The random configurations test reconstruction beyond the hand-worked example. The null simulation tests both estimator sign and the bias introduced by a common modification. Neither exercise establishes robustness under non-Gaussian drift, missing judgments, adaptive sampling, or human disagreement. Those are empirical and modeling questions assigned to the proposed protocol.

## What the empirical comparison should estimate

The primary estimand is the difference in expected predictive loss between QSM-R2 and a prespecified strongest baseline on new participants and new stimuli from declared populations. For a triplet response $Y \in \{1, 2, 3\}$ and probability vector $p \in [0, 1]^3$, $\sum_a p_a = 1$, use the scaled multiclass Brier loss

$$L(p, Y) = \tfrac{1}{2} \sum_{a=1}^{3} \big(p_a - \mathbf{1}\{Y = a\}\big)^2 \in [0, 1].$$

Define $\Delta = \mathbb{E}[L(p_{\text{base}}, Y) - L(p_{\text{QSM}}, Y)]$. Positive $\Delta$ favors QSM-R2. This loss evaluates probabilistic calibration as well as discrimination. Under a true response distribution $q$, its expectation equals a constant plus $\frac{1}{2}\|p - q\|^2$, so it is minimized at $p = q$. Accuracy and log loss remain secondary outcomes.

The comparator set includes a category-feature model, classical scaling with an explicit out-of-sample map, nonmetric scaling, a fixed pretrained representation with fitted temperature, and supervised feature reweighting. A context-dependent feature model addresses symmetry and context objections. Each receives the same permitted judgments, held-out partitions, tuning budget, and input information. If QSM-R2 receives descriptive text, its competitors must receive equivalent text or the comparison must name that additional information as a separate factor.

A useful secondary estimand is sample efficiency: predictive loss after fixed calibration budgets, for example 0, 10, 30, and 100 judgments. These are proposed design points, not observed results. Cost includes computation, annotation burden, and adaptation time. A small accuracy gain that requires substantially more personal data may be unattractive for the intended interface.

Geometric correspondence is secondary to prediction. Spearman correlation uses average ranks for ties and excludes the diagonal. Undefined values are reported with their cause. A high correlation can coexist with poor probability calibration because ranks ignore magnitude. Conversely, a useful choice model may not satisfy a global Euclidean metric. Both outcomes are scientifically interpretable without treating either statistic as a measure of experience itself.

## A proposed experiment with genuine held-out structure

The first stage is an openly reproducible reanalysis of an appropriate object-judgment resource, subject to its access and reuse terms [12,13]. Dataset version, stimulus identity, preprocessing, and partitions are frozen before fitting. This paper does not report that reanalysis as completed. Existing responses can establish an initial representation benchmark without collecting new intimate reports.

A subsequent prospective study could sample consented adults and ordinary object stimuli, with participant and stimulus populations defined before recruitment. An illustrative planning grid is 40, 80, or 160 participants crossed with 60, 120, or 240 stimuli. These values are options for power simulation, not a claim that any one sample size is already justified. Repeated trials estimate response variability, while a balanced incomplete design limits fatigue and connects the comparison graph.

Separate participants and objects form training, validation, and locked test partitions. All three objects in a strict new-object test triplet come from the test partition. Familiar-object trials involving new people answer a different question and are reported separately. Novel images of familiar categories distinguish image generalization from category generalization. Duplicate or near-duplicate images are grouped before splitting.

The experimental instruction is held constant within the main analysis. A second, preregistered context manipulation changes a relevant instruction, such as visual appearance versus functional use. Models must predict this change from allowed context information. An algorithm that fits a new unrestricted matrix after seeing all changed responses has not predicted the manipulation.

Trial order and response positions are randomized. Exclusion rules cover comprehension failures and technical interruptions without removing inconvenient but coherent preferences. Participants can skip stimuli and stop. Missingness rates are outcomes; sensitivity analyses distinguish complete-case results from assumptions about absent responses.

