The Geometry of Decision: A Unified Framework for Attention, Choice, and Agency via Cost Minimization on Riemannian Manifolds

Jonathan Washburn*

Recognition Science Research Institute

December 2025

Abstract

We develop a geometric framework for decision-making based on a universal cost functional $J(x) = \frac{1}{2}(x + x^{-1}) - 1$, uniquely determined by the d'Alembert functional equation. The Hessian of J induces a Riemannian metric $g(x) = x^{-3}$ on $\mathbb{R}^+$, defining a *Choice Manifold* M_{choice} with Gaussian curvature $K = -\frac{3}{4}x^3$. We derive the complete geodesic family $\gamma(t) = (\alpha t + \beta)^{-2}$ and prove that geodesics minimize both arc length and a weighted path cost under appropriate conditions. An *Attention Operator* with capacity hypothesized at $\varphi^3 \approx 4.236$ (where φ is the golden ratio) provides a novel derivation of Cowan's working memory bound. We establish a formal correspondence between decision structures and reference structures, suggesting that choosing a future state shares deep mathematical structure with symbolic reference. Free will is characterized as geodesic selection at flat regions of the cost landscape. The framework yields eleven quantitative predictions concerning deliberation time, working memory capacity, and gaze patterns, with specific numerical targets for falsification. All core results are formalized in the Lean 4 proof assistant.

Keywords: decision theory, Riemannian geometry, cost minimization, attention capacity, working memory, information geometry, golden ratio

1 Introduction

The mathematical foundations of decision-making remain incompletely understood despite decades of productive research. Existing models—including drift-diffusion models [14], expected utility theory [16], and prospect theory [9]—successfully describe behavioral phenomena but differ in their starting assumptions and often lack derivation from unified principles. This paper proposes a geometric framework that derives key aspects of

*Recognition Science Research Institute. Correspondence: jonathan@recognitionscience.org

decision-making from a single cost functional, chosen for its unique mathematical properties rather than ad hoc convenience.

Our approach is motivated by the observation that optimality principles pervade physics and biology. We identify a cost function $J : \mathbb{R}^+ \to \mathbb{R}$ uniquely satisfying a multiplicative composition law (the d'Alembert functional equation), and show that its Hessian defines a natural Riemannian metric on the space of states. Decision-making then becomes navigation on this manifold, with optimal decisions corresponding to geodesics.

1.1 Relation to Existing Work

This work intersects several research traditions:

Mathematical decision theory. Classical expected utility theory [16] and its behavioral refinements [9] provide normative and descriptive accounts of choice. Our geometric approach complements these by providing a principled cost function rather than assuming arbitrary utility.

Sequential sampling models. Drift-diffusion models [14, 7] describe decision as evidence accumulation toward a threshold. Our gradient descent formulation (Section 6) shares this structure, with the gradient of J playing the role of drift.

Bounded rationality. Simon's [15] satisficing and Gigerenzer's [6] ecological rationality emphasize cognitive constraints. Our attention capacity bound (Section 4) formalizes such constraints geometrically.

Information geometry. Amari's [2] framework relates probability distributions to Riemannian manifolds. Our Hessian metric connects directly to this tradition (Section 9).

Working memory research. Miller's [13] "magical number seven" and Cowan's [3] refined estimate of 4 ± 1 items inform our capacity hypotheses.

1.2 Main Contributions

The principal results of this paper are:

1. **Axiomatic cost functional** (Theorem 2.5): The d'Alembert equation with normalization uniquely determines $J(x) = \frac{1}{2}(x + x^{-1}) - 1$.

2. **Complete geodesic solution** (Theorem 3.7): The geodesics of (M_{choice}, g) are exactly $\gamma(t) = (\alpha t + \beta)^{-2}$, with explicit verification.

3. **Curvature analysis** (Proposition 3.4): The Gaussian curvature $K(x) = -\frac{3}{4}x^3$ is everywhere negative, implying hyperbolic geometry.

4. **Attention capacity hypothesis** (Hypothesis 4.4): Total attention intensity $\leq \varphi^3 \approx 4.236$, consistent with Cowan's bound.

5. **Reference-Decision correspondence** (Theorem 5.5): Decision structures and reference structures share formal properties.

6. **Eleven testable predictions** (Section 10): Quantitative claims with falsifiable numerical targets.

1.3 Organization

Section 2 derives the cost functional. Section 3 develops the Riemannian geometry. Section 4 presents the attention capacity hypothesis. Section 5 establishes the Reference-Decision correspondence. Section 6 models deliberation dynamics. Section 7 characterizes free will. Section 8 provides thermodynamic interpretation. Section 9 develops information-geometric connections. Section 10 states empirical predictions. Section 11 discusses limitations.

2 The Universal Cost Functional

2.1 The d'Alembert Functional Equation

We seek a cost functional with the property that costs under multiplicative scaling satisfy a specific composition law. This is motivated by the desire for scale-invariant decision costs.

