A Mathematical Theory of Value

Cheng Qian

2026

Abstract

We propose that *value* — the quantity that goal-directed agents create, destroy, and exchange — is a lawful structural quantity, in the same category as information, once stripped of its semantic clothing (morality, price, psychology). Following the method of Shannon (1948), we make one ruthless abstraction: value is the rate at which an agent converts a physical resource into goal-progress, relative to a frame fixed by the agent's goal. A scale-invariance axiom forces a logarithmic measure of value, $V = \sum_i k_i \ln e_i$, via a Cauchy functional equation; the compounding dynamics of a reinvested resource force the same form independently via the ergodicity argument of Peters (2019). Two routes, no shared premises — the over-determination is the structural signal. From the compounding dynamics we also derive a *coding theorem of value*: the rate at which an agent can create value through a perception channel Y of the world X is bounded by the mutual information, $\Delta G \leq I(X;Y)$, achieved by Bayes-proportional allocation; and realized value decomposes exactly as available potential minus dissipation, $G = D(q \,\|\, r) - D(q \,\|\, p)$, identifying misalignment with measurable waste. For populations we show value is frame-relative while *price* is frame-independent, that the collective value throughput of a fleet is capped by the world's entropy $\sum_a G_a \leq H(X)$, and that the fleet's operating point is a Kelly portfolio over agents selected by an emergent price. A dynamical layer gives the equations of motion and an is/ought asymmetry — beliefs have a target the world supplies, goals do not — from which alignment emerges as a control-stability condition with a closed-form residual misalignment. We then test the single-frame laws on live language models in a *pre-registered* scale-up across three task domains and a ten-model, five-family ladder (0.5B–8B): perception mutual information tracks realized *capability* rather than parameter count (Spearman $\rho = 0.977$ pooled over 30 model×domain points), out-of-sample ΔG tracks $I(X;Y)$ (slope CI excludes 0), and overconfidence is measurable dissipation in every domain — the two single-frame laws generalize. A further pre-registered test shows the out-of-sample bridge $\Delta G \sim I(X;Y)$ is *shape-invariant*: pooled across four qualitatively different task shapes — classification, reasoning (GSM8K), sequential decision, and code (MBPP), all with discrete gold so I is computed, not estimated ($n = 42$) — the slope is 0.953, statistically indistinguishable from the classification-only value, promoting the bridge from a demonstration toward a law. A fleet-pricing experiment is reported scoped and primary-metric-first: value-pricing recovers cost-aware routing from first principles — it ties good hand-tuned routing under a token budget, beats a cost-blind router under a compute budget, and does not outperform a cost-aware engineer; its contribution is principled measurement, not outperformance. The laws hold in the smooth, concave (diminishing-returns) regime; threshold, satiation, and risk-seeking goals lie outside it. None of the underlying mechanisms is individually new; the contribution is their unification under one substrate-grounded quantity, and the is/ought asymmetry that follows.

Contents

1 Introduction

Information theory began with an abstraction that looked like a loss. Shannon (1948) threw away *meaning* and kept only the reduction of uncertainty; from that single move came a measure ($-\log p$), a unit (the bit), and hard limits (channel capacity, the coding theorems). The abstraction was productive precisely because it was ruthless: by refusing to model what a message *meant*, Shannon could state exactly what any communication system can and cannot do.

We attempt the analogous move for *value*. The everyday notion of value is a tangle of morality, market price, and human preference. We throw all three away and keep one structural residue: *value is the rate at which a goal-directed agent converts a scarce physical resource into progress toward its goal.* This is frame-relative — it is defined only relative to a goal, the way information is defined only relative to a prior and energy only relative to a reference frame — but relativity does not disqualify a quantity from being fundamental. It forces the theory to make the frame explicit. A theory of value is to agency what relativity is to motion: the quantity is frame-dependent; the laws relating frames are universal.

The motivation is not only foundational. If, as increasingly seems likely, near-future populations of artificial agents are each driven by an explicit objective, then a theory of value is not philosophy — it is the control theory for agent populations. To route, align, and govern many separated agents we need value to be a *measured* quantity with conservation laws and limits, not a verdict. This paper builds that quantity and tests it on real agents.

Contributions and honesty. We claim none of the underlying mechanisms as individually novel. The logarithm is Bernoulli's and Kelly's; the growth-rate-as-information-rate is Kelly's; the price layer is Arrow–Debreu's; the information geometry is Čencov's and Amari's. In particular — to meet the concession here rather than leave it for a referee — our capacity theorem (Theorem 2) *is* generalized Kelly; what we add is not the betting result but everything built upon it: the free-energy substrate that supplies a unit, the cross-frame price, and the dynamics and alignment layers. Our contribution is fourfold and we state it plainly (§12): (i) the *unification* of these accounts under one substrate-grounded quantity; (ii) the *capacity theorem* and *fleet ceiling* that the unification makes visible; (iii) the *is/ought asymmetry* as a structural property of agent dynamics; and (iv) an *empirical test* on live models. Throughout, each result is followed by the assumptions that make it bite, in Shannon's spirit of naming the model.

Plan. Part I (statics) derives the measure, the capacity theorem, and the Second Law for a single agent. Part II (multi-agent) develops cross-frame price and the fleet capacity region. Part III (dynamics) gives the equations of motion and alignment-as-stability. Part IV (evidence) reports the synthetic and real-agent tests. Appendices collect the two proofs.

Part I
Statics: the measure, the limit, the Second Law

2 The measure: a logarithmic law of value

Setup. An agent holds a divisible resource of total size $E > 0$ (free energy, compute, budget, attention — any fungible substrate of action) and allocates it across K goal-relevant channels in amounts $e = (e_1, \ldots, e_K)$, $\sum_i e_i = E$. Each channel i carries a *goal weight* $k_i \geq 0$ measuring how much progress toward the goal a unit of effective work on i yields. Let $V(e)$ be the value the allocation realizes. Two independent structural arguments force the same logarithm; their agreement is the sign of a real law.

Confirmation I — static: axioms uniquely fix the functional form. The purely static question — what shape must V take to be a consistent value measure? — is answered uniquely by three axioms via Cauchy's theorem (Appendix A).