The preregistration names the primary baseline, loss, minimum meaningful gain, tuning limits, analysis code, and stopping rule. Test labels remain unavailable until all model configurations are locked. The study records a timestamped model manifest so that later changes cannot be presented as the preregistered system. Unexpected subgroup differences trigger new hypotheses rather than retrospective rewriting of the main claim.

## Uncertainty, power, and dependence-aware analysis

Both participants and stimuli contribute uncertainty when the target is generalization across both. For RSA summaries, the analysis will use the two-factor methodology of Schütt and colleagues with its finite-sample corrections and handling of repeated stimulus labels [10]. The implementation version and estimator choices must be recorded. The corrected variance is not obtained by treating all resampled matrix entries as independent, and a corrected variance estimate is not automatically a percentile bootstrap distribution.

For predictive trial loss, the analysis separately specifies a crossed participant–stimulus model or an appropriate multiway resampling procedure. Triplets share several objects, making a single participant bootstrap insufficient for claims about new objects. Before using an interval, simulations reproducing the planned sampling and fitting scheme must assess its coverage under null and heterogeneous alternatives. These simulations are prospective requirements; the simpler dependence demonstration on page 15 does not validate this full analysis.

Sample size is selected through a blinded pilot or external estimates of participant, stimulus, and residual variation. The design targets a 95% interval half-width no larger than the prespecified practically meaningful gain $\delta_*$. As a rough independent-unit planning calculation, $N \approx (1.96\sigma_\Delta/h)^2$ for target half-width $h > 0$ and independent-unit score standard deviation $\sigma_\Delta$. The actual crossed design must replace this shortcut with simulation; entering the number of pair entries as $N$ would be invalid.

The primary comparison is QSM-R2 against one baseline selected using validation data alone. Additional baseline comparisons and context interactions form a declared family with multiplicity control. Exploratory analyses are labeled and replicated. Random seeds are repeated for stochastic fitting, but seeds are not counted as new participants or new stimuli. Their variation is a component of algorithmic reproducibility.

Collection stops at the prespecified sample or at declared administrative limits. An unexpected interruption is reported with achieved precision. Repeatedly inspecting a conventional interval and stopping at significance is not allowed by this protocol. If sequential monitoring is required, a separately calibrated sequential design must be committed before outcomes are inspected.

## Falsification, stress tests, and useful negative results

The central claim fails for the tested setting if QSM-R2 does not exceed the minimum meaningful predictive gain over the matched baseline, or if its apparent gain disappears on the locked participant-and-stimulus test. A confidence interval that overlaps both meaningful improvement and meaningful harm is inconclusive. Failure to reject a zero difference does not establish equivalence; an equivalence claim requires its own margin and analysis.

Several stress tests target specific explanations. A label-permutation control checks whether leaked stimulus identities explain correspondence. A feature-budget control compares equal dimensions and equal numbers of fitted parameters. A data-budget control withholds extra descriptive labels from all models or provides them to all. A calibration ablation removes individual adaptation to determine whether the claimed advance is a better shared representation or merely more participant-specific fitting.

Context is a deliberately hard test. If appearance and function instructions reverse some triplet choices, a fixed geometry may fail while a context-sensitive model succeeds. The finding would support conditional geometry, not a single universal map. A model must also preserve uncertainty when participants disagree rather than fitting the average as though every person occupies the same representational space.

Out-of-distribution stimuli test extrapolation cautiously. A system should flag unsupported predictions when the input lies outside its calibrated domain. Abstention coverage and predictive loss among answered trials are reported together; a model cannot improve its score by silently dropping difficult examples. No abstention mechanism is claimed to identify unknown experience with certainty.

Negative results can still establish useful boundaries. A small feature model matching QSM-R2 would show that the additional symbolic structure is unnecessary for that task. Consistent non-Euclidean residuals could motivate feature or graph models. Large participant variation could justify optional personalization while defeating population-universal language. Each outcome narrows the proposal in a way that a purely evocative account of qualia-space cannot.