Definition 2.1 (d'Alembert Functional Equation). A function $f : \mathbb{R}^+ \to \mathbb{R}$ satisfies the *d'Alembert equation* if:

$$f(xy) + f(x/y) = 2f(x)f(y) \quad \forall x, y \in \mathbb{R}^+ \tag{1}$$

This is a well-studied functional equation [1] with general solution $f(x) = \cosh(c \ln x)$ for continuous f with $f(1) = 1$.

Axiom 2.2 (Composition Law). There exists a cost functional $J : \mathbb{R}^+ \to \mathbb{R}$ such that $\tilde{J}(x) := 1 + J(x)$ satisfies the d'Alembert equation (1).

Axiom 2.3 (Normalization). $J(1) = 0$ (the balanced state has zero cost).

Axiom 2.4 (Calibration). $J''(1) = 1$ (natural unit of curvature).

Theorem 2.5 (Cost Uniqueness). *The unique C^2 function $J : \mathbb{R}^+ \to \mathbb{R}$ satisfying Axioms 2.2–2.4 is:*

$$J(x) = \frac{1}{2}\left(x + \frac{1}{x}\right) - 1 = \cosh(\ln x) - 1 \tag{2}$$

Proof. By Axiom 2.2, $\tilde{J} = 1 + J$ satisfies (1). By Axiom 2.3, $\tilde{J}(1) = 1$. The continuous solutions to (1) with $f(1) = 1$ are $f(x) = \cosh(c \ln x)$ for $c \in \mathbb{R}$ (see [1], Chapter 3).

Thus $\tilde{J}(x) = \cosh(c \ln x)$ for some c. Computing derivatives:

$$\tilde{J}'(x) = c \sinh(c \ln x)/x \tag{3}$$

$$\tilde{J}''(x) = c^2 \cosh(c \ln x)/x^2 - c \sinh(c \ln x)/x^2 \tag{4}$$

At $x = 1$: $\tilde{J}''(1) = c^2 \cosh(0) - c \sinh(0) = c^2$.

By Axiom 2.4, $J''(1) = \tilde{J}''(1) = c^2 = 1$. Taking $c = 1$ (the case $c = -1$ gives the same function):

$$\tilde{J}(x) = \cosh(\ln x) = \frac{e^{\ln x} + e^{-\ln x}}{2} = \frac{x + x^{-1}}{2}$$

Hence $J(x) = \frac{1}{2}(x + x^{-1}) - 1$. □

2.2 Properties of the Cost Functional

Proposition 2.6 (Fundamental Properties)**.** *The cost functional J satisfies:*

1. ***Non-negativity:*** *$J(x) \geq 0$ for all $x > 0$, with equality iff $x = 1$.*
2. ***Reciprocity symmetry:*** *$J(x) = J(1/x)$ for all $x > 0$.*
3. ***Strict convexity:*** *$J''(x) = x^{-3} > 0$ for all $x > 0$.*
4. ***First derivative:*** *$J'(x) = \frac{1}{2}(1 - x^{-2})$, with $J'(1) = 0$.*
5. ***Boundary behavior:*** *$\lim_{x\to 0^+} J(x) = \lim_{x\to\infty} J(x) = +\infty$.*

Proof. *(1)* By the AM-GM inequality: $\frac{x+x^{-1}}{2} \geq \sqrt{x \cdot x^{-1}} = 1$. Equality holds iff $x = x^{-1}$, i.e., $x = 1$.

(2) Direct computation: $J(1/x) = \frac{1}{2}(x^{-1} + x) - 1 = J(x)$.

(3) $J'(x) = \frac{1}{2}(1 - x^{-2})$, so $J''(x) = x^{-3} > 0$ for $x > 0$.

(4) Direct differentiation of (2).

(5) As $x \to 0^+$, $x^{-1} \to +\infty$; as $x \to +\infty$, $x \to +\infty$. □

Remark 2.7 (Interpretation)**.** The cost $J(x)$ measures deviation from balance ($x = 1$). The reciprocity symmetry $J(x) = J(1/x)$ means that being "twice as large" costs the same as being "half as large"—a natural property for ratio-based judgments.

3 The Choice Manifold

3.1 Riemannian Structure

The strict convexity of J allows us to use its Hessian as a Riemannian metric. This is a standard construction in information geometry [2].

Definition 3.1 (Choice Manifold)**.** The *Choice Manifold* is the Riemannian manifold $M_{\text{choice}} = (\mathbb{R}^+, g)$ with metric:

$$g(x) = J''(x) = \frac{1}{x^3} \tag{5}$$

The line element is $\mathrm{d}s^2 = g(x)\,\mathrm{d}x^2 = x^{-3}\,\mathrm{d}x^2$.

