Schooled Minds for Standardized Worlds: A Model of Cognitive Variation

Matthew Cashman (cashman@mit.edu)[1], T.C. Zeng[2] & Ivan Kroupin[3]
[1]Sloan School of Management, Massachusetts Institute of Technology
[2]Department of Psychology, Harvard University
[3]Comparative Cultural Psychology Group, Max Planck Institute for Evolutionary Anthropology

Abstract

We present a formal account of cognitive variation between schooled and unschooled populations by modeling minds as constrained mappers of different types of territories. Using Rate Distortion Theory, we model minds as learning a finite codebook of reusable tools (concepts, heuristics, schemas) under limits on capacity, attention, and per-use costs. Schooled worlds provide standardized information ecologies, enabling compact, portable representations (“Red Means Stop”). Low-standardization ecologies reward high-fidelity, niche-tuned maps that sacrifice portability (“Tracking deer by the river is hard when the wind blows from the water”). We introduce a standardization weight (α) capturing developmental exposure and derive distinct optima under different ontogenetic distributions. Environmental standardization is defined by the slope of the distortion—rate curve: in highly standardized settings, small informational investments yield broad performance gains, favoring abstraction; in low-standardization settings, gains require encoding nuanced local detail. The framework predicts trade-offs, universally helpful tools, and mismatch losses across ecologies.

Keywords: culture; cultural evolution; rate-distortion theory

Introduction

The human mind is a map-maker. It builds compressed representations of the world around it in order to navigate effectively (Sims, 2016; Tishby et al., 2005). Given that schooled and unschooled populations show striking differences on a variety of cognitive dimensions (Kroupin et al., 2024; Luriâ, 1976; Scribner & Cole, 1981), we might then wonder: is this evidence of some fundamental difference in cognitive abilities, or perhaps— is this simply a matter of unschooled maps being a poor fit to schooled territories? We propose that one feature driving the difference between schooled and unschooled minds is a difference in the scope of the map an individual builds: schooled minds are tuned to represent schooled environments, which are themselves evolved and deliberately built to be highly standardized—and so, very easily compressible Red means STOP in all developed environments. In contrast, unschooled minds are tuned to represent environments which are much less standardized—environments where an additional bit spent encoding a feature of the environment captures much less than in a more standardized environment. In addition to this, the representations in schooled minds are much more portable, being largely re-deployable to novel standardized environments. This is in contrast to unschooled minds, which are tuned to a specific, idiosyncratic environment and which therefore struggle to represent aspects of novel environments.

To characterize this difference, we might draw an analogy to 3D modeling in computer graphics. A “low-poly” model uses a small number of polygons to approximate a shape; it is computationally efficient and portable, capturing the essential structure of an object (e.g., a generic tree or building) while discarding its specific irregularities. Conversely, a “high-poly” model uses a vast number of polygons to capture every texture and unique detail. It is a very high fidelity model of a particular tree, but is computationally expensive and difficult to use on trees in general. It may not just be a maple, it may be a *specific* maple - and it is therefore not a good representation of an oak. Simultaneously, a low-poly model of a rectilinear steel, glass, and concrete building may in fact be quite high fidelity because the building is in fact built of simple polygons. We argue that human cognition faces a similar trade-off in the face of environments that are more or less amenable to representation with simple shapes.

Schooled minds approximate a universal atlas for modern, technologically-advanced (WEIRD; Henrich, 2020) environments. While they may reflect a particularly good fit to a person’s specific environment relative to others (e.g., New York vs. Vienna), New York and Vienna have many standardized features (like "Red Means Stop") such that a mind adapted to New York functions just fine if dropped into Vienna. To continue the computer graphics metaphor, schooled minds invest in “low-poly” abstractions—generic polygons like right angles, categories, and logic—that work reasonably well everywhere in modernity, especially in the built world of modernity which has been engineered (or which has evolved) such that it is in fact built of generic polygons. Unschooled minds, in contrast, are more like detailed local guidebooks. They are optimized to produce a high-fidelity model of a specific place and manner of being. They invest in high-fidelity representations of specific textures, local causal chains, and cues that are specific to a given niche. They provide maximum competence locally, but fail to generalize. We formalize this using Rate-Distortion Theory (RDT), treating the mind as an encoder optimizing a codebook $\mathbf{\Phi}$ under a capacity constraint.