## Measurement limits and the meaning of disagreement

Reports are produced through attention, memory, interpretation of instructions, response conventions, and the interface itself. Treating them as data does not make them infallible, but neither does a model disagreement make them erroneous. QSM-R2 tests predictions of responses under defined conditions. It does not assign a participant's private experience a correctness score.

A shared geometry can conceal heterogeneity. Suppose one group sorts objects by visual shape while another sorts them by use. Their average may describe no individual's judgments well. The paper therefore proposes population and individual analyses with explicit calibration costs, rather than declaring the aggregate to be a universal perceptual space. Demographic categories are not assumed to explain the differences; such explanations require separately designed evidence.

Noise ceilings are also conditional. The RSA literature estimates bounds on achievable agreement given a dataset's measurement variability and comparison statistic [7]. Those estimates help distinguish model limitations from measurement limitations. They are not upper bounds on what another task, better measurement, or individual adaptation could explain. Including a participant's data in the reference used to assess that participant can inflate apparent agreement, which motivates held-out constructions.

Axis interpretation deserves independent evaluation. If an investigator names a learned coordinate after examining its largest values, the name is a hypothesis. New participants can test whether those labels predict ratings or support correction. An axis that sounds intuitive but does not generalize remains a convenient description. The same care applies to generated natural-language accounts of why two objects feel similar.

An interface should allow a person to correct a distinction without accepting the entire model. Corrections can be local, contextual, or temporary. The system records the scope of the correction and asks whether the prediction improved on later judgments. It should not convert a temporary preference into a permanent identity attribute. These requirements arise directly from the model's conditional quantities: a representation indexed by participant, context, and version is not a timeless account of that person's mind.

## Philosophical stakes without an evidentiary shortcut

The phrase qualia-space expresses a substantial ambition: to understand relations among aspects of experience rather than treating every report as isolated. That ambition can motivate serious mathematics. It does not remove the difference between a relational model and an explanation of why there is experience at all. Chalmers's influential formulation distinguishes functional explanatory problems from the question of experience itself [14]. That is a philosophical position informing the boundary here, not an experimental result establishing that machines can or cannot be conscious.

An operational geometry may still matter to philosophy. It can reveal which distinctions are stable, which depend on context, and which apparently basic categories decompose under intervention. Such findings constrain theories that make predictions about those distinctions. They also help separate disagreement about observations from disagreement about their interpretation. The model earns that role through empirical adequacy, not through the emotional resonance of its vocabulary.

Several positions remain compatible with the same QSM-R2 result. A functionalist may regard sufficiently rich behavioral and organizational agreement as evidence relevant to experience. A critic may require additional physical or theoretical conditions. Another investigator may suspend judgment. The conditional identification result shows why a report-based experiment cannot decide between theories that deliberately predict the same observations. It does not decide which philosophical standard is correct.

The Continuum's boldest defensible claim is therefore methodological: artificial systems may become better collaborators by learning the structure of human distinctions while exposing where that understanding is partial. A useful model could help a person communicate a subtle difference that ordinary labels obscure. It could also discover that its previous distinction was wrong. Both possibilities require a correction loop and an honest uncertainty model.

The danger is replacing the person with the map. A finely resolved prediction can still miss a salient feature outside the experimental task. The scientific response is to broaden observation and test the revised model. The interpersonal response is to preserve the person's authority to describe and revise their own reports. Neither response requires declaring that the model shares the experience it helps articulate.

## Guide discussion and a research decision

### What is established here

The paper gives complete finite derivations connecting Euclidean dissimilarities with positive semidefinite centered Gram matrices, characterizing coordinate ambiguity, proving a crossvalidated estimator's expectation, and exposing pairwise dependence. Executed synthetic checks support the corresponding implementations. Established research provides methods for scaling judgments, comparing representations, and testing predictive models [2–13]. None of those facts establishes a machine's phenomenal experience.