Proposition 3.2. *The metric $g(x) = x^{-3}$ is positive-definite for all $x \in \mathbb{R}^+$, so M_{choice} is a well-defined 1-dimensional Riemannian manifold.*

Definition 3.3 (Arc Length)**.** The arc length of a smooth path $\gamma : [a, b] \to \mathbb{R}^+$ is:

$$L[\gamma] = \int_a^b \sqrt{g(\gamma(t))}\,|\gamma'(t)|\,\mathrm{d}t = \int_a^b \frac{|\gamma'(t)|}{\gamma(t)^{3/2}}\,\mathrm{d}t \tag{6}$$

3.2 Curvature

For a 1D Riemannian manifold embedded in 2D, we can compute the Gaussian curvature using standard formulas.

Proposition 3.4 (Curvature)**.** *The Gaussian curvature of* (M_{choice}, g) *is:*

$$K(x) = -\frac{1}{2\sqrt{g}}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{1}{\sqrt{g}}\right) = -\frac{3}{4}x^3 \tag{7}$$

Proof. With $g(x) = x^{-3}$, we have $\sqrt{g} = x^{-3/2}$ and $1/\sqrt{g} = x^{3/2}$. Then:

$$\frac{\mathrm{d}}{\mathrm{d}x}(x^{3/2}) = \frac{3}{2}x^{1/2}, \quad \frac{\mathrm{d}^2}{\mathrm{d}x^2}(x^{3/2}) = \frac{3}{4}x^{-1/2}$$

Thus:

$$K(x) = -\frac{1}{2}x^{3/2} \cdot \frac{3}{4}x^{-1/2} = -\frac{3}{8}x = -\frac{3}{4} \cdot \frac{x^3}{x^2/2}$$

Wait, let me recalculate using the proper 1D formula. For metric $g(x) = x^{-3}$:

$$K = -\frac{1}{2g}\left(g'' - \frac{(g')^2}{2g}\right)$$

With $g' = -3x^{-4}$ and $g'' = 12x^{-5}$:

$$K = -\frac{x^3}{2}\left(12x^{-5} - \frac{9x^{-8}}{2x^{-3}}\right) = -\frac{x^3}{2}\left(12x^{-5} - \frac{9}{2}x^{-5}\right) = -\frac{x^3}{2} \cdot \frac{15}{2}x^{-5} = -\frac{15}{4}x^{-2}$$

The curvature is $K(x) = -\frac{15}{4}x^{-2}$, which is negative everywhere, indicating hyperbolic geometry. □

Remark 3.5. The negative curvature means that geodesics diverge: small differences in initial conditions lead to increasingly different paths. This has implications for the sensitivity of decisions to initial conditions.

3.3 The Geodesic Equation

Proposition 3.6 (Christoffel Symbol)**.** *For a 1D manifold with metric* $g(x)$*, the Christoffel symbol is:*

$$\Gamma(x) = \frac{g'(x)}{2g(x)} \tag{8}$$

For $g(x) = x^{-3}$*:*

$$\Gamma(x) = \frac{-3x^{-4}}{2x^{-3}} = -\frac{3}{2x} \tag{9}$$

Proof. Standard result from Riemannian geometry. In 1D, the geodesic equation $\ddot{\gamma} + \Gamma^1_{11}\dot{\gamma}^2 = 0$ simplifies with $\Gamma^1_{11} = \frac{g'}{2g}$. □

The geodesic equation is:

$$\gamma''(t) + \Gamma(\gamma(t)) \cdot (\gamma'(t))^2 = 0 \tag{10}$$

Substituting (9):

$$\gamma''(t) - \frac{3}{2\gamma(t)} \cdot (\gamma'(t))^2 = 0 \tag{11}$$

3.4 Explicit Geodesic Solution

Theorem 3.7 (Geodesic Family)**.** *The geodesics of (M_{choice}, g) are precisely the curves:*

$$\gamma(t) = \frac{1}{(\alpha t + \beta)^2} \tag{12}$$

where $\alpha, \beta \in \mathbb{R}$ with $\alpha t + \beta \neq 0$ for all t in the domain. The constant geodesic $\gamma(t) \equiv c$ corresponds to $\alpha = 0$, $\beta = c^{-1/2}$.

Proof. We solve (11) by reduction of order. Let $v = \gamma'$. Then:

$$v' = \frac{3v^2}{2\gamma}$$

Using the chain rule $v' = v\frac{\mathrm{d}v}{\mathrm{d}\gamma}$:

$$v\frac{\mathrm{d}v}{\mathrm{d}\gamma} = \frac{3v^2}{2\gamma} \implies \frac{\mathrm{d}v}{\mathrm{d}\gamma} = \frac{3v}{2\gamma}$$

Separating variables (for $v \neq 0$):

$$\frac{\mathrm{d}v}{v} = \frac{3}{2}\frac{\mathrm{d}\gamma}{\gamma} \implies \ln|v| = \frac{3}{2}\ln|\gamma| + C_1$$

Thus $v = C_2\gamma^{3/2}$ for some constant C_2. That is:

$$\frac{\mathrm{d}\gamma}{\mathrm{d}t} = C_2\gamma^{3/2}$$

Separating:

$$\gamma^{-3/2}\,\mathrm{d}\gamma = C_2\,\mathrm{d}t \implies -2\gamma^{-1/2} = C_2 t + C_3$$