Throughout, we use “map”, “model”, and “tool” as broad functional terms. Our basic assumption is that, in a given ecology, having representations that support efficient action is generally advantageous—though we do not claim that cognition optimizes for veridical truth, nor that all adaptive behavior requires explicit internal models. Formally, our distortion function $d(x, r)$ should be read as a reduced-form proxy for downstream costs (errors, missed opportunities, coordination failures) incurred when an agent acts on representation r in situation x, holding fixed whatever policy maps representations to actions. Likewise, a "tool" ϕ can be a concept, heuristic, procedure, or schema: any reusable primitive that helps compress observations into action-relevant summaries. We aspire to a model of representational tradeoffs under constraints, not a complete account of decision-making.

We also draw on Kroupin and Zeng (2024) for the cultural–ecological motivation and for their terminology around depth of standardization: the extent to which everyday coordination is mediated by reusable, explicitly taught interfaces (artifacts, procedures, symbols) that can be recombined across contexts. In their framing, "Lean" systems have high depth of standardization and contain many "Small Games" that can be navigated with relatively context-independent "thin rules" (e.g., forms, grades, traffic lights, timetables), whereas "Rich" systems have low depth of standardization and contain more "Big Games" governed by context-dependent "thick rules". We treat these distinctions as applying to environments E (and to an individual's ontogenetic mix P_α^{ont}), not as a claim about "more information" in any absolute sense. In our notation, "high-standardization" corresponds to the Lean/Small-Game/thin-rule/high-legibility end of their spectrum, and "low-standardization" to the opposite end. Higher standardization concentrates action-relevant uncertainty into shared cues, while lower standardization leaves more task-relevant variance in the residual. For simplicity we use *standardization* throughout the paper.

Our formulation is related to the Information Bottleneck framework (Tishby et al., 2005), which studies representations that compress observations while preserving information about an explicitly specified relevance variable. Here, we do not assume a fixed relevance target; instead, relevance is implicit in the ecology-specific distortion function $d(x, r)$, interpreted as downstream action cost. Our approach is also aligned with cognitive applications of rate–distortion theory (Sims, 2016), but we shift focus from within-task perceptual encoding to the learning and maintenance of a finite, reusable toolbox (a codebook $\mathbf{\Phi}$) across a distribution of ecologies. To this we contribute a cultural–ecological scope: standardization in ontogeny (captured by α) changes which abstractions are worth storing, and this predicts systematic mismatch losses when a codebook optimized for one ecology is deployed in another. In addition, we highlight the coevolution of schooled minds and built environments, and the deliberate construction of easily-compressible worlds as key factors in enabling the success of the schooled mind.

Empirically, a large cross-cultural and historical literature documents systematic cognitive differences between schooled and unschooled populations, and recent syntheses argue that these schooling effects pervade many domains studied in cognitive science (Kroupin et al., 2024). In Luria's classic studies, unschooled adults often treat syllogistic premises and taxonomic categories as inappropriate in the absence of concrete experience, preferring practical judgments over abstract inference (Luriâ, 1976). Yet later work shows that performance depends on how tasks invite context-stripping: when participants are prompted to reason in an explicitly hypothetical frame, unschooled adults can reason from unfamiliar or counterfactual premises much more accurately (Dias et al., 2005). In literacy research, Scribner and Cole argue that schooling and literacy cultivate specific cognitive routines rather than a uniform increase in general capacity, with strongest benefits on tasks that mirror schooled practices (Scribner & Cole, 1973, 1981). More broadly, education and literacy systematically shape performance even on "nonverbal" neuropsychological measures, challenging assumptions of cultural fairness (Ardila et al., 2010; Rosselli & Ardila, 2003), and quasi-experimental evidence suggests additional schooling can causally raise IQ scores (Brinch & Galloway, 2012).

Theoretical Framework

The Universe of Environments ($\mathcal{E}$)

Let $\mathcal{E}$ be the set of all possible environments. Each environment $E \in \mathcal{E}$ corresponds to a specific probability distribution $p(x \mid E)$ over observations $x \in \mathcal{X}$ (e.g., "village market", "forest hunt", "fifth-grade classroom").

We distinguish between two broad classes of environments based on the statistical structure of $p(x \mid E)$:

- *High-standardization environments:* These environments concentrate probability mass near a low-dimensional manifold determined by a few explicit rules and shared interfaces. The distribution $p(x \mid E)$ is highly compressible (e.g., traffic lights, arithmetic problems).
- *Low-standardization environments:* These environments have high-dimensional, slowly decaying structure. The distribution $p(x \mid E)$ is complex and fractal; predicting the next observation requires tracking many variables (e.g., wind, scent, season).