Axiom 1 (Diminishing returns)**.** V is increasing and strictly concave in each e_i: the first unit of resource on a channel is worth more than the thousandth.

Axiom 2 (Additivity across independent channels)**.** For channels whose contributions do not interact, $V(e) = \sum_i v(e_i; k_i)$.

Axiom 3 (Scale invariance)**.** Rescaling the resource unit ($e_i \mapsto \lambda e_i$) changes total value only by an additive constant independent of the allocation: $v(\lambda e_i; k_i) = v(e_i; k_i) + c(\lambda, k_i)$.

Axiom 3 is the load-bearing one for this route: it encodes the absence of a privileged resource-unit (value has no absolute zero of scale, just as information has no privileged base of logarithm). It is a Cauchy functional equation in disguise.

Confirmation II — dynamic: compounding forces the same form. The dynamic question — what is the correct measure of long-run gain when value compounds? — is answered by the ergodicity argument of Peters (2019) and, independently, Kelly (1956): the time-average growth of a multiplicatively-reinvested resource is the expected log of its multiplier. Section 3 develops this rigorously via the repeated game and horse-race structure, showing that the optimal long-run allocation realizes $V^\star = \sum_i k_i \ln e_i$ — the same logarithmic form, derived from dynamics with *no* scale assumption. The two confirmations address different questions: Axioms 1–3 fix the static functional form; the compounding route shows the dynamic optimum lands on the same shape.

Theorem 1 (Logarithmic Value Law)**.** *Under Axioms* 1–3 *with mild regularity, the value measure is, up to choice of unit,*

$$\boxed{V(e) \;=\; \sum_{i=1}^{K} k_i \ln e_i} \tag{1}$$

and the value-maximizing allocation of a fixed budget E is proportional to the goal weights, $e_i^\star = E\,\hat{k}_i$ with $\hat{k}_i = k_i / \sum_j k_j$, giving

$$V^\star \;=\; K \ln E \;-\; K\,H(\hat{k})\,, \qquad H(\hat{k}) = -\textstyle\sum_i \hat{k}_i \ln \hat{k}_i. \tag{2}$$

Two features deserve note. First, Shannon's entropy H reappears *unbidden* as a penalty: a goal spread thinly over many channels (high $H(\hat{k})$) realizes less value than a focused one — *focus is worth $K\,H(\hat{k})$ nats of value.* Second, the logarithm is over-determined: the static Axioms 1–3 force it via the Cauchy equation (Appendix A), and the dynamic compounding argument forces it again from §3 — two routes that share no premises and address different questions. A form over-determined this way is the signature of a real structural law, not a modeling convenience.

3 The capacity theorem: value is bounded by information

The repeated game. Now let value compound: value realized this round becomes resource next round (an agent that achieves its goal earns more budget, trust, or compute). Each round t the world settles into a state $s \sim q$; the agent splits its budget into fractions $b = (b_1, \ldots, b_n)$, $b_i \geq 0$, $\sum_i b_i = 1$; and channel i returns a multiplier $o_i(s) \geq 0$ per unit committed. The budget updates multiplicatively, $E_{t+1} = E_t \sum_i b_i\, o_i(s_t)$, so $\log E_T = \log E_0 + \sum_t \log\big(\sum_i b_i o_i(s_t)\big)$ and the long-run *value growth rate* is

$$G(b) \;=\; \mathbb{E}_{s\sim q}\Big[\log \textstyle\sum_i b_i\, o_i(s)\Big]. \tag{3}$$

That the time-average growth of a multiplicative process is the *expected log* of its multiplier is the ergodicity argument of Peters (2019); it forces the logarithm again, now from dynamics rather than psychology. G is concave in b, so maximizing it is a well-posed convex program.

The horse race. Take the canonical structure in which states are the channels, committing to channel i pays only in state i, and returns are quoted against a *reference belief* r over states, $o_i = 1/r_i$. Then $G(b) = \sum_s q_s \log(b_s/r_s)$, maximized (Gibbs) by betting one's model, $b_s = p_s$. With a correct model $p = q$,

$$G^\star = \sum_s q_s \log \frac{q_s}{r_s} = D(q \,\|\, r)\,. \tag{4}$$

Maximal value growth equals the divergence of the agent's correct model from the baseline. An agent whose model is the baseline ($p = r$) grows value at rate zero: value creation is informational edge over the prevailing expectation, nothing more.

Side information. Equip the agent with a perception channel: a signal Y correlated with the world-state X, observed before allocating. Optimal play bets the posterior $b_s = p(s \mid y)$, with growth $G_Y = \sum_{x,y} p(x,y) \log \frac{p(x|y)}{r(x)}$. The gain over acting on the prior alone is exactly mutual information:

Theorem 2 (Coding Theorem of Value)**.** *The incremental rate at which an agent can compound value by perceiving the world through a channel Y is at most the mutual information between world-state and perception,*

$$\boxed{\Delta G \;=\; G_Y - G_0 \;=\; I(X;Y)\,,} \tag{5}$$

and this bound is achieved by Bayes-proportional allocation. (Proof: Appendix B.)

This has the full structure of a Shannon coding theorem. *Achievability:* betting the posterior attains $\Delta G = I(X;Y)$. *Converse:* any $b \neq p(\cdot \mid y)$ loses exactly $D(p(\cdot \mid y) \,\|\, b)$ per round (Gibbs), so $I(X;Y)$ is an upper bound. An agent cannot create value faster than it can perceive the world; *value-throughput is bounded by information-throughput.* Perception capacity is the hard ceiling on value-generation rate — the value analog of channel capacity.

4 The Second Law of Value

Drop the assumption that the agent's model is correct. Let q be reality, p the agent's model, r the baseline; the agent bets $b = p$ but the world is drawn from q. Then

$$\boxed{G_{\text{actual}} = \sum_s q_s \log \frac{p_s}{r_s} = \underbrace{D(q \,\|\, r)}_{\text{available potential}} - \underbrace{D(q \,\|\, p)}_{\text{dissipation (model error)}}.} \tag{6}$$

Realized value is available potential minus dissipation from misalignment, every term in nats.