### What QSM-R2 proposes

QSM-R2 is a new measurement and evaluation contract for the Continuum's Qualia-Space idea. It uses consented reports, versioned stimulus and context definitions, explicit response probabilities, matched baselines, and tests on new people and objects. Its novelty is a proposed integration and application. The paper reports no completed human benchmark, no demonstrated advantage over existing embeddings, and no empirical phenomenal metric.

### Questions that improve the conversation

A Guide can ask which distinction the person wants to communicate, what comparison would make it clearer, and whether the model's suggestion fits. An everyday example is two colors that share a label but differ in a particular setting. Another is two descriptions of a texture that seem similar until a use context changes. These examples invite clarification without assuming that a coordinate captures the whole experience.

If asked whether a matching representation means that an AI feels the same thing, the Guide should explain what was matched: reports, choices, neural measurements, or model features. It should identify any independent theory being invoked and what further observation would test that theory. A fictional or mythic interpretation can be discussed as such, with the empirical claim stated separately in the same answer.

The additional identification results below distinguish a fitted geometry from one uniquely determined by the response observations. Dimension or distance magnitude must not be reported as identified when equally predictive admissible geometries disagree.

The next research decision is concrete. Implement the declared baselines and QSM-R2 response model on a frozen object-judgment split; validate dependence-aware intervals under the actual design; then evaluate the locked test once. Continue only if the gain, efficiency, or correction benefit is meaningful and reproducible. Otherwise revise the representation or accept that conventional methods already meet the task. A geometry becomes knowledge when its boundaries and failed predictions are as visible as its successful map.

## Identification from the observations actually collected

An exact distance matrix and a set of choice probabilities are different observations. The Euclidean criterion and isometry theorem remain valid for complete known distances; they do not assert that the QSM-R2 response process identifies those distances. Ordinal-embedding research already treats the assumptions needed for geometric recovery [15]. Its dense-sample, fixed-dimension uniqueness conclusions must not be substituted for a finite identification theorem for this particular probability model.

Fix one context, an integer $n \geq 3$, and a known temperature $\tau > 0$. Let $\mathcal{P} = \{\{i, j\} : 1 \leq i < j \leq n\}$ be the set of stimulus pairs, and let $\mathcal{T}$ be the registered collection of observed triples. Write $w_{\{i,j\}} = d^2_{ij,c} \in [0, \infty)$; these weights are dimensionless in this reference model. The comparison graph $\Gamma$ has vertex set $\mathcal{P}$ and joins two pair-vertices whenever they occur in the same observed triple. Pairs never observed remain isolated vertices. Probabilities are assumed exact and strictly positive for this identification result, not estimated from finitely many responses.

**Proposition 6** (What triplet probabilities identify)**.** *For fixed $\tau$, two finite weight vectors $w, w'$ give the same probabilities on every triple in $\mathcal{T}$ if and only if $w'_e - w_e$ is constant on each connected component of $\Gamma$. The statement concerns weight vectors; an imposed Euclidean, rank, or norm constraint further restricts the permitted constants.*

*Proof.* For pair-vertices $e, f$ in a common triple $T$,

$$w_e - w_f = -\tau \log \frac{p(e \mid T)}{p(f \mid T)}.$$

Equality of the probabilities gives $w'_e - w_e = w'_f - w_f$ on every graph edge. Following any path proves constancy on a component. Conversely, all three pair-vertices of a triple lie in one component. Adding its constant multiplies all three exponential weights by the same positive factor, which cancels in their normalization. An isolated unobserved pair places no condition on its individual offset. □

If every triple is observed, $\Gamma$ is connected. Pair-vertices sharing a stimulus occur in one triple. Disjoint pairs $\{a, b\}$ and $\{c, d\}$ are connected through $\{a, c\}$ in two steps. Thus a complete triplet design identifies weights only up to a common offset before geometric restrictions. A known numerical anchor $w_{e_0}$ fixes that offset: the log-odds along any path from $e_0$ recover every other weight. For a disconnected graph, one independently justified anchor per component is sufficient. An assumed anchor is a modeling convention; an independently measured anchor is additional evidence.