Both categories can be high in raw sensory complexity, and agents will tend to operate near their cognitive constraints. What distinguishes high- from low-standardization in this paper is task-relevant compressibility: highly standardized environments supply shared interfaces that collapse action-relevant uncertainty efficiently, while low-standardization environments leave more action-relevant uncertainty in the residual even after attending to the best available cues. For instance, E_{market} might have fixed prices with a fixed rule structure for buying and selling, while E_{barter} generates highly variable social exchanges where the things exchanged depend on relationship, reputation, and local scarcity.

We also distinguish between the environment where an agent learns and the environment where they are tested: $P_\alpha^{\text{ont}}(E)$ is the *ontogenetic* distribution encountered during development. The parameter $\alpha \in [0, 1]$ acts as a *standardization weight*, shifting mass toward high-standardization environments. In contrast, $P^{\text{test}}(E)$ is the *test* distribution where performance is evaluated (e.g., a lab task, a forest). From this, we model two main sources of variation. First are *ecologies* ($E \in \mathcal{E}$), varying in degree of standardization (e.g., exams and bureaucratic offices vs. informal markets and subsistence hunting; (Kroupin & Zeng, 2024)). Second are *life histories* (P_α^{ont}), the mixture of ecologies a person regularly encounters while growing up (high α = many standardized settings; low α = mostly idiosyncratic, local settings).

What is optimized

For each life history, we assume the mind learns a finite *toolbox* or codebook $\mathbf{\Phi}$ (concepts, heuristics, schemas, representations). This toolbox is chosen to keep distortion low so that everyday actions succeed in the environments the person actually faces, to respect a fixed capacity limit K that determines which tools occupy the codebook, and to control moment-to-moment effort so that applying tools does not demand excessive attention on each new situation.

Our formal objective combines error cost, captured by how often the compressed mental map leads to worse outcomes in a given ecology ($d(x,r)$), with rate cost, measured by how many "bits" of attention or working memory are needed to apply the toolbox in that ecology ($I_E(X;R)$), under the capacity constraint that the agent can learn and keep only a fixed budget of distinct tools ($|\mathbf{\Phi}| = K$). Different schooling regimes and ecologies correspond to different points in this trade-off: schooled minds invest more in portable tools that work across many standardized contexts, while unschooled minds invest more in high-fidelity tools finely tuned to a particular niche.

The Map and The Territory

As a first analytical result, we formalize the map/territory distinction to make precise the representational commitments that generate the predictions below.

The Cognitive Codebook ($\mathbf{\Phi}$) We model the mind's knowledge as a *codebook* or set of prototypes, denoted $\mathbf{\Phi} = \{\phi_1, \phi_2, \ldots, \phi_K\}$. In Rate-Distortion Theory, a codebook is a finite set of representation points (centroids) used to approximate the continuous input space. Here, we use the term to represent the fundamental cognitive primitives—concepts, heuristics, categories, etc.—that an agent uses to parse the world. Just as a vector quantizer maps a continuous signal to the nearest code vector, the mind maps complex environmental stimuli to discrete mental states. We model the mind as having a finite capacity K (representing memory, attention, or learning time). Any observation x in the world is encoded (compressed) by selecting a tool $\phi \sim \pi(\cdot \mid x)$ from this codebook and mapping it via:

$$R = f(x; \mathbf{\Phi}, \phi)$$

where f is the encoding function (e.g., nearest-neighbor match or probabilistic assignment). When we suppress the selection policy, we write $R = f(x; \mathbf{\Phi})$.

Here, a codebook element $\phi \in \mathbf{\Phi}$ denotes a *cognitive tool*: a learned symbol, procedure, heuristic, schema, or callable routine that can be invoked by the encoder. Formally, each ϕ is a representation point in $\mathcal{R}$ together with an associated decoding/acting routine, so the encoder can make a hard assignment (choose a single ϕ) or a soft assignment (mix over ϕ via $\pi(\phi \mid x)$).

We distinguish the discrete tool budget $K = |\mathbf{\Phi}|$ from an information-rate capacity C (e.g., a constraint on $I(X;R)$). In this paper, K is the binding resource for cross-group comparisons, while C captures limits on encoding bandwidth or deployment precision. Soft assignments can be interpreted as using more bits (higher $I(X;R)$), but we treat C as fixed or implicit and focus on how different environments allocate the K slots.