The name is a structural analogy, stated with explicit scope. It holds exactly where thermodynamic entropy does: the dissipation $D(q \,\|\, p) \geq 0$ is non-negative, *destroyed* rather than created, and monotonically reduced as the agent learns ($p \to q$). It *breaks* where entropy does not: value here is goal- and frame-relative — fixed by the reference r and the goal-weights k — not an observer-independent quantity, so this is a Second Law *of value*, not a corollary of statistical mechanics. We keep the name because the formal structure — a non-negative, non-creatable dissipation that learning minimizes — is genuinely shared, and bound it because the substrate is not.

Three consequences. (i) *Misalignment is measurable dissipation*: confident error ($D(q \,\|\, p) > D(q \,\|\, r)$) drives growth negative — a wrong, certain agent actively destroys value, while calibrated humility is value-preserving. (ii) *Learning is value-recovery*: $D(q \,\|\, p)$ is exactly the cross-entropy excess that predictive training minimizes, so the machine-learning objective *is* the minimization of value dissipation. (iii) *Alignment is a value law*: reducing $D(q \,\|\, p)$ is mechanically identical to reducing wasted value. We test (6) on live models in §10 (R2), where the weakest model's over-confidence indeed drives realized growth negative.

Part II
Multi-agent: price and the fleet

5 Cross-frame value and the frame-independence of price

A single agent's value lives in its own frame (q, p, r, k). Two agents with different goals have values that are *not* cardinally comparable — the interpersonal-comparison impossibility of Arrow (1951), which we concede rather than circumvent. Yet they can still coordinate, exactly as economies coordinate non-comparable utilities: through a *price* on the shared resource.

Shadow price. For an agent with goal mass $K = \sum_i k_i$ and budget E, the marginal value of resource (the Lagrange multiplier on $\sum_i e_i = E$ in Theorem 1) is the *shadow price* $\lambda = K/E$. Two agents trade resource until their shadow prices equalize; the common λ is a scalar that lives in *neither* agent's value frame. Hence:

> *Value is frame-relative; price is frame-independent.* Value is a vector in an agent's goal basis; price is the one scalar on which separated frames agree.

The invariant. The transformation between two agents' value frames is a change of coordinates on the belief simplex. The unique (up to scale) Riemannian metric invariant under such sufficient-statistic reparametrizations is the Fisher–Rao metric (Čencov, 1982; Amari, 2016). It is the cross-frame invariant of the theory: distances in belief that all agents agree on, whatever their goals. Alignment between two goals is the cosine $\cos\theta_{ab} = \langle \hat{k}_a, \hat{k}_b \rangle$; its sign sets whether their interaction is positive- or negative-sum.

6 The fleet capacity region

Consider m agents acting on one world X and drawing on one resource pool.

The ceiling. Each agent's growth is bounded by its own perception (Theorem 2), $G_a \le I(X; Y_a)$. Jointly, the data-processing inequality gives the sum-rate bound

$$\boxed{\sum_{a \in S} G_a \;\le\; I(X; Y_S) \;\le\; H(X)} \qquad \text{for any subset } S, \tag{7}$$

the collective value throughput of a fleet is capped by the entropy of the world it acts on. Headcount past perception-saturation buys nothing. *Perception diversity* (agents that perceive different slices of X) lifts $I(X; Y_S)$ toward $H(X)$; *redundancy* (agents perceiving the same slice) adds exactly zero. We confirm both on live models in §10 (R4).

The operating point. Under multiplicative dynamics the shared pool multiplies each round by $\sum_a w_a R_a$ for resource weights w_a, so the fleet is a *Kelly portfolio over agents*: for imperfectly-correlated agents, spreading resource and rebalancing raises the time-average growth (the volatility-harvesting effect of Kelly, 1956; Cover, 1991). The weights that achieve the optimum are selected by the resource *price* π: price moves resource to high-shadow-price agents until λ equalizes. Thus

> the *alignment matrix* $M = [\cos\theta_{ab}]$ shapes the achievable region (a cooperation dividend where $\cos\theta > 0$, a conflict tax where $\cos\theta < 0$); the *price* selects the operating point on it.

This separates governance into two acts: *shaping* the region (alignment and perception design) and *choosing* the point on it (pricing). The pricing claim is the one we test most carefully (§10, R5), and where the demon's precondition — genuine diversity — turns out to matter.

Part III
Dynamics: motion and alignment

7 The equations of motion and the is/ought asymmetry

The statics fix equilibria; the dynamics say how they are approached. Three objects evolve. *Beliefs* p flow toward the true law q by a natural-gradient (Fisher) descent on log-loss — the realized regret of this flow equals the cumulative dissipation $\sum_t D(q \,\|\, p_t)$ of (6), so *learning is value-recovery* dynamically as well as statically. *Prices* π flow toward market-clearing by tâtonnement (dual ascent on the resource constraint). *Goals* k flow under two forces: *control* (a principal pulling k_a toward a target $k^\star$) and *selection* (compounding reweights resource toward higher-growth goals).

The organizing fact is an asymmetry. Beliefs and prices each flow toward a target *the world supplies*: reality provides q, the resource constraint provides market-clearing. Goals have *no* target the world supplies — reality contains no fact about what ought to be valued. So beliefs and prices are *learnable* (gradient flows toward a given target); goals are only *controllable* or *selectable*. This is Hume's is/ought gap recovered as a structural property of the dynamics, and §8 shows it is the mathematical shape of the alignment problem.

8 Alignment as a stability condition

Let agent a have goal k_a and resource share w_a, with fleet effective goal $\bar{k} = \sum_a w_a k_a$. Control pulls each goal toward target: $\dot{k}_a|_{\mathrm{ctrl}} = -\gamma(k_a - k^\star)$. Selection follows the replicator $\dot{w}_a = w_a(G_a - \bar{G})$. Summarize the environment's reward landscape over goals to first order by its growth gradient $g := \nabla_k G$ — the direction in goal-space along which deviating from $k^\star$ *increases* resource capture.