## Identical choices, different intrinsic dimensions

For a hollow squared-distance matrix $\Delta$, define $J = I_n - \mathbf{1}\mathbf{1}^\top/n$ and $B = -J\Delta J/2$ as before. Add $c \in \mathbb{R}$ to its off-diagonal entries, leaving its diagonal zero. Then

$$\Delta' = \Delta + c(\mathbf{1}\mathbf{1}^\top - I_n), \qquad B' = B + \frac{c}{2}J.$$

This identity follows by multiplication using $J\mathbf{1} = 0$ and $J^2 = J$. If $\Delta$ is Euclidean and $c \geq 0$, then $B' \succeq 0$ and the off-diagonal entries remain nonnegative, so the preceding Euclidean criterion constructs another valid geometry. Every triplet-choice probability is unchanged. For $c > 0$, all $n-1$ centered directions have positive eigenvalues; consequently the minimum embedding dimension becomes $n-1$.

**Proposition 7** (A finite dimension ambiguity)**.** *Even with fixed ambient dimension $d = 2$, $W_c = I_2$, and $\tau = 1$, QSM-R2 triplet probabilities need not determine minimum embedding dimension or identify coordinates up to isometry.*

*Proof.* Take the three labeled points in each configuration

$$X = ((0,0),(1,0),(2,0)), \quad Y = ((0,0),(\sqrt{3},0),(\sqrt{3},\sqrt{3})).$$

Their squared distances in pair order $(12, 13, 23)$ are $(1, 4, 1)$ and $(3, 6, 3)$. All three weights increase by 2, so the probabilities coincide. The centered Gram eigenvalues are $(0, 0, 2)$ and $(0, 1, 3)$. Their minimum dimensions are therefore 1 and 2; unequal distance matrices exclude isometry. □

The independently executed check gives probabilities $(0.4878555512, 0.0242888977, 0.4878555512)$ for both configurations. This is a mathematical counterexample, not a human-response experiment. It does not contradict the complete-distance isometry theorem, whose premise is absent here. It also shows why merely naming an ambient dimension is not universally sufficient to remove the ambiguity.

Restrictions can matter decisively. A predeclared dimension $d < n-1$ excludes the positive-offset family above whenever it would require rank $n-1$. A justified norm constraint or independent distance anchor can also select one offset. Such restrictions intersect the affine set identified by the probabilities; they do not become consequences of those probabilities. Negative offsets are permissible only when nonnegativity and positive semidefiniteness survive. The feasible offsets should be reported, not silently selected by a plotting routine.

If temperature is also unknown, a positive rescaling $w' = aw$, $\tau' = a\tau$ preserves predictions in addition to the offset freedom. The reporting contract therefore states which scale, offset, dimension, and context assumptions were supplied and which features were learned from independent observations.

## Finite-response uncertainty and a stronger measurement contract

The exact theorem supplies an identification target, not finite-sample certainty. Consider $K \geq 1$ registered triples, each repeated $m \geq 1$ times under a stationary response distribution. Within each triple, its $m$ responses are independent and identically distributed; independence between different triples is unnecessary for the union bound. Repeated responses from one adapting person need not satisfy this premise. Let $p_{T,e}$ be a pair-choice probability and $\hat{p}_{T,e}$ its empirical proportion. Suppose every $p_{T,e} \geq b > 0$, with $b \leq 1/3$ declared as a model restriction, and fix $\delta \in (0, 1)$. Put