Tools may be present but only partially learned. We model this with a proficiency parameter $s_\phi \in [0,1]$ that enters distortion as $d(x, \phi; s_\phi)$, with monotone improvement in proficiency (i.e., $s'_\phi > s_\phi \Rightarrow d(x,\phi;s'_\phi) \le d(x,\phi;s_\phi)$ for all x). This lets a tool occupy a slot yet contribute imperfectly to compression and action.

For example, one element ϕ_k in $\mathbf{\Phi}$ might correspond to a very general schema like "official document" (covering passports, permits, and certificates). Another might correspond to a highly specific local script such as "how to bargain with the same fish seller every Tuesday". Both are "tools", but the first is portable and at least somewhat useful across many contexts, while the second is finely tuned to a particular niche. High fidelity in a familiar niche sacrifices portability: each slot buys low distortion only in a small region of $\mathcal{X}$ (or a particular environment), leaving fewer slots for cross-context abstractions.

Separately, an agent can possess highly general, portable theories (e.g., basic physical laws) that are correct in principle but operationally irrelevant because applying them would require costly observation, binding, and computation. In our framework this shows up as a high deployment cost C_{deploy} (Section), not as high-fidelity representation. Throughout, we treat the capacity parameter K as fixed across schooled/unschooled comparisons; the difference is what occupies the K slots and how tools are deployed.

The Utility of a Polygon: Distortion Reduction We define the utility of a basis element $\phi \in \mathbf{\Phi}$ in environment E as the marginal reduction in distortion it provides:

$$\Delta D(\phi, E) = \min_{r \in \mathbf{\Phi} \setminus \{\phi\}} \mathbb{E}_{x \sim p(x|E)}[d(x,r)] - \min_{r \in \mathbf{\Phi}} \mathbb{E}_{x \sim p(x|E)}[d(x,r)]$$

Intuitively, $\Delta D(\phi, E)$ asks: if we remove this particular tool ϕ from the mind's toolbox, how much worse does the person do, on average, in environment E? If performance barely changes, the tool is not very useful in that environment; if performance worsens sharply, the tool is doing important work there.

Because of this, net usefulness depends on deployability. The definition above isolates the *representational* value of ϕ assuming it can be deployed appropriately. In practice, a tool may reduce distortion in principle but still be unused if its deployment costs are too high. A simple way to express this is

$$U(\phi, E) = \Delta D(\phi, E) - \eta\, \mathbb{E}_{x \sim p(x|E)}\left[C_{\text{deploy}}(\phi, x, E)\right],$$

where C_{deploy} aggregates the costs of instantiating and applying a tool (attention, observation, variable binding, computation), and $\eta > 0$ weights these deployment costs relative to distortion reduction.

The Cost of Deployment: Availability vs. Capability Possessing a basis function ϕ is not enough; the agent must also know *when* to apply it. We distinguish between the *storage cost* of a tool (included in the capacity constraint $|\mathbf{\Phi}|$) and the *deployment cost* of applying that tool to a specific observation x.

We model the decision to use a tool as a selection policy $\pi(\phi \mid x)$. The difficulty of applying an abstract tool (like a logical syllogism) to a detailed, concrete observation (like a conversation about bears) lies in the cost of directing attention appropriately. In a low-standardization ("Rich") environment, the abstract concept ϕ (e.g., "All bears are X") can fail to capture the relevant variance (e.g., "This specific bear is angry and very near"), resulting in high distortion in spite of the cognitive cost of storing ϕ. We operationalize this as the KL-divergence between the ecology's habitual input distribution and the attention-weighted distribution required by the tool:

$$C_{\text{suppress}}(\phi, E) = \mathrm{D}_{\mathrm{KL}}\big(p_E(x) \,||\, p_E(x \mid \text{attention}_\phi)\big)$$

Formally, let attention_ϕ define a reweighting of features such that $p_E(x \mid \text{attention}_\phi) \propto p_E(x) a_\phi(x)$, where $a_\phi(x)$ upweights the dimensions relevant to ϕ and downweights the rest. The KL term then measures how much the agent's habitual input distribution must be "bent" to make ϕ applicable. Intuitively, this may be how unnatural it feels, in this ecology, to treat only the features relevant to ϕ as important and to ignore the rest. In Luria's classic syllogism tasks (Luriâ, 1976), the experimenter asks the participant to attend only to the abstract logical form ("all bears in the North are white..."), and to ignore their detailed empirical knowledge about actual bears. For someone whose daily survival depends on tracking concrete animal behavior, this kind of context-stripping is not just unusual; it is cognitively costly.