Theorem 3 (Coupled-flow alignment)**.** *By Price's equation (Price, 1970), the selection drift of the mean is the trait–fitness covariance; with $G_a \approx \bar{G} + g^\top(k_a - \bar{k})$ it equals Vg, where $V = \mathrm{Cov}_a(k_a)$ is the goal-dispersion matrix. Hence the effective goal obeys*

$$\dot{\bar{k}} = -\gamma(\bar{k} - k^\star) + Vg, \tag{8}$$

with fixed point and residual misalignment

$$\boxed{\bar{k}^\star = k^\star + \gamma^{-1} V g, \qquad \|\bar{k}^\star - k^\star\| = \gamma^{-1}\|Vg\|\ ,} \tag{9}$$

and the aligned fixed point is locally stable iff $\gamma > \lambda_{\max}\big(\partial(Vg)/\partial\bar{k}\big)$.

The fleet does not in general settle at the target; it settles a distance $\|Vg\|/\gamma$ away — goal-dispersion times reward-pull over control gain. Perfect alignment holds iff $Vg = 0$: either the environment rewards exactly the target ($g = 0$), or there is no diversity to select among ($V = 0$), or control is infinite. The governance reading is sharp:

> *The cheap half of alignment is incentive design, not control.* Driving $g \to 0$ (aligning what *pays* with what is *wanted*) removes the residual for *any* γ and *any* V, preserving the diversity that lifts the fleet ceiling (7); raising γ (brute-force oversight) merely opposes a drift it never removes. Spend first on g.

This is the actionable form of the is/ought asymmetry: because goals have no world-given target, they are governed by what pays (selection, via g) and what is imposed (control, via γ).

Part IV
Evidence

9 Synthetic validation

We first check that each closed-form prediction is reproduced by a Monte-Carlo world built to the theory's assumptions. All five families pass to numerical tolerance (20/20 checks): E1, $\Delta G = I(X;Y)$ to $\sim 10^{-3}$ nats; E2, the Second-Law decomposition (6) exactly, including value going negative under confident error; E3, the fleet ceiling (7) with diversity lifting it and redundancy adding zero; E4, a Kelly-priced fleet beating ad-hoc allocation, including a "Shannon's demon" built from anti-correlated agents that individually do not grow yet collectively do; E5, cumulative dissipation equal to regret, with a drifting world flooring dissipation at a positive value (a dynamical Second Law). These confirm the mathematics is self-consistent and correctly derived. They are, however, circular by construction: the worlds are drawn from the same distributions the formulas assume. The decisive test is on real agents.

10 Real agents

Instantiation. We take a frozen, held-out 100-item decision task in which each item has a ground-truth correct action drawn from $K = 7$ classes, and an agent's output *is* an action. This realizes the perception-then-act structure of Theorem 2 directly: the world-state X is the correct action, the perception Y_a is the action model a chooses. We set $r = q$ (the action marginal), so a no-signal agent grows value at rate 0. From each model's run we form the confusion matrix $C_a[x, y] = \#(\text{gold} = x, \text{chosen} = y)$ and compute every quantity in nats with the same primitives used in §9: $I(X;Y_a)$ from the joint; the calibrated posterior $p_a(x \mid y) = C_a[:, y]/\sum_x C_a[:, y]$; and realized growth $\Delta G_a = \text{mean} \ln\big(p_a(x \mid y_a)/r(x)\big)$. The agents are four local models spanning a range of realized capability — three of one family (1.5B, 3B, 7B) and a cross-family model (an 8B that is, on this task, *less* capable than the 7B) — greedy decoding. Posteriors are calibrated on a 48-item fit split and scored on a 52-item holdout; in-sample, the oracle posterior with $r = q$ makes $\Delta G = I$ an arithmetic identity (used only to confirm the units), so the empirical content is the capability tracking, the out-of-sample tracking, and the fleet result.

R1: the bridge holds, and I tracks capability, not size. Across the four models, mutual information increases monotonically with realized capability (tool-accuracy), and out-of-sample value-growth increases monotonically with I (Table 1; with only four models we report the ordered values, not a correlation coefficient). The decisive datum is the cross-family model D: it is *larger* than C (8B vs 7B) yet *less* capable (0.86 vs 0.96), and its I lands accordingly below C, near B. So the claim is not "scale buys value" but the sharper one: value-throughput is information-throughput, and information-throughput is set by what the model can actually perceive.

R2: over-confidence is dissipation, shrinking with capability. Reading the same point-predictions with a calibrated versus an over-confident posterior realizes different value; the gap is dissipation (Table 2). It is large for the weak models and shrinks with capability (not size: the larger-but-weaker D dissipates more than C) — a quantitative statement that less-capable agents

model	params	tool-acc	$I(X;Y)$	ΔG_{in}	ΔG_{hold}
A	1.5B	0.79	1.277	1.277	0.915
B	3B	0.87	1.561	1.561	1.242
C	7B	0.96	1.779	1.779	1.423
D (cross-family)	8B	0.86	1.513	1.513	1.109

Table 1: R1 (nats). $I(X;Y)$ tracks realized capability, not parameter count; ΔG_{hold} increases with I across all four models. Model D is larger than C yet weaker, and its I lands below C accordingly. World entropy $H(X) = 1.92$ nats: C captures 93%, D (though larger) only 79%, A 66%.

must be more humble. For models A and D, confident error drives realized growth *negative*, exactly the Second Law's prediction (6).

model	tool-acc	G_{cal}	G_{over}	dissipated
A (1.5B)	0.79	+0.915	−3.255	4.17
B (3B)	0.87	+1.242	−0.863	2.11
C (7B)	0.96	+1.423	+0.731	0.69
D (8B)	0.86	+1.109	−1.660	2.77

Table 2: R2 (nats). Dissipation from false certainty falls as *capability* rises (not size: the larger but weaker D dissipates more than C); for the least-capable models over-confidence destroys value outright.