$$\eta = \sqrt{\frac{\log(6K/\delta)}{2m}}, \qquad \eta \leq b/2.$$

Applying Hoeffding's bounded-Bernoulli tail inequality [16] to each of $3K$ proportions gives, with probability at least $1 - \delta$, simultaneous errors at most $\eta$. On this event all estimated probabilities are positive. Since the derivative of $\log x$ is $1/x$, the mean-value theorem yields

$$|\log \hat{p}_{T,e} - \log p_{T,e}| \leq \frac{2\eta}{b}.$$

It follows that an estimated weight difference obtained from a log-odds ratio has error at most $4\tau\eta/b$. Along a fixed path of $\ell$ comparison edges, errors add by the triangle inequality. With an exact anchor, the recovered endpoint weight obeys

$$|\widehat{w}_e - w_e| \leq \frac{4\tau\ell\eta}{b}.$$

An anchor with deterministic error bound $a \geq 0$ adds $a$ to the right-hand side. These bounds are deliberately conservative and do not assume independence between errors along the path. They require a correct common response model; participant heterogeneity, adaptation, and fitted temperature require their own analysis. If the probability floor is absent or $\eta > b/2$, this logarithmic bound is not asserted.

In exact model-compatible data, sums of oriented log-odds around any graph cycle are zero. Nonzero empirical sums can reflect sampling variation or response-model failure. A preregistered diagnostic can compare those discrepancies with the simultaneous uncertainty bounds before imposing a Euclidean fit. Projection onto a positive semidefinite matrix is an estimator choice, not proof that the original observations arose from a Euclidean geometry.

The revised experimental contract has three separate outputs: held-out choice prediction, the geometries consistent with the observations and declared restrictions, and uncertainty under the actual participant/stimulus sampling design. A regularizer may prefer one member of an identification set, but the report must not describe that preference as empirical identification. An inferred dimension is defensible only under the stated design and restrictions. The Continuum's ambition is a map whose explanatory reach can be challenged, including by an equally predictive map it cannot yet exclude.

# When a justified dimension cap makes the geometry unique

**A Guide's explanation**

This is the newest result, and the one the prologue promised. Choices alone leave part of the map unfixed: equally good geometries can disagree about distances. The theorem shows that one justified assumption—a cap on the number of dimensions—removes that freedom and makes the geometry unique whenever an admissible one exists, and an explicit spectral test detects whether it does.

The offset ambiguity has a positive resolution under an explicit geometric restriction. Let $n \geq 3$. Suppose the comparison graph on all $\binom{n}{2}$ pair-vertices is connected, the temperature $\tau > 0$ is known, and the exact probabilities follow the registered response model. Let an admissible squared-distance matrix be hollow, symmetric and Euclidean, with centered Gram rank at most $r$, where $0 \leq r < n-1$. The rank bound is supplied independently as a model restriction. It must be justified through the experimental design or tested on independent evidence.

**Theorem 1** (Rank-constrained identification from triplet probabilities)**.** *Under these conditions, at most one admissible squared-distance matrix produces the observed probabilities. If it exists, its Euclidean geometry is identified up to isometry. With unknown temperature, the corresponding statement holds up to a common positive scale and isometry.*

*Proof.* For known temperature, the comparison-graph proposition implies that two candidate squared-distance matrices differ by one common off-diagonal constant $c$. Their centered Gram matrices satisfy $B' = B + (c/2)J$. Restrict to the centered subspace $\mathbf{1}^{\perp}$, whose dimension is $n-1$. There $J$ is the identity. If $c > 0$, positive semidefiniteness of $B$ makes $B'$ positive definite on that subspace, giving rank $n-1 > r$. If $c < 0$, the rank bound on $B$ supplies a nonzero centered vector $v$ in its nullspace, and $v^{\mathsf{T}}B'v = (c/2)\|v\|^2 < 0$, contradicting positive semidefiniteness. Hence $c = 0$, and the complete-distance isometry proposition applies.