Attentional suppression is only one component of what makes a tool usable. More generally, we define a per-use deployment cost for applying tool ϕ to observation x in environment E:

$$\begin{aligned} C_{\text{deploy}}(\phi, x, E) = {} & C_{\text{suppress}}(\phi, E) + C_{\text{obs}}(\phi, E) \\ & + C_{\text{bind}}(\phi, x) + C_{\text{comp}}(\phi, x). \end{aligned} \tag{1}$$

Here C_{obs} captures the cost of gathering the inputs the tool requires (measurement, data collection), C_{bind} captures the cost of mapping a messy situation onto the tool's variables (search / variable binding), and C_{comp} captures the cost of carrying out the computations or simulations the tool prescribes. For example, germ theory can reduce distortion across many environments, but its net usefulness depends on whether the context supplies cheap observation and binding (instrumentation, diagnostic proxies, sanitation routines) so that C_{obs} and C_{bind} do not swamp $\Delta D(\phi, E)$. For germ theory, standardization can reduce these observation and binding costs by supplying shared measurement conventions and institutional scaffolding. In contrast, standardization cannot reduce costs enough to make fundamental physical laws useful day-to-day despite their being strictly correct.

We include these deployment costs as explicit penalties in the optimization problem below, alongside rate costs and a fixed toolbox budget.

Schooled minds have learned to suppress context (set attention weights to 0 for non-relevant features). This is "cheap" because it is practiced daily. Unschooled minds have learned that context is *always* relevant in low-standardization ecologies. Suppressing context (setting weights to 0) is cognitively expensive or "unnatural" because their prior $p(\text{context} \mid \text{signal})$ is high.

This explains the gap in competence or performance observed by Luria 1976. The Uzbek peasant possesses the latent capacity for logic (it is not discarded), but the deployment cost is prohibitively high in natural conversation. When the experimenter lowers this cost by reframing the task (e.g., "Imagine a world where..."), they artificially reduce the friction of context-stripping, allowing the tool to be deployed.

The Optimization Problem

With the representational vocabulary fixed, we can state the optimization that yields contrasting theoretical optima under different developmental ecologies.

We model the cognitive agent as an encoder $q(r|x)$ that maps observations $x \in \mathcal{X}$ to mental representations $r \in \mathcal{R}$. Following Rate-Distortion Theory ((Berger, 1971; Imel & Zaslavsky, 2024)), the agent seeks to minimize distortion $d(x, r)$ subject to constraints on cognitive resources. Here, $q(r \mid x)$ is induced by selecting a tool via $\pi(\phi \mid x)$ and mapping deterministically through f, i.e., $q(r \mid x) = \sum_{\phi \in \Phi} \pi(\phi \mid x)\, \mathbf{1}[r = f(x; \Phi, \phi)]$, where Φ denotes the learned codebook $\mathbf{\Phi}$.

We distinguish between three distinct components of cognitive cost:

- *Capacity Constraint* ($|\mathbf{\Phi}| = K$)*:* A fixed budget of tools the agent can learn and keep available (the opportunity cost of one tool is another tool not learned).
- *Transmission Rate* ($I(X; R \mid \mathbf{\Phi}, \pi)$)*:* The working memory or attentional cost of using the codebook to encode a specific observation on the fly.
- *Deployment Cost* (C_{deploy})*:* the per-use cost of instantiating and applying tools (attention, observation/data-gathering, variable binding/search, computation/simulation).

Schooling environments push agents to fill a fixed tool budget with portable, abstract tools and to learn deployment policies that make these tools cheap to use. In our model, standardized contexts are those where task-relevant uncertainty is mediated by shared cues, so x can often be encoded at low rate and low deployment cost with minimal distortion. Conversely, low-standardization ecologies push agents to allocate more of the same fixed tool budget to locally tuned tools and to spend more rate and deployment cost per situation to minimize distortion in a particular niche.

We introduce a mixing parameter $\alpha \in [0, 1]$, representing the standardization weight (i.e., developmental exposure to standardized environments). The agent optimizes a codebook $\mathbf{\Phi}$ to minimize the expected loss over its *ontogenetic* distribution $P_\alpha^{\text{ont}}(E)$:

$$\begin{aligned}\mathbf{\Phi}^*(\alpha) = \underset{\substack{\mathbf{\Phi},\pi \\ |\mathbf{\Phi}|=K}}{\arg\min} \Bigg\{ \mathbb{E}_{E\sim P_\alpha^{\text{ont}}} \Bigg[& I_E(X; R \mid \mathbf{\Phi}, \pi) \\ & + \beta\, \mathbb{E}_{\substack{x\sim p(x|E) \\ \phi\sim\pi(\cdot|x)}} d(x, f(x; \mathbf{\Phi}, \phi)) \\ & + \gamma\, \mathbb{E}_{\substack{x\sim p(x|E) \\ \phi\sim\pi(\cdot|x)}} C_{\text{deploy}}(\phi, x, E) \Bigg] \Bigg\}\end{aligned} \tag{2}$$

