A Preference for Traceable Agency

An Axiomatization of Expected Utility plus Mutual Information*

Daniele Condorelli
University of Warwick

25th August 2026

Abstract

An action can matter twice: through the payoff it produces and the visible trace it leaves. A channel maps actions into consequences; preferences rank action lotteries in each channel. Data-processing conditions and a recordwise sure-thing principle, with standard axioms, characterize expected utility plus λ times the mutual information between action and consequence. Relative to a fixed payoff-utility scale, the theorem identifies λ, the price of trace in payoff units; expected utility is the case $\lambda = 0$. The criterion reads the coupling of act and consequence, not consequences alone; randomisation can be strictly optimal. When channel rows have pairwise-disjoint supports, the optimal lottery has multinomial-logit probabilities; when traces overlap, substitution depends on the similarity of their consequence distributions and independence of irrelevant alternatives can fail.

---

*Email: d.condorelli@gmail.com. All parts of this manuscript have been written with assistance from a multi-model LLM workflow; the author personally contributed very little to the writing of the proofs beyond encouragement. A machine-checked Lean 4 proof of Theorem 1 is available at `https://github.com/dcond/TraceableAgency`. All claims, errors, and omissions remain the sole responsibility of the author.

1 Introduction

Shower upon him every earthly blessing ... He would even risk his cakes and would deliberately desire the most fatal rubbish, the most uneconomical absurdity ... simply in order to prove to himself—as though that were so necessary—that men still are men and not the keys of a piano. ... If you say that all this, too, can be calculated and tabulated ... then man would purposely go mad in order to be rid of reason and gain his point!

— Fyodor Dostoevsky, *Notes from the Underground*, Part I, Chapter VIII [18]

Dostoevsky's Underground Man faces a choice problem. The cakes are material payoff, and he will risk them so that his conduct can attest to being his own, not the mechanical implication of a table. Yet the proof he seeks is self-defeating. An act fixed in advance cannot provide it, and deliberate rebellion, too, can be "calculated and tabulated." So he would "purposely go mad in order to be rid of reason"—not because madness is what he values, but because it is his last imagined way of preventing the table from closing before he acts. This paper takes the evidentiary idea literally. In Dostoevsky's example, conduct is observed directly: the act supplies its own evidence. We consider the general case, in which an act is observed only through a possibly noisy payoff-bearing outcome and an explicit visible record.

For visible consequences to carry information about an act, two conditions are needed. There must be uncertainty beforehand about which act will occur, and the consequence must discriminate among the possible acts. Either condition can fail on its own. A lottery over acts leaves no trace when every act generates the same distribution of visible consequences; a consequence that perfectly reveals the act conveys nothing new when the act was already known. We call the intrinsic preference for the two together a preference for *traceable agency*. The paper axiomatizes it jointly with ordinary preferences over payoff risk.

Let A be a finite set of actions, O a finite set of payoff-bearing outcomes, R a finite set of explicit visible records, and $K : A \to \Delta(O \times R)$ the channel relating actions to visible consequences. An action lottery $q \in \Delta(A)$ gives the distribution of the realised action. The primitive is a family $\{\succeq_K\}_K$, with one binary relation $\succeq_K$ on $\Delta(A)$ for each fixed channel K. The statement $q \succeq_K p$ means that q is weakly preferred to p in environment K. Eight axioms are imposed on this family. To compare lotteries governed by different channels, the channels are placed side by side as blocks of one larger environment; duplicating a channel or adding an unused block does not alter the comparison at issue. Two processing axioms say that the agent never gains when her record is garbled or her action variable is processed noisily, and a benchmark comparison rules out indifference to trace. A recordwise sure-thing principle for records structured in stages supplies the cardinal structure. No audience appears in the model, and the taste is content-neutral: it prices how much the consequence reveals, not what it reveals.