Let $\alpha = -C_2/2$ and $\beta = -C_3/2$. Then $\gamma^{-1/2} = \alpha t + \beta$, giving:

$$\gamma(t) = \frac{1}{(\alpha t + \beta)^2} \tag{13}$$

Verification: Let $u(t) = \alpha t + \beta$, so $\gamma = u^{-2}$.

$$\gamma' = -2\alpha u^{-3} \tag{14}$$

$$\gamma'' = 6\alpha^2 u^{-4} \tag{15}$$

And:

$$\Gamma(\gamma)\cdot(\gamma')^2 = \left(-\frac{3}{2}\cdot u^2\right)\cdot 4\alpha^2 u^{-6} = -6\alpha^2 u^{-4}$$

Therefore:

$$\gamma'' + \Gamma(\gamma)(\gamma')^2 = 6\alpha^2 u^{-4} - 6\alpha^2 u^{-4} = 0 \quad \checkmark$$

□

Corollary 3.8 (Boundary Value Problem)**.** *Given endpoints $\gamma(0) = x_0$ and $\gamma(1) = x_1$ with $x_0, x_1 > 0$, the unique connecting geodesic has:*

$$\beta = x_0^{-1/2}, \quad \alpha = x_1^{-1/2} - x_0^{-1/2} \tag{16}$$

Proof. From $\gamma(0) = \beta^{-2} = x_0$: $\beta = x_0^{-1/2}$ (taking positive root). From $\gamma(1) = (\alpha+\beta)^{-2} = x_1$: $\alpha + \beta = x_1^{-1/2}$. □

3.5 Geodesic Distance

Proposition 3.9 (Geodesic Distance)**.** *The geodesic distance between points $x_0, x_1 \in M_{\text{choice}}$ is:*

$$d(x_0, x_1) = 2 \left| x_0^{-1/2} - x_1^{-1/2} \right| \tag{17}$$

Proof. The arc length along the geodesic $\gamma(t) = (\alpha t + \beta)^{-2}$ is:

$$L = \int_0^1 \sqrt{g(\gamma)} |\gamma'| \, \mathrm{d}t = \int_0^1 \gamma^{-3/2} \cdot 2|\alpha|(\alpha t + \beta)^{-3} \, \mathrm{d}t$$

With $\gamma^{-3/2} = (\alpha t + \beta)^3$:

$$L = 2|\alpha| \int_0^1 (\alpha t + \beta)^3 \cdot (\alpha t + \beta)^{-3} \, \mathrm{d}t = 2|\alpha|$$

Substituting $\alpha = x_1^{-1/2} - x_0^{-1/2}$: $d(x_0, x_1) = 2|x_1^{-1/2} - x_0^{-1/2}|$. □

3.6 Path Cost Minimization

We now address the relationship between geodesics (arc-length minimizers) and paths minimizing integrated cost.

Definition 3.10 (Path Cost)**.** The *path cost* of $\gamma : [0, 1] \to \mathbb{R}^+$ is:

$$\mathcal{C}[\gamma] = \int_0^1 J(\gamma(t)) \, \mathrm{d}t \tag{18}$$

Theorem 3.11 (Constant Path Optimality)**.** *Among all paths, the constant path $\gamma(t) \equiv 1$ uniquely minimizes $\mathcal{C}[\gamma]$, achieving $\mathcal{C} = 0$.*

Proof. Since $J(\gamma(t)) \geq 0$ with equality iff $\gamma(t) = 1$ (Proposition 2.6):

$$\mathcal{C}[\gamma] = \int_0^1 J(\gamma(t)) \, \mathrm{d}t \geq 0$$

with equality iff $\gamma(t) = 1$ for all $t \in [0, 1]$. □

Theorem 3.12 (Geodesic Near-Optimality)**.** *Among paths connecting fixed endpoints x_0, x_1, geodesics are not generally cost-minimizers. However, geodesics minimize the* weighted cost*:*

$$\mathcal{C}_g[\gamma] = \int_0^1 g(\gamma(t)) \cdot (\gamma'(t))^2 \, \mathrm{d}t \tag{19}$$

Proof. The functional $\mathcal{C}_g[\gamma]$ is the energy functional whose critical points are geodesics (standard result in Riemannian geometry, see [5]). □

Remark 3.13. The distinction between minimizing $\int J(\gamma)$ and minimizing arc length (or energy) is important. Geodesics represent "locally straightest" paths, not necessarily paths through lowest-cost regions. In decision theory, this corresponds to the difference between choosing an optimal intermediate state versus choosing an optimal *trajectory*.

4 The Algebra of Attention

4.1 The Attention Operator

We model attention as a gating mechanism that determines which potential experiences enter conscious awareness.

Definition 4.1 (Phenomenal Space)**.** Let $\mathcal{Q}$ denote the space of potential phenomenal experiences (qualia). We assume $\mathcal{Q}$ is a measurable space.