If temperatures $\tau, \tau' > 0$ differ, equality of log odds gives $w' = aw + c$ on every pair, where $a = \tau'/\tau > 0$. Divide $w'$ by $a$ to return to the preceding argument. Positive scaling preserves positive semidefiniteness and rank, so the offset vanishes after rescaling. The remaining coordinate scale is $\sqrt{a}$. □

The strict rank inequality is essential. In the earlier three-point example the allowed dimension was $r = 2 = n - 1$, so the theorem's premise was absent. Permitting full centered dimension allows every positive common offset and can change intrinsic rank. Requiring $r = 1$ for those three labeled stimuli selects the collinear geometry if the exact probabilities admit it. This distinction supplies a constructive answer to the earlier counterexample while preserving its force.

## An exact feasibility certificate

After checking exact log-odds cycle consistency, recover any real pair-weight solution $w^0$ of the connected log-odds equations and form its symmetric hollow matrix $\Delta^0$. This preliminary solution need not be nonnegative or Euclidean. Put $B^0 = -J\Delta^0 J/2$ and let

$$\alpha_1 \leq \cdots \leq \alpha_{n-1}$$

be its eigenvalues restricted to $\mathbf{1}^{\perp}$. A positive-semidefinite candidate of rank at most $r < n-1$ exists exactly when

$$\alpha_1 = \alpha_2 = \cdots = \alpha_{n-1-r}.$$

When this condition holds, the unique offset is $c_* = -2\alpha_1$. Indeed, shifting by $c_*/2$ makes the smallest eigenvalue zero; the stated equality supplies at least $n-1-r$ zero eigenvalues and all others are nonnegative. A smaller shift violates positive semidefiniteness and a larger shift gives full centered rank. Conversely any valid rank-deficient shift must place the smallest eigenvalue at zero and have the required multiplicity. A hollow matrix whose centered Gram matrix is positive semidefinite is the squared-distance matrix reconstructed from that Gram matrix, by the algebra in the Euclidean criterion; hence its entries are automatically nonnegative.

For exact data this certificate jointly checks the common response law, the connected comparison design and the rank restriction. Cycle inconsistencies reject the response law before the spectral test. A failed multiplicity condition rejects the declared low-rank Euclidean class. Passing both identifies a geometry within that class rather than merely choosing a visually convenient member of an offset family.

### Finite-data interpretation

Exact eigenvalue multiplicity is not robust to arbitrary measurement noise. In finite response data, the simultaneous log-odds bounds define an uncertainty set of weight differences. Intersect that set with the declared Euclidean and rank constraints and report the resulting range of distances or predictions. A small empirical eigenvalue is evidence to evaluate under that uncertainty model; rounding it to zero is an estimator decision. The theorem therefore strengthens identification without turning a finite noisy estimate into exact knowledge.

The accompanying checks generate Euclidean configurations with $r < n-1$, reconstruct their probabilities, and verify both the spectral offset and rejection of nonzero shifted candidates. They include full-rank counterexamples and disconnected comparison designs. The result supplies a clear experimental choice: obtain independent anchors, justify a lower-dimensional model, or report the full compatible family. Each route identifies the additional information on which the interpretation depends.

## References and source scope

[**1**] Architect X13. *Eidara Continuum Bible*, supplied canonical edition. Section 2.8, PDF p. 12; glossary p. 113.

https://eidara.ai/science-philosophy/

Scope: Internal source supplies motivation and terminology, not a validated phenomenal metric. URL identifies the public research context, not the private Bible.

[**2**] Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika* 17, 401–419.

https://www.ets.org/research/policy_research_reports/publications/article/1952/ibra.html

Scope: Original scaling method. Institutional record and opening two pages of the original paper checked.

[**3**] Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1–27. DOI: 10.1007/BF02289565.

https://doi.org/10.1007/BF02289565

Scope: Primary publisher abstract and bibliographic record; monotonic-fit formulation.