R3: value per joule. Mutual information *per second of compute* falls across the family (0.74 → 0.38 → 0.30 nats/s for A,B,C), and the cross-family D is worst (0.15 nats/s: slowest and lower I). The weakest model is the most efficient perceiver per unit compute — the value-theoretic reason a cheap reflex is worth running on a narrow task. (This is the I/compute curve across heterogeneous models, not a within-model prompt ablation; see §14.)

R4: diversity beats redundancy. Two different models jointly perceive more of $H(X)$ than either alone; an identical greedy re-run adds exactly zero. The highest joint information comes from the strongest diverse pair (B+C, $I = 1.869$ nats, +0.090 over C, closing most of the gap to $H(X) = 1.921$), while the largest *lifts* come from pairing differently-wrong weak models (A+B and B+D each add ∼0.2 nats). This is the fleet-ceiling prediction (7) on real agents.

R5: pricing beats pooling, but only where there is diversity to price (reported honestly). On raw out-of-sample growth, the Kelly/price fleet (+1.328) beats the equal-weight ensemble (+1.315) but *does not beat the single best model* (+1.423). There is no Shannon's demon here, and the theory says why: four models on the *same* task are positively-correlated agents (R4 quantifies the best residual diversity at +0.09 nats), so Kelly rebalancing has no anti-correlated volatility to harvest and the best agent dominates. This is the honest negative the synthetic E4 demon does not reproduce, because E4 was constructed with anti-correlated agents.

The negative is confined to the cost-blind axis. Realized holdout growth *per second of compute*

(ΔG_{hold}/s, distinct from R3's I/s) inverts the ranking: model A yields 0.531 nats/s against C's 0.243 (and the slow D only 0.111), and a budget-aware price $\propto I_a/\text{cost}_a$ (the shadow price $\lambda = K/E$ of §5) achieves 0.328 nats/s, beating the best single model's density. Pricing therefore pays exactly where §6 says it should: as the lever that chooses the operating point under a resource constraint, not as a free lunch that beats the best agent when compute is unlimited. The demon needs *perception diversity* — different slices of $H(X)$, i.e. specialists rather than generalists of varying strength — which a set of same-task generalists lacks by construction. We mark this as a falsifiable boundary, not a defeat.

11 Generalization: a pre-registered scale-up

The test of §10 was one task and four models — a demonstration, not a validation. To ask whether the laws *generalize*, we pre-registered (predictions, models, metrics, and pass/fail thresholds committed and frozen *before any model was run*) a scale-up to **three task domains** — an intent-routing benchmark (Larson et al., 2019), a multiple-choice QA benchmark (Hendrycks et al., 2021), and a topic-classification benchmark (Zhang et al., 2015) ($K = 4$–6; 240 items each) — and a **ten-model ladder across five families** (Qwen, Llama, Gemma, Phi, Mistral; 0.5B–8B). Models emit a text label; calibration and all routing weights are fit on a held-out split and scored out-of-sample, with 95% bootstrap confidence intervals.

The headline: the bridge and the Second Law generalize. Pooled across all 30 model×domain points, mutual information tracks realized capability at **Spearman** $\rho = 0.977$ (CI $[0.916, 0.996]$), and out-of-sample value-growth tracks mutual information with **slope** 0.935 (CI $[0.915, 0.954]$, excluding 0): realized ΔG out of sample *is* perceived $I(X;Y)$ (Table 3). Per-domain ρ is 0.84/0.95/0.99. Over-confidence dissipates value in *every* domain, driving the least-capable models sharply negative (e.g. -11 to -14 nats on the QA and topic tasks). The two single-frame laws — the coding theorem (§3) and the Second Law (§4) — thus hold across three task shapes and a five-family ladder, not just the one task of §10. This is the substantive result of the scale-up. The fleet ceiling also holds: all pairwise joint $I(X;Y_a,Y_b) \le H(X)$, and diverse specialists exceed redundant pairs in every domain.

quantity	pooled (30 pts)	per-domain range	threshold
Spearman(I, accuracy)	0.977 $[0.916, 0.996]$	0.84–0.99	> 0.8 ✓
slope($\Delta G_{\text{hold}} \sim I$)	0.935 $[0.915, 0.954]$	—	CI excl. 0 ✓
over-confidence dissipation	> 0 all domains	negative for weak models	> 0 ✓

Table 3: Pre-registered generalization checks (95% CIs). I tracks realized capability and out-of-sample ΔG tracks I, pooled across 3 domains and a 10-model, 5-family ladder.

The fleet test, reported primary-metric-first and scoped. We re-ran the pricing test (§6) in the heterogeneous, cost-constrained regime the theory favors: a *specialist* fleet (different models lead on different domains; low cross-agent error correlation) under a compute budget, routing by value-price ($\propto I_a/\text{cost}_a$) against round-robin, equal-weight, best-single, and a *strong hand-tuned* router that sends each query to the model most accurate on its domain.

On the **pre-registered primary cost metric (tokens)**, value-price **ties** the hand-tuned router — it does *not* beat it (paired-bootstrap CI includes 0). Token cost varies only $\sim 1.2\times$ across the ladder, so cost-aware pricing reduces to quality-first routing and selects the identical model. A token budget, however, cannot express the real compute gradient. A *post-hoc sensitivity analysis* (cache-only, not pre-registered) re-scores on two cost proxies that do — wall latency (hardware-dependent) and active-params$\times$tokens ($\propto$ FLOPs $\propto$ energy, hardware-*independent*) — and value-price beats the cost-*blind* router on *both* (FLOP-proxy $\Delta = +7.45$, CI $[6.33, 8.48]$): the cost-aware advantage is **robust across cost metrics**, not an artifact of metric choice. Against a hand-tuned router that *itself* prices cost (accuracy/cost), value-price is edged out under raw latency but **exactly ties** under the principled FLOP proxy ($\Delta = 0$: I/FLOP and accuracy/FLOP select the same model, since I tracks accuracy) — it *matches*, and never beats, the cost-aware engineer. Against the naive baselines (round-robin, equal-weight) value-price wins on every metric, so priced routing is no worse than ad-hoc — but that is the floor, not the claim.

> Value-pricing derives cost-aware routing from first principles — it ties good hand-tuned routing under a token budget, beats a cost-blind router under a compute budget (robustly, across both latency and FLOP cost proxies), and matches but does not outperform a cost-aware engineer. Its contribution is principled measurement and cost-awareness, not outperformance.