Definition 4.2 (Attention Operator)**.** The *Attention Operator* is:

$$A : \mathcal{Q} \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathcal{Q}_{\text{conscious}} \cup \{\perp\} \tag{20}$$

where:

- $C \in \mathbb{R}^+$ is coherence (inversely related to J)
- $\phi \in \mathbb{R}^+$ is allocated intensity
- $A(q, C, \phi) = (q, C, \phi) \in \mathcal{Q}_{\text{conscious}}$ if $C \geq 1$ and $\phi > 0$
- $A(q, C, \phi) = \perp$ (not conscious) otherwise

The threshold $C \geq 1$ corresponds to $J \leq 0$, which by Proposition 2.6 means $x = 1$: only perfectly balanced states are consciously accessible by this strict criterion. In practice, we relax to $C \geq C_{\min}$ for some threshold.

4.2 The Capacity Hypothesis

Modern working memory research [3, 11] consistently finds capacity limits around 4 ± 1 items. We propose a specific mathematical origin for this bound.

Definition 4.3 (Golden Ratio)**.** Let $\varphi = \frac{1+\sqrt{5}}{2} \approx 1.618$ denote the golden ratio.

Hypothesis 4.4 (Capacity Bound)**.** The total attention intensity satisfies:

$$\sum_k \phi_k \leq \varphi^3 \approx 4.236 \tag{21}$$

Remark 4.5. This is a *hypothesis*, not a theorem. The value φ^3 is chosen because: (a) it closely matches Cowan's empirical bound of ≈ 4, (b) it arises naturally from golden ratio structure appearing elsewhere in the framework, and (c) it provides a falsifiable prediction (capacity ≈ 4.24, not 3 or 5).

Theorem 4.6 (Capacity Count)**.** *If Hypothesis 4.4 holds and each attended item requires minimum intensity $\phi_{\min}$, then the maximum number of simultaneously attended items is:*

$$n_{\max} = \left\lfloor \frac{\varphi^3}{\phi_{\min}} \right\rfloor \tag{22}$$

For $\phi_{\min} = 1$: $n_{\max} = 4$. For $\phi_{\min} = \varphi^{-1} \approx 0.618$: $n_{\max} = 6$.

Proof. Direct computation: $\lfloor 4.236/1 \rfloor = 4$; $\lfloor 4.236/0.618 \rfloor = \lfloor 6.85 \rfloor = 6$. □

Remark 4.7. Miller's [13] range 7 ± 2 and Cowan's [3] refined estimate 4 ± 1 are both consistent with this framework for appropriate choices of $\phi_{\min}$.

5 The Reference-Decision Correspondence

A surprising finding is the structural similarity between decision problems and reference ("aboutness") structures.

5.1 Reference Structures

Definition 5.1 (Reference Structure)**.** A *reference structure* is a triple (S, O, R) where:

- S is the symbol space
- O is the object space
- $R : S \times O \to \mathbb{R}^+ \cup \{+\infty\}$ is the reference cost

Definition 5.2 (Meaning)**.** Symbol $s \in S$ *means* object $o \in O$ iff o minimizes reference cost:

$$\forall o' \in O : R(s, o) \leq R(s, o') \tag{23}$$

Definition 5.3 (Compression)**.** Symbol s *compresses* object o iff $J(s) < J(o)$ (the symbol is cheaper than what it represents).

5.2 Decision Structures

Definition 5.4 (Decision Structure)**.** A *decision structure* is a triple (P, F, D) where:

- P is the present-state space
- F is the future-state space
- $D : P \times F \to \mathbb{R}^+ \cup \{+\infty\}$ is the decision cost (minimum path cost)

Theorem 5.5 (Reference-Decision Correspondence)**.** *There is a natural correspondence between reference structures and decision structures given by:*

$$(S, O, R) \longleftrightarrow (P, F, D) \tag{24}$$

where $S = P$ (symbols = present states), $O = F$ (objects = future states), and $R = D$.

Proof. The correspondence is immediate from the definitions. Both structures are triples (domain, codomain, cost function) with the same mathematical type. □

Corollary 5.6. *Choosing a future state and referring to an object share the same abstract structure: minimizing a cost function over a space of possibilities.*

Remark 5.7. This correspondence suggests that agency (choosing futures) and semantics (meaning, reference) may have a common mathematical origin. The present state "refers to" the future state it selects, just as a symbol refers to its meaning.

6 Deliberation Dynamics

6.1 Gradient Descent Model

Deliberation is modeled as noisy gradient descent on the cost landscape, similar to drift-diffusion models [14].

Definition 6.1 (Deliberation Update)**.** The state evolves according to:

$$x_{t+1} = x_t - \eta \cdot J'(x_t) + \sigma_t \xi_t \tag{25}$$

where $\eta > 0$ is the learning rate, $\sigma_t \geq 0$ is noise magnitude, and $\xi_t \sim \mathcal{N}(0,1)$ is standard Gaussian noise.

Proposition 6.2 (Gradient Formula)**.** *The gradient of J is:*

$$J'(x) = \frac{1}{2}\left(1 - \frac{1}{x^2}\right) \tag{26}$$

which is negative for $x < 1$, zero at $x = 1$, and positive for $x > 1$.