Here $I_E(X; R \mid \mathbf{\Phi}, \pi)$ denotes the mutual information between X and R induced by the encoder and deployment policy when $X \sim p(x \mid E)$. The objective follows the classic RD Lagrangian convention: the rate term plus β times expected distortion, while deployment cost is scaled separately by γ and is not multiplied by β. The constraint $|\mathbf{\Phi}| = K$ encodes a fixed tool budget shared across schooled/unschooled comparisons. If $\mathcal{E}$ is discrete, then $\mathbb{E}_{E\sim P_\alpha^{\text{ont}}}[\cdots]$ reduces to $\sum_{E\in\mathcal{E}} P_\alpha^{\text{ont}}(E)[\cdots]$.

One ambiguity is where the situation-specific parameter values required to instantiate a tool appear in the objective. If we treat them as part of the representation R, then the bits needed to specify them are charged in the rate term $I_E(X; R \mid \mathbf{\Phi}, \pi)$. By contrast, C_{deploy} is intended to capture overhead that makes those bits available in the first place (observation, variable binding/search, computation); we separate these terms to avoid double-counting the same informational content.

Crucially, on this model "schooled" and "unschooled" minds do not differ in their internal optimization process or capacity parameters. They differ only in the distribution $P_\alpha^{\text{ont}}(E)$ they were exposed to during development.

The Schooled Optimum (High α) In schooled, high-standardization societies, much of the ontogenetic environment is organized around standardized interfaces (Kroupin & Zeng, 2024). In our model, these interfaces collapse many contextual factors into shared cues, so a given amount of cognitive bandwidth yields large reductions in distortion in many everyday contexts. The distribution $P_\alpha^{\text{ont}}(E)$ is concentrated on standardized contexts (classrooms, bureaucracies, grocery stores, what an elevator is and how to use it).

- *Strategy:* The agent learns a "low-poly" encoder. Because the environment is standardized, x can be compressed into a very low-rate representation r (e.g., a category label or a variable x) with minimal distortion.
- *Result:* The basis set $\mathbf{\Phi}^*_{\text{school}}$ consists of portable, abstract tokens.
- *The Mismatch Cost:* When this agent enters a low-standardization environment E_{raw} (where $P^{\text{test}}(E_{raw}) = 1$), their low-rate encoder filters out "noise" that is actually "signal.", and they lack some cognitive tools useful in that niche altogether. The distortion $d(x, r)$ spikes because the abstract basis set cannot reconstruct the high-frequency details of the local niche.

The Unschooled Optimum (Low α) In unschooled, low-standardization societies, the agent faces a specific, locally idiosyncratic niche E_{local} where task-relevant variance is spread across many interacting variables and is less mediated by standardized cues; reducing distortion typically requires encoding more local detail per situation.

- *Strategy:* The agent optimizes for *local fidelity*. They allocate their bit-budget (capacity) to encode the specific variances of E_{local}.
- *Result:* A "high-fidelity" map $\mathbf{\Phi}^*_{\text{unschool}}$. The mutual information $I(X; R)$ is high, but specific to $p(x|E_{local})$.
- *The Portability Cost:* Because the codebook is optimized for the specific statistics of E_{local}, it generalizes poorly. Insofar as $\mathbf{\Phi}^*_{\text{unschool}}$ contains representations that could be applied to novel environments, the "deployment cost" of using these dense, entangled representations in a standardized, abstract context (e.g., a syllogism task) is prohibitive. It simply does not contain many useful abstractions (like "Red Means Stop") that would reduce distortion in such environments.