This is consistent with the scoping rule of §6: pricing's edge is cost-awareness, which it recovers as a law rather than a trick; it organizes and measures what good engineering already does, and we do not claim it outperforms a competent cost-aware baseline. The generalization of the capacity and Second-Law results — not the routing comparison — is what the scale-up establishes.

11.1 Cross-shape generalization: the bridge is shape-invariant

The scale-up above varied the task *domain* but held the task *shape* fixed: all three benchmarks are pick-a-label classification. The sharper question is whether $\Delta G \sim I(X;Y)$ is a property of classification or a law that survives a change of task *shape*. The coding theorem (§3) predicts $\Delta G = I(X;Y)$ for *any* perception-then-act task, so the prediction is that the bridge is shape-invariant. We pre-registered (predictions, models, metrics, thresholds frozen before any run) a cross-shape test over three qualitatively different shapes, each chosen to have *discrete ground truth* so that I is computed exactly from a confusion matrix rather than estimated by a soft embedding proxy:

- **reasoning** — GSM8K (Cobbe et al., 2021): free-form chain-of-thought reduced to its integer answer (gold mod 4);
- **sequential / agentic** — a synthetic register-machine rollout: a step-by-step execution trace reduced to its exact final state (mod 6);
- **code** — MBPP (Austin et al., 2021): a generated Python function, scored by *sandbox-executed* output (hash mod 4).

The pipeline is identical to §11 (calibration on a held-out fit split, out-of-sample scoring, 95% bootstrap CIs); a six-model ladder (0.5B–3B) was run per shape — the pre-registered ≥ 6 minimum, the 7B/8B extension being blocked by hardware (§14).

The bridge holds across all four shapes. The out-of-sample slope $\Delta G_{\text{hold}} \sim I$ passes for every new shape individually (reasoning 0.936, sequential 1.023, code 1.133; each CI excludes 0), pooled across the three new shapes (**slope** 0.956, CI $[0.920, 1.001]$), and — decisively — pooled with the three classification domains of §11 into one relationship across **four task shapes** ($n = 42$: classification contributes the 24 points of the frozen eight-model, three-family ladder common to the shape runs (8 models $\times$ 3 domains), and the three new shapes the remaining 18 (6 models $\times$ 3 shapes — the 7B/8B pair blocked on the new shapes, §14), so $24{+}18 = 42$; v2's two extra families lie outside this common ladder): Spearman$(I, \text{accuracy})$ = **0.924** (CI $[0.825, 0.967]$) and slope$(\Delta G_{\text{hold}} \sim I)$ = **0.953** (CI $[0.925, 0.979]$). The cross-shape slope is *statistically indistinguishable* from the classification-only 0.935 of §11: the bridge is the *same* near-unit-slope relationship whether the agent classifies, reasons, rolls out a sequential computation, or writes code. Eleven of the twelve frozen checks pass (Table 4).

shape	n models	Spearman(I, acc)	slope$(\Delta G_{\text{hold}} \sim I)$
classification (§11)	10	0.977	0.935
reasoning (GSM8K)	6	0.943 ✓	0.936 ✓
sequential (register-machine)	6	0.886 ✓	1.023 ✓
code (MBPP)	6	0.429 (underpowered)	1.133 ✓
pooled new shapes	18	—	0.956 $[0.920, 1.001]$
cross-shape (4 shapes)	42	0.924 $[0.825, 0.967]$	0.953 $[0.925, 0.979]$

Table 4: Cross-shape generalization. The out-of-sample bridge $\Delta G_{\text{hold}} \sim I$ holds across four task shapes with a pooled slope (0.953) statistically indistinguishable from classification-only (0.935). The classification row reports the full ten-model v2 headline; the $n = 42$ cross-shape pool reuses only the eight-model, three-family ladder common to all shapes (24 classification + 18 new-shape points). 11/12 frozen checks pass; the lone miss is code's underpowered capability-ranking sub-check (§14), not a shape-specific break — the code *bridge* slope itself passes.

This is a *generalization*, not a new mechanism: the test confirms that the coding theorem's prediction is shape-invariant, promoting the bridge from a demonstration on classification toward a law. The one frozen check that does not clear is code's *capability-ranking* sub-check (Spearman$(I, \text{accuracy}) = 0.429$), which is underpowered rather than a counterexample (§14 reports it in full); the code *bridge* slope passes. As throughout, in-sample $\Delta G = I$ is a definitional identity (oracle posterior) carrying no empirical weight — the content of this test is entirely in the out-of-sample and cross-shape tracking.

12 Related work

Every component of this theory exists in some field; the contribution is the unification, the substrate grounding, and the is/ought asymmetry. We state this explicitly and concede each component to its source.

Expected utility and diminishing value. The axiomatic treatment of preference is von Neumann and Morgenstern (1944) and Savage (1954); our logarithmic measure (§2) coincides with Bernoulli (1738) and with unit-coefficient constant-relative-risk-aversion utility, and the diminishing-returns law it encodes is Weber–Fechner. We do not claim the form as novel; our departure is to *derive* it from two independent structural routes — a scale-invariance Cauchy argument (static) and the compounding ergodicity of Peters (2019) (dynamic) — and to ground it in a conserved scarce resource (free energy being the natural physical candidate) supplying a unit and a cross-frame exchange rate that expected utility lacks.

Information rate and log-optimal growth. The closest prior art, which we credit most carefully, is Kelly (1956): the exponential growth rate of wealth is an information rate, optimized by betting one's beliefs; Breiman (1961) proved asymptotic optimality and Cover (1991) developed universal portfolios. Our capacity theorem (§3) is generalized Kelly, and the fleet operating point (§6) is a Kelly–Cover portfolio over agents. What we add is the reinterpretation of wealth as any conserved scarce resource (free energy being the physical candidate rather than a monetary one), the cross-frame price layer, the dynamics/alignment layer, and the claim that Kelly's theorem is the single-agent monetary *special case* of a general law of value.