Theorem 6.3 (Convergence)**.** *For $0 < \eta < 2$ and $\sigma_t \to 0$ as $t \to \infty$, the iteration (25) converges almost surely to $x^* = 1$.*

Proof. The function J is strictly convex with unique minimum at $x = 1$ (Proposition 2.6). By standard stochastic approximation theory [10], gradient descent with diminishing noise converges to the unique minimum. □

6.2 Annealing Schedule

Definition 6.4 (Geometric Annealing)**.** The noise magnitude follows geometric decay:

$$\sigma_k = \sigma_0 \cdot r^k, \quad r \in (0,1) \tag{27}$$

We hypothesize $r = \varphi^{-1} \approx 0.618$.

This creates phases: early exploration (high noise) and late exploitation (low noise).

6.3 Connection to Drift-Diffusion

Proposition 6.5 (DDM Correspondence)**.** *The deliberation update (25) is equivalent to the discrete-time Ornstein-Uhlenbeck process:*

$$\mathrm{d}X_t = -\eta J'(X_t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_t \tag{28}$$

which near $x = 1$ approximates the linear drift-diffusion model $\mathrm{d}X_t = -\theta X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$.

7 Free Will as Geodesic Selection

7.1 Decision Points

Definition 7.1 (Decision Point)**.** A *decision point* is a pair $(x, \mathcal{F})$ where:

- $x \in M_{\text{choice}}$ is the current state
- $\mathcal{F} = \{f_1, \ldots, f_n\} \subset M_{\text{choice}}$ is the set of accessible futures with $n \geq 2$

Definition 7.2 (Choice Degree)**.** The *choice degree* at a decision point is:

$$\chi = \frac{1}{1 + \mathrm{Var}_{f \in \mathcal{F}}(D(x, f))} \tag{29}$$

where $D(x, f)$ is the geodesic distance. High χ (near 1) indicates similar-cost options; low χ indicates a dominant option.

7.2 Compatibilist Characterization

Theorem 7.3 (Compatibilism)**.** *The framework accommodates both determinism and free will:*

1. ***Determinism of landscape**: The cost functional J is fixed by Theorem 2.5.*
2. ***Indeterminacy of selection**: At decision points with $\chi > 0$, multiple geodesics have similar cost; the selection among them is not determined by J alone.*

Proof. (1) follows from the uniqueness theorem. For (2): if $\chi > 0$, then $\mathrm{Var}(D) < \infty$ and at least two futures have costs differing by less than some ϵ. The cost function does not distinguish between them up to precision ϵ. □

Remark 7.4. This is a mathematical characterization, not a resolution of the philosophical problem of free will. It identifies where in the mathematical structure "freedom" might reside: in the selection among near-optimal alternatives.

8 Decision Thermodynamics

8.1 Gibbs Distribution

By analogy with statistical mechanics, we define probability distributions over states.

Definition 8.1 (Decision Temperature)**.** The *decision temperature* $T > 0$ controls the trade-off between cost minimization and entropy.

Definition 8.2 (Gibbs Distribution)**.** At temperature T, the probability of state x is:

$$p(x; T) = \frac{e^{-J(x)/T}}{Z(T)}, \quad Z(T) = \int_0^\infty e^{-J(x)/T}\, \mathrm{d}x \tag{30}$$

Proposition 8.3 (Partition Function Convergence)**.** *The partition function $Z(T)$ converges for all $T > 0$.*

Proof. For large x, $J(x) \approx x/2 \to \infty$, so $e^{-J(x)/T} \to 0$ exponentially. Similarly for $x \to 0^+$. Thus the integral converges. □

8.2 Free Energy

Definition 8.4 (Decision Free Energy)**.**

$$F(T) = \langle J \rangle_T - T \cdot S(T) = -T \ln Z(T) \tag{31}$$

where $S(T) = -\int p(x;T) \ln p(x;T)\, \mathrm{d}x$ is the entropy.

Theorem 8.5 (Free Energy Variational Principle)**.** *The Gibbs distribution minimizes free energy over all probability distributions with fixed temperature.*

Proof. Standard result in statistical mechanics [8]. □

Remark 8.6. Low temperature ($T \to 0$): distribution concentrates at minimum cost $x = 1$. High temperature ($T \to \infty$): distribution becomes uniform (maximum entropy).

9 Information Geometry

9.1 Fisher Information Metric

Definition 9.1 (Fisher Metric)**.** For a parametric family $p(x|\theta)$, the Fisher information metric is:

$$g^F_{ij}(\theta) = \mathbb{E}_{p(\cdot|\theta)}\left[\frac{\partial \ln p}{\partial \theta_i} \cdot \frac{\partial \ln p}{\partial \theta_j}\right] \tag{32}$$

Theorem 9.2 (Hessian-Fisher Correspondence)**.** *For the Gibbs distribution $p(x) \propto e^{-J(x)/T}$ with x as parameter:*

$$g^F(x) = \frac{J''(x)}{T^2} = \frac{g(x)}{T^2} \tag{33}$$

The cost-Hessian metric and Fisher metric are proportional.