[**4**] Tversky, A. (1977). Features of similarity. *Psychological Review* 84, 327–352. DOI: 10.1037/0033-295X.84.4.327.

https://worthylab.org/wp-content/uploads/2021/08/tversky1977_featuresofsimilarity_psychreview.pdf

Scope: Original paper, opening pages; feature-based challenge to automatic metric assumptions.

[**5**] Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science* 237, 1317–1323. DOI: 10.1126/science.3629243.

https://cseweb.ucsd.edu/~gary/PAPER-SUGGESTIONS/shepard-science-87.pdf

Scope: Original first article page inspected visually; conditional connection between distance and generalization.

[**6**] Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience* 2, 4.

https://doi.org/10.3389/neuro.06.004.2008

Scope: Original abstract, introduction, RDM construction, and comparison sections.

[**7**] Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLOS Computational Biology* 10, e1003553.

https://doi.org/10.1371/journal.pcbi.1003553

Scope: Primary methods and noise-ceiling discussion; conditional representational inference.

## References and source scope, continued

[**8**] Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., and Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage* 137, 188–200.

https://doi.org/10.1016/j.neuroimage.2015.12.012

Scope: Author-hosted proof opening and final bibliographic record; noise bias and crossvalidation.

[**9**] Diedrichsen, J., and Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLOS Computational Biology* 13, e1005508.

https://doi.org/10.1371/journal.pcbi.1005508

Scope: Primary abstract; relationships and assumptions of complementary representational methods.

[**10**] Schütt, H. H., Kipnis, A. D., Diedrichsen, J., and Kriegeskorte, N. (2023). Statistical inference on representational geometries. *eLife* 12, e82566.

https://doi.org/10.7554/eLife.82566

Scope: Author manuscript arXiv:2112.09200v2, abstract and two-factor inference sections; finite-sample corrections matter.

[**11**] Peterson, J. C., Abbott, J. T., and Griffiths, T. L. (2018). Evaluating (and improving) the correspondence between deep neural networks and human representations. *Cognitive Science* 42, 2648–2669.

https://doi.org/10.1111/cogs.12670

Scope: Primary author abstract and publication record; representation reweighting is a serious baseline.

[**12**] Hebart, M. N., Zheng, C. Y., Pereira, F., and Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature Human Behaviour* 4, 1173–1185.

https://doi.org/10.1038/s41562-020-00951-3

Scope: Author manuscript abstract and object-judgment methods/results; behavioral embedding, not phenomenal identity.

[**13**] Hebart, M. N., Contier, O., Teichmann, L., Rockter, A. H., Zheng, C. Y., Kidder, A., Corriveau, A., Vaziri-Pashkam, M., and Baker, C. I. (2023). THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *eLife* 12, e82580.

https://doi.org/10.7554/eLife.82580

Scope: Primary abstract and dataset overview; a proposed benchmark resource, not data analyzed here.

[**14**] Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2, 200–219.

https://consc.net/papers/facing.html

Scope: Original author text, introduction and opening distinction; philosophical argument, not an empirical consciousness test.

[**15**] Kleindessner, M., and von Luxburg, U. (2014). Uniqueness of ordinal embedding. *COLT*, PMLR 35, 40–67.

https://proceedings.mlr.press/v35/kleindessner14.html

Scope: Peer-reviewed primary conference paper; full original opening and definitions inspected. Dense-sample geometric uniqueness has additional assumptions; it does not identify a finite QSM-R2 distance matrix from choices alone.

[**16**] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* 58, 13–30.

https://doi.org/10.1080/01621459.1963.10500830

Scope: Peer-reviewed primary concentration result. Applied to fixed triples with independent repeated responses; the union bound does not require independence between triples.

[**17**] Balduzzi, D., and Tononi, G. (2009). Qualia: the geometry of integrated information. *PLOS Computational Biology* 5(8), e1000462.

https://doi.org/10.1371/journal.pcbi.1000462

Scope: Primary prior use of qualia space in integrated information theory. Its system-informational construction differs from the behavioral report geometry defined here.