Predictions

Environmental Standardization and Cultural Practice

To connect the optima to environmental structure, we characterize environmental standardization as a property of the distortion–rate curve that shapes which tools are worth learning and deploying. Let $R_E(D)$ be the standard Rate-Distortion function for the source distribution $p(x|E)$, representing the minimum rate required to achieve distortion D. Conversely, let $D_E(R)$ be the Distortion-Rate function, representing the minimum distortion achievable with rate R. An operational definition of environmental standardization can be had via the derivative of the Distortion-Rate function:

$$\text{Std}(E) = \left|\frac{\partial D_E(R)}{\partial R}\right| \quad \text{at a target rate } R^*.$$

At the RDT optimum for the RD part of Eq. 2, the envelope condition implies $|\partial D_E(R)/\partial R| = 1/\beta$. Thus, steeper $D_E(R)$ (larger $|\partial D_E/\partial R|$) corresponds to a smaller β, i.e., each extra bit yields a larger reduction in distortion. Under high standardization, the $D(R)$ curve is steep. A small investment in information (e.g., knowing the traffic rule "Red Means Stop") reduces distortion massively. The environment is "compressible." Schooled minds optimize for these steep gradients, achieving low distortion with minimal rate. The environment has been built, or has evolved, to facilitate this sort of compression. Under low standardization, the $D(R)$ curve is shallow (heavy-tailed). To reduce distortion in a social interaction or a hunt, one must encode a vast amount of nuance; each additional bit yields diminishing returns. While there can be useful regularities ("Jaguars usually hunt at night"), they are often less portable across contexts and less mediated by shared interfaces. Unschooled minds optimize to traverse this shallow curve.

Highly standardized environments exhibit steeper distortion–rate drops, so as α increases the optimal codebook shifts toward low-poly tokens that exploit those gradients. This improves performance in standardized environments but increases mismatch loss when the learned codebook is deployed in low-standardization environments. When α is lower, the trade-off reverses and higher-fidelity codebooks are favored. Taken together, these analytical components yield concrete performance predictions about cross-context testing.

Alignment with Empirical Findings

Our framework offers a unified interpretation of several robust findings in the cross-cultural and historical cognitive literature (e.g., (Kroupin et al., 2024); (Kroupin & Zeng, 2024)).

- *Luria's Syllogisms and Empirical Bias:* Unschooled participants often reject the premise of counterfactual syllogisms, preferring empirical verification (Luriâ, 1976). In our model, this reflects a high deployment cost (C_{suppress}). The abstract premise requires suppressing the high-dimensional prior $p(x)$ (where bears are brown and context matters) in favor of a narrow, counter-factual attention distribution. For a mind optimized for low-standardization ecologies, this suppression is prohibitively expensive. Prompting strategies ("Imagine a world where...") effectively lower C_{suppress} by explicitly re-weighting the attention distribution, temporarily moving the task toward the high-standardization end of the spectrum where the unschooled agent's performance improves.
- *Raven's Matrices and Fluid Intelligence:* Raven's Progressive Matrices are a widely used nonverbal test consisting of sets of pattern-completion items of increasing difficulty, intended to measure "eductive" (meaning-making) ability (Raven & Raven, 2003). In our terms, these tasks sit near the high-standardization end of our spectrum ($|\partial D/\partial R|$ is steep). Schooled minds, possessing a codebook $\mathbf{\Phi}^*_{\text{school}}$ optimized for such standardized structures, operate near the Rate-Distortion frontier, whereas unschooled minds, whose codebooks are tuned for the heavy-tailed distributions of the natural world, use capacity representing irrelevant features of the test setting and so show lower fluid intelligence scores.
- *The Flynn Effect as a Standardization Trajectory:* The secular rise in IQ scores (Flynn Effect) tracks long-run increases in environmental standardization (Flynn, 1987). One driver of the difference, in addition to things like much improved nutrition, is changes in average cognitive toolkit. As the macro-scale parameter α increases, the ontogenetic distribution P_α^{ont} shifts toward standardized structures. Later cohorts therefore acquire codebooks $\mathbf{\Phi}$ composed of more portable basis functions characteristic of standardized environments. This shift naturally predicts rising performance on lab-style tasks—which rely on these portable bases—followed by a plateau as environments become saturated with standardization.