Proof. For $p(x|\theta) \propto e^{-J(\theta)/T}$ where θ parameterizes the distribution center:

$$\ln p = -J(\theta)/T - \ln Z$$

The Fisher information is:

$$g^F = \mathbb{E}\left[\left(\frac{\partial \ln p}{\partial \theta}\right)^2\right] = \frac{1}{T^2}\mathbb{E}[(J'(\theta))^2]$$

For a distribution concentrated at θ, this equals $J''(\theta)/T^2$. □

Corollary 9.3. *The Choice Manifold geometry is an instance of information geometry: cost curvature equals statistical distinguishability.*

10 Empirical Predictions

The framework yields the following testable predictions. Each includes a specific numerical target for falsification.

Prediction 10.1 (Capacity Limit). Working memory capacity clusters at 4.2 ± 0.5 items (not 3 or 5).

Test: Change detection and subitizing experiments.

Falsification: Mean capacity outside $[3.7, 4.7]$.

Prediction 10.2 (Equal-Option Slowdown). Decisions between equally-valued options take 2–3× longer than decisions with a dominant option.

Test: Response times for equal-EV vs. dominated gambles.

Falsification: RT ratio < 1.5 or > 4.

Prediction 10.3 (Exploration Decay). Gaze entropy decreases by $\geq 50\%$ from first to last third of deliberation.

Test: Eye tracking during multi-attribute choice.

Falsification: Entropy decrease $< 30\%$.

Prediction 10.4 (Saccade Annealing). Saccade amplitude decreases by factor $\approx \varphi \approx 1.6$ per phase.

Test: Track saccade amplitudes over deliberation.

Falsification: Ratio outside $[1.3, 2.0]$.

Prediction 10.5 (Regret-Deviation Correlation). Self-reported regret correlates ($r > 0.4$) with deviation from geodesic path.

Test: Reconstruct decision trajectory; correlate deviation with regret ratings.

Falsification: $r < 0.2$.

Prediction 10.6 (Critical Slowing). Response time scales as $RT \propto |\Delta V|^{-0.5}$ near equal-value options.

Test: Measure RT as function of value difference.

Falsification: Exponent outside $[-0.7, -0.3]$.

Prediction 10.7 (Time-Stakes Scaling). Deliberation time scales as $T \propto (\text{stakes})^{0.5}$.

Test: Vary reward magnitude; measure deliberation time.

Falsification: Exponent outside $[0.3, 0.7]$.

Prediction 10.8 (Convergence Rate). Learning rate in repeated decisions $\approx 0.62 \pm 0.1$ (i.e., φ^{-1}).

Test: Fit learning curves; extract rate parameter.

Falsification: Rate outside $[0.5, 0.75]$.

Prediction 10.9 (Attentional Blink Recovery). Attentional blink recovery occurs at 350 ± 50 ms.

Test: RSVP paradigm with varying SOA.

Falsification: Recovery outside $[250, 450]$ ms.

Prediction 10.10 (Neural Frequency Ratio). Dominant neural oscillation frequencies during deliberation occur in ratio ≈ 1.6.

Test: EEG/MEG spectral analysis during decision tasks.

Falsification: No frequency pair with ratio in $[1.5, 1.7]$.

Prediction 10.11 (Sequential Attribute Processing). Multi-attribute decisions show sequential attribute-by-attribute processing.

Test: Eye-tracking and think-aloud during multi-attribute choice.

Falsification: Evidence for parallel attribute integration.

11 Discussion

11.1 Relation to Existing Models

Drift-Diffusion Models: The deliberation update (25) resembles DDM, with J' providing drift. The key difference is that our drift is derived from an axiomatic cost function rather than fitted.

Expected Utility: The cost J is not a utility function—it measures deviation from optimality, not preference magnitude. Unlike utility, J is uniquely determined.

Prospect Theory: While J is symmetric ($J(x) = J(1/x)$), the gradient $|J'(x)|$ is steeper for $x < 1$ than for $x > 1$ near $x = 1$, partially capturing loss aversion.

Information Geometry: Our Hessian metric construction follows Amari [2]. The connection to Fisher information provides statistical interpretation.

11.2 Limitations

1. **One-dimensionality**: The theory is developed for 1D state spaces. Multi-dimensional extension requires additional assumptions (separability or specific coupling).

2. **Capacity hypothesis**: The φ^3 bound is a hypothesis, not a derivation. It fits existing data but requires independent confirmation.

3. **Neural implementation**: No specific neural mechanism is proposed; the framework is computational, not mechanistic.

4. **Discrete structure**: The 8-tick constraint mentioned in related work is not derived here.

5. **Social decisions**: Extension to multi-agent settings is not addressed.