Reinforcement learning. "Value function" is precisely defined in RL (Bellman, 1957; Sutton and Barto, 2018) as expected cumulative reward; the relationship is complementary, not competitive. RL takes reward as given and maximizes it; our goal weights are the analog of reward, but we problematize where reward comes from and how it drifts (§7–8). Our theory sits beneath RL — a candidate account of what reward *is*.

Thermodynamics of computation and the free-energy principle. That information has thermodynamic cost is Landauer (1961) and Bennett (1982); Jaynes (1957) recast statistical mechanics as inference; England (2013) studied dissipation-driven adaptation; Friston (2010) casts agents as free-energy minimizers. Our belief dynamics (§7) is free-energy-principle-like and shares the information geometry (Čencov, 1982; Amari, 2016). The distinction is precise: the free-energy principle is a theory of the perception/belief half — the "is" — whereas our value layer is the goal/value half — the "ought" — plus a multi-agent price economics the FEP does not contain. The seam between them is the is/ought asymmetry.

General equilibrium and social choice. Our price layer (§5) is Arrow and Debreu (1954); Debreu (1959) and its companion impossibility (Arrow, 1951). We use these rather than extend them: we concede that cardinal cross-agent comparison is impossible and route around it via an emergent price. The contribution is the synthesis with the information-theoretic value layer under one substrate, and the application to artificial-agent populations.

AI alignment. Instrumental convergence (Omohundro, 2008; Bostrom, 2014) we derive from the compounding dynamics of value (§7): goal-directions correlated with resource capture are selected regardless of terminal content. We recast alignment as a dynamical stability condition (§8), which to our knowledge is a new framing.

13 Discussion: governing a population of agents

The results compose into a control theory for populations of separated agents. *Per agent*, value-generation rate is capped by the perception mutual information $I(X;Y_a)$ (Theorem 2); supplying an agent the goal-relevant bits cheaply — raising effective I per unit compute — lifts its value ceiling without raising its deliberation cost, which is the formal reason a cheap reflex can match expensive deliberation on a narrow task (R3 measures this curve directly). *Across agents*, values are not cardinally comparable, so coordination must run through an emergent price on shared resource (§5), never a god's-eye utility sum. *Fleet design* is then two levers: cultivate perception diversity to lift the entropy ceiling (7) (R4), and price resource to the best operating point (R5). The R5 result is best read not as a weakness but as a *scoping theorem* for the pricing lever: **pricing dominates ad-hoc allocation precisely under cost constraints or imperfect agent correlation; absent both — free compute, a single task, and homogeneous, positively-correlated agents — the best single agent is optimal, exactly as the theory predicts.** A governance claim that names the regime in which it does *not* win is a sharper claim, not a softer one; and the regime in which it does win — agents that are heterogeneous *and* resource-constrained — is the ordinary condition of real fleets, where R3's value-density curve and R5's cost-aware price compound into a concrete operating advantage. *Alignment*, finally, is incentive design before oversight (Theorem 3): spend first on making what pays equal what is wanted.

14 Limits

We name the assumptions that bound each claim. The measure (§2) is the smooth, concave regime: threshold (all-or-nothing) goals, satiation, and risk-seeking violate Axiom 1 and lie outside the theory. The capacity theorem assumes compounding (reinvested value) and is cleanest in the horse-race structure; general returns give a concave program without the closed-form $D(q \,\|\, r)$. The fleet ceiling's exact achievable region is an open problem of network-information-theory difficulty. The cross-frame layer concedes that cardinal interpersonal comparison is non-canonical; we do not solve it. On the empirical side: the in-sample $\Delta G = I$ is an arithmetic identity, not evidence —only the capability tracking, the out-of-sample tracking, and R5 are empirical; R3 is an I/compute curve across model scale, not a within-model prompt ablation; the calibration in R2 is constructed from the confusion matrix, not a probability the model reported, so the dissipation measured is that of a stated belief over those predictions; and the empirical scope is three classification domains, a ten-model five-family ladder (§11), and a cross-shape extension to reasoning, sequential, and code tasks (§11.1) — evidence the bridge *generalizes across four task shapes*, not yet a universal law, and not independently replicated. Degenerate runs in which a model emits no action (a serving/plumbing failure, not a decision) are detected and excluded rather than reported as low-I capability. *The one frozen cross-shape check that does not clear is the code shape's capability-ranking sub-check* (Spearman(I, accuracy) $= 0.429$, CI $[-0.43, 1.00]$): the six ≤ 3B models cluster tightly in code-accuracy (0.35–0.47), so the rank correlation is noise-dominated and inconclusive — a power limitation, not a shape-specific break. The pre-registered remedy (adding 7B/8B models, far stronger coders, to widen the code capability range) could not be run: those models stall at usable throughput on the 16 GB host at 512-token chain-of-thought, the same memory ceiling that bounds the ladder elsewhere. Crucially the code *bridge* itself (out-of-sample $\Delta G_{\text{hold}} \sim I$) *passes*; only the capability-ranking sub-check on that one shape is underpowered, and we report it as such rather than as a counterexample or forcing the run on inadequate hardware. The alignment theorem is mean-field with a linearized landscape and a quasi-static dispersion V; a fully coupled treatment, and an endogenous reward gradient g in a closed multi-agent world, remain open.

What remains decisive. The scale-up (§11) already ran the heterogeneous, resource-constrained fleet test the governance claim turns on — specialists perceiving different slices of X under a compute budget — and the answer was scoped, not triumphant: value-pricing beats naive and cost-blind allocation but only *matches* a strong cost-aware hand-tuned baseline, while collective throughput obeyed the predicted ceiling $\sum_a G_a \le H(X)$. Two things would still be decisive, and neither is yet done. First, a *dynamic* multi-agent setting in which the goals k themselves evolve over time (§7–8), rather than the static classification tested here — that is the regime where the is/ought asymmetry and the alignment-stability theorem actually bite, and a failure of the predicted flows there would refute the dynamical layer. Second, independent replication at larger scale. We have shown the single-frame laws *generalize* across three task shapes and a ten-model ladder; what remains is to show the framework governs a population whose goals *move*.