Discussion

The schooled/unschooled distinction reflects a trade-off between generalization and fidelity. Schooled minds carry a universal atlas that allows them to function adequately in any standardized port of call, but they lack the detailed representations required to function effectively in less standardized environments. Modernity can be viewed as a co-evolutionary process where we adopt increasingly abstract cognitive maps and simultaneously terraform environments into standardized structures (building roads, standardizing time, regulating behavior) so that our maps become accurate. The schooling effect is the result of measuring cognition within the walls of this constructed, standardized world. The objective function we lay out suggests that, given a limited mental toolbox, one should choose a set of tools $\mathbf{\Phi}$ that (i) keep everyday mistakes small, on average, in the environments the person actually grows up in (P_α^{ont}), (ii) do not require too much moment-to-moment attentional effort to apply (the rate term), and (iii) fit within a fixed toolbox budget ($|\mathbf{\Phi}| = K$). Different values of α correspond to different life histories: a high-α agent spends much of childhood and adolescence in standardized settings (school, bureaucracy), whereas a low-α agent spends more time in irregular, locally idiosyncratic ecologies. Future work might investigate this model's predictions with longitudinal or cross-cultural studies that vary exposure to standardized institutions and task standardization, and with agent-based or neural simulations that learn rate–distortion codebooks under different α and K. Alternative accounts (capacity, familiarity, modularity, ecological rationality, gadgets, and cultural learning; e.g., (Arumugam et al., 2022; Boyd & Richerson, 1985; Dias et al., 2005; Fodor, 1983; Gigerenzer & Todd, 1999; Heyes, 2018)) make distinct predictions about how performance should vary with task standardization and deployment costs; our framework predicts the strongest dissociations when low-standardization tasks demand high C_{suppress} or when the codebook lacks portable abstractions that pay off in standardized tasks. Discriminating tests therefore hinge on manipulating standardization, exposure (α), and deployability rather than format alone. A limitation is that α and K are latent and may be non-identifiable from cross-sectional data, so decisive evidence requires within-culture designs that perturb exposure or deployment costs directly.

References

Ardila, A., Bertolucci, P. H., Braga, L. W., Castro-Caldas, A., Judd, T., Kosmidis, M. H., Matute, E., Nitrini, R., Ostrosky-Solis, F., & Rosselli, M. (2010). Illiteracy: The Neuropsychology of Cognition Without Reading. *Archives of Clinical Neuropsychology*, *25*(8), 689–712. https://doi.org/10.1093/arclin/acq079

Arumugam, D., Ho, M. K., Goodman, N. D., & Roy, B. V. (2022, October 30). *On Rate-Distortion Theory in Capacity-Limited Cognition & Reinforcement Learning*. arXiv: 2210.16877 [cs]. https://doi.org/10.48550/arXiv.2210.16877

Berger, T. (1971). *Rate distortion theory: A mathematical basis for data compression*. Prentice-Hall.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. University of Chicago Press.

Brinch, C. N., & Galloway, T. A. (2012). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences*, *109*(2), 425–430. https://doi.org/10.1073/pnas.1106077109

Dias, M., Roazzi, A., & Harris, P. L. (2005). Reasoning From Unfamiliar Premises: A Study With Unschooled Adults. *Psychological Science*, *16*(7), 550–554. https://doi.org/10.1111/j.0956-7976.2005.01573.x

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*(2), 171–191. https://doi.org/10.1037/0033-2909.101.2.171

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Henrich, J. P. (2020). *The WEIRDest people in the world: How the West Became Psychologically Peculiar and Particularly Prosperous* (First edition). Farrar, Straus and Giroux.

Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press.

Imel, N., & Zaslavsky, N. (2024). Optimal compression in human concept learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*(0). Retrieved August 15, 2024, from https://escholarship.org/uc/item/7pc1g61d

Kroupin, I., Davis, H. E., Zeng, T. C., & Henrich, J. (2024, August 13). *Back to school: How formal education has shaped almost all of cognitive science*. https://doi.org/10.31234/osf.io/tqy2n

Kroupin, I., & Zeng, T. C. (2024, August 13). *From Richer to Leaner cultures and minds: How depth of standardization as an evolutionary dynamic drives changes in technology, social structure, and cross-cultural cognitive variation*. https://doi.org/10.31234/osf.io/mxfcv

Luriâ, A. R. (1976). *Cognitive development: its cultural and social foundations* (M. Cole, Ed.; M. Lopez-Morillas, Trans.). Harvard university press.

Raven, J., & Raven, J. (2003). Raven progressive matrices. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 223–237). Kluwer Academic/Plenum Publishers.

Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition*, *52*(3), 326–333. https://doi.org/10.1016/S0278-2626(03)00170-2

Scribner, S., & Cole, M. (1973). Cognitive Consequences of Formal and Informal Education. *Science, New Series*, *182*(4112), 553–559. http://www.jstor.org/stable/1737765

Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Harvard University Press.

Sims, C. R. (2016). Rate–distortion theory and human perception. *Cognition*, *152*, 181–198. https://doi.org/10.1016/j.cognition.2016.03.020

Tishby, N., Pereira, F. C., & Bialek, W. (2005, September 17). *The information bottleneck method*. arXiv: physics/0004057. https://doi.org/10.48550/arXiv.physics/0004057