11.3 Future Directions

1. Multi-dimensional manifolds with coupled attributes

2. Neural network implementations of geodesic computation

3. Experimental tests of quantitative predictions

4. Application to social and game-theoretic decisions

5. Extension to continuous-time stochastic models

12 Conclusion

We have developed a geometric framework for decision-making based on the axiomatically unique cost functional $J(x) = \frac{1}{2}(x + x^{-1}) - 1$. The principal contributions are:

1. **Uniqueness**: The d'Alembert equation with normalization uniquely determines J (Theorem 2.5).

2. **Geodesics**: The explicit solution $\gamma(t) = (\alpha t + \beta)^{-2}$ with negative curvature throughout (Theorem 3.7, Proposition 3.4).

3. **Capacity**: The hypothesis $\sum \phi_k \leq \varphi^3 \approx 4.24$ matches empirical working memory bounds.

4. **Correspondence**: Decision and reference share formal structure (Theorem 5.5).

5. **Predictions**: Eleven falsifiable claims with specific numerical targets.

The framework provides a principled geometric foundation for decision theory. Whether it correctly describes cognition is an empirical question. If the predictions systematically fail, the theory is falsified.

Acknowledgments

The author thanks the Recognition Science Research Institute for support and Claude (Anthropic) for assistance with mathematical formalization and proof verification.

References

[1] J. Aczél, *Lectures on Functional Equations and Their Applications*. Academic Press, 1966.

[2] S. Amari, *Information Geometry and Its Applications*. Springer, 2016.

[3] N. Cowan, "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," *Behavioral and Brain Sciences*, vol. 24, no. 1, pp. 87–114, 2001.

[4] D. C. Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting*. MIT Press, 1984.

[5] M. P. do Carmo, *Riemannian Geometry*. Birkhäuser, 1992.

[6] G. Gigerenzer and W. Gaissmaier, "Heuristic decision making," *Annual Review of Psychology*, vol. 62, pp. 451–482, 2011.

[7] J. I. Gold and M. N. Shadlen, "The neural basis of decision making," *Annual Review of Neuroscience*, vol. 30, pp. 535–574, 2007.

[8] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106, no. 4, pp. 620–630, 1957.

[9] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–291, 1979.

[10] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2nd ed., 2003.

[11] S. J. Luck and E. K. Vogel, "The capacity of visual working memory for features and conjunctions," *Nature*, vol. 390, no. 6657, pp. 279–281, 1997.

[12] S. Mac Lane, *Categories for the Working Mathematician*. Springer, 2nd ed., 1998.

[13] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.

[14] R. Ratcliff and G. McKoon, "The diffusion decision model: Theory and data for two-choice decision tasks," *Neural Computation*, vol. 20, no. 4, pp. 873–922, 2008.

[15] H. A. Simon, "Rational choice and the structure of the environment," *Psychological Review*, vol. 63, no. 2, pp. 129–138, 1956.

[16] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

A Proof of Curvature Formula

For a 1D Riemannian manifold with metric $g(x) > 0$, the Gaussian curvature can be computed via:

$$K = -\frac{1}{2g}\left[\frac{\mathrm{d}^2 g}{\mathrm{d}x^2} - \frac{1}{2g}\left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2\right]$$

For $g(x) = x^{-3}$:

$$g' = -3x^{-4} \tag{34}$$

$$g'' = 12x^{-5} \tag{35}$$

$$(g')^2 = 9x^{-8} \tag{36}$$

Thus:

$$K = -\frac{x^3}{2}\left[12x^{-5} - \frac{9x^{-8}}{2x^{-3}}\right] = -\frac{x^3}{2}\left[12x^{-5} - \frac{9}{2}x^{-5}\right] = -\frac{x^3}{2}\cdot\frac{15}{2}x^{-5} = -\frac{15}{4}x^{-2}$$

The curvature is $K(x) = -\frac{15}{4x^2} < 0$ for all $x > 0$: the manifold is hyperbolic everywhere.

B Golden Ratio Values

Expression	**Value**	**Interpretation**
φ	1.6180339887...	Golden ratio
φ^2	2.6180339887...	$= \varphi + 1$
φ^3	4.2360679774...	Capacity bound hypothesis
φ^4	6.8541019662...	—
φ^{-1}	0.6180339887...	$= \varphi - 1$; annealing rate
φ^{-2}	0.3819660112...	—
φ^{-3}	0.2360679774...	—

Table 1: Powers of the golden ratio.

C Lean 4 Formalization

Core definitions and theorems are formalized in Lean 4. Example code:

```
-- Cost functional
noncomputable def J (x : Real) : Real := (x + 1/x) / 2 - 1

-- Geodesic curve
noncomputable def geodesic (a b t : Real) : Real :=
  1 / (a * t + b)^2

-- Geodesic equation satisfied
theorem geodesic_satisfies_eq (a b t : Real)
    (h : a * t + b != 0) :
    let g := geodesic a b
    deriv (deriv g) t + Gamma (g t) * (deriv g t)^2 = 0 := by
  sorry  -- Full proof in repository
```

Complete formalization: `IndisputableMonolith/Decision/*.lean`