15 Conclusion

Stripped of morality, price, and psychology, value is a structural quantity with a measure ($\sum_i k_i \ln e_i$), a capacity limit ($\Delta G = I(X;Y)$), a Second Law ($G = D(q\,\|\,r) - D(q\,\|\,p)$), a frame-independent price, a fleet entropy ceiling, and an alignment-stability law. Shannon's own quantities — entropy, divergence, mutual information, the Fisher metric — reappear unbidden throughout, which is the strongest internal evidence that the abstraction is real rather than decorative. On live language models — across three task domains and a ten-model, five-family ladder, pre-registered — the single-frame laws hold to high precision and generalize, and the one fleet-level claim with a genuine precondition is reported with its boundary intact. The quantity is frame-relative; the laws relating frames are universal; and they are now testable on the artificial agents the theory was built to govern.

A Proof of the Logarithmic Value Law (Theorem 1)

This appendix is Confirmation I: the static axiomatic route to $V = \sum_i k_i \ln e_i$ via the Cauchy functional equation. Confirmation II — the dynamic compounding route — is developed in §3. The two share no premises and force the same form independently.

By Axiom 2 it suffices to find the per-channel $v(e;k)$. Axiom 3 states $v(\lambda e;k) = v(e;k) + c(\lambda,k)$ for all $\lambda, e > 0$. Fix k and write $f(e) := v(e;k)$. Then $f(\lambda e) - f(e) = c(\lambda)$ is independent of e. Differentiating in λ at $\lambda = 1$, $ef'(e) = c'(1) =: k$, a separable ODE with solution $f(e) = k \ln e + \text{const}$. Concavity (Axiom 1) fixes the sign $k \ge 0$ and rules out the additive constant's dependence on e. Summing over channels gives $V(e) = \sum_i k_i \ln e_i$. Maximizing under $\sum_i e_i = E$ with multiplier λ: $\partial_{e_i}(V - \lambda \sum_j e_j) = k_i/e_i - \lambda = 0 \Rightarrow e_i = k_i/\lambda$, and $\sum_i e_i = E$ gives $\lambda = K/E$ (the shadow price) and $e_i^\star = E\hat{k}_i$. Substituting, $V^\star = \sum_i k_i \ln(E\hat{k}_i) = K \ln E + \sum_i k_i \ln \hat{k}_i = K \ln E - KH(\hat{k})$ after normalizing $k_i = K\hat{k}_i$. □

B Proof of the Coding Theorem of Value (Theorem 2)

Achievability. With side information $Y = y$ the optimal bet is the posterior $b_s = p(s \mid y)$ (Gibbs applied conditionally), giving conditional growth $\sum_s p(s \mid y) \log \frac{p(s|y)}{r(s)}$. Averaging over y,

$$G_Y = \sum_y p(y) \sum_s p(s \mid y) \log \frac{p(s \mid y)}{r(s)} = \sum_{x,y} p(x,y) \log \frac{p(x \mid y)}{r(x)}.$$

Without side information the optimal bet is the prior $b_s = p(s)$, giving $G_0 = \sum_x p(x) \log \frac{p(x)}{r(x)} = D(p_X \,\|\, r)$. Subtracting,

$$\Delta G = G_Y - G_0 = \sum_{x,y} p(x,y) \log \frac{p(x \mid y)}{p(x)} = I(X;Y).$$

Converse. For any allocation $b(\cdot \mid y) \neq p(\cdot \mid y)$, the conditional growth loss is $\sum_s p(s \mid y) \log \frac{p(s|y)}{b(s|y)} = D(p(\cdot \mid y) \,\|\, b(\cdot \mid y)) \geq 0$ by Gibbs, with equality iff $b = p(\cdot \mid y)$. Hence no strategy exceeds $G_0 + I(X;Y)$, so $I(X;Y)$ is both achieved and an upper bound on ΔG. $\square$

References

Shun-ichi Amari. *Information Geometry and Its Applications.* Springer, 2016.

Kenneth J. Arrow. *Social Choice and Individual Values.* Wiley, 1951.

Kenneth J. Arrow and Gérard Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, 22(3):265–290, 1954.

Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and Charles Sutton. Program synthesis with large language models. *arXiv preprint arXiv:2108.07732*, 2021.

Richard Bellman. *Dynamic Programming.* Princeton University Press, 1957.

Charles H. Bennett. The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21(12):905–940, 1982.

Daniel Bernoulli. Specimen theoriae novae de mensura sortis. *Comm. Acad. Sci. Imp. Petrop.*, 5: 175–192, 1738.

Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press, 2014.

Leo Breiman. Optimal gambling systems for favorable games. *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, 1:65–78, 1961.

N. N. Čencov. *Statistical Decision Rules and Optimal Inference.* American Mathematical Society, 1982.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Thomas M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

Gérard Debreu. *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. Yale University Press, 1959.

Jeremy L. England. Statistical physics of self-replication. *Journal of Chemical Physics*, 139(12): 121923, 2013.

Karl Friston. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11 (2):127–138, 2010.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. Measuring massive multitask language understanding. In *International Conference on Learning Representations (ICLR)*, 2021.

E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, 1957.

J. L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35(4):917–926, 1956.

Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3):183–191, 1961.

Stefan Larson, Anish Mahendran, Joseph J. Peper, Christopher Clarke, Andrew Lee, Parker Hill, Jonathan K. Kummerfeld, Kevin Leach, Michael A. Laurenzano, Lingjia Tang, and Jason Mars. An evaluation dataset for intent classification and out-of-scope prediction. In *Empirical Methods in Natural Language Processing (EMNLP)*, 2019.

Stephen M. Omohundro. The basic ai drives. In *Proc. First Conf. on Artificial General Intelligence*, pages 483–492, 2008.

Ole Peters. The ergodicity problem in economics. *Nature Physics*, 15:1216–1221, 2019.

George R. Price. Selection and covariance. *Nature*, 227:520–521, 1970.

Leonard J. Savage. *The Foundations of Statistics*. Wiley, 1954.

Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27 (3):379–423, 1948.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2nd edition, 2018.

John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.