The main result is an if-and-only-if characterization. A family $\{\succeq_K\}_K$ satisfies the eight axioms if and only if there are a nonconstant payoff utility $u : O \to \mathbb{R}$ and $\lambda > 0$ such that, for every finite joint channel K and every $q, p \in \Delta(A)$,

$$q \succeq_K p \quad \Longleftrightarrow \quad \mathbb{E}_{qK}[u(O)] + \lambda I_{qK}(A; O, R) \geq \mathbb{E}_{pK}[u(O)] + \lambda I_{pK}(A; O, R),$$

where $I_{qK}(A; O, R)$ is the mutual information between the action and its visible consequence. We call such families *trace-tempered*. The utility u is unique up to positive affine transformations; after fixing that scale and the logarithm base, λ is measured in payoff units per unit of information, and the same pair prices comparisons across channels, embedded as blocks of one larger environment. A sign trichotomy completes the picture: reversing all three axioms that carry the taste for trace characterizes $\lambda < 0$—a preference for deniability—while replacing them by indifference clauses yields expected utility, $\lambda = 0$.

The information term has a simple interpretation. Observing (O, R) induces a posterior belief about the realised action; traceability is the expected reduction in that uncertainty, measured by Shannon entropy H. If visible consequences are independent of actions, the value is zero. If the action is known in advance because q is degenerate, the value is also zero. If the visible-consequence supports of the actions in the support of q are pairwise disjoint, the value is $H(q)$—Dostoevsky's case, in which conduct is observed directly.[1] Thus the criterion is not a taste for randomisation as such. Randomisation matters only because it creates uncertainty about agency for the visible consequence to resolve.

The taste is for the coupling of act and consequence. A four-action channel turns this into a direct test. Every action pays the same amount and writes one of two symbols on the record. Action a_1 always writes the first symbol and a_2 the second; a_3 and a_4 each write either symbol with equal probability. The choice is between a fair coin over $\{a_1, a_2\}$ and a fair coin over $\{a_3, a_4\}$. The two coins randomise identically and induce the same joint distribution of payoff and record. Under the first coin the record identifies the realised action; under the second it reveals nothing. Any criterion that reads only the distribution of consequences is indifferent between the coins; so is a taste for randomisation as such. Traceable agency predicts a strict preference for the first. Because the record affects neither payoffs nor any later decision, the preference cannot be instrumental.[2] The comparison identifies a value of act–consequence dependence; by itself it does not distinguish trace from control or identify the psychological source of that value.[3]

This comparison has not been run. The nearest evidence comes from two experiments. Bartling, Fehr, and Herz elicit each principal's point of indifference between retaining and delegating the right to choose a project and the effort devoted to it. At that point, the delegation lottery has a certainty equivalent 16.7% higher on average than the lottery produced by the principal's own choices, when both are valued outside the delegation setting [5]. This is consistent with a preference for trace: principals are willing to pay to keep the payoff lottery dependent on their own choices. However, the same evidence may instead reflect autonomy. Mistry and Liljeholm, and later Norton and Liljeholm, ask subjects to select a room containing two slot machines and then gamble in that room. Subjects are more likely to choose a high-divergence room over a zero-divergence room when they choose between its machines themselves than when a computer chooses for them, holding current expected payoffs and outcome entropies fixed [30, 31]. With equal weight on the two machines, the divergence between their token distributions is the mutual information between the machine and the token. This also admits a trace interpretation: in self-play, the token is evidence about a voluntarily selected machine. But token values could change after the room was selected, allowing subjects in self-play to adapt their machine choices. Thus trace is confounded with the value of deciding in the first design and with that of adapting in the second. The test proposed above removes both confounds.

Section 2 sets out the domain and the eight axioms. Section 3 states the representation theorem, with a proof sketch, and derives uniqueness, the sign trichotomy, and the independence of the axioms. Section 4 presents a two-treatment test that can reject the model; derives deliberate randomisation, with multinomial logit as the boundary case when channel-row supports are pairwise disjoint; prices garblings of the record; and measures λ from choices. Section 5 reviews the literature, in particular rational inattention, where mutual information is a cost of attention rather than an object of taste, and the axiomatic characterisations of information indices. Section 6 discusses interpretation and scope. The proofs are in the appendices: Appendix A gives one forward proof of Theorem 1, including its constant-payoff characterisation, and Appendix B contains the remaining results.

[1] If the visible-consequence supports of all channel rows are pairwise disjoint, the unique optimal lottery has multinomial-logit weights. Overlap can make independence of irrelevant alternatives fail; see §4.2.

[2] A second treatment completes the design: repeat the choice in a channel that writes a fair record independently of the realised action. Every representation then assigns both flips the same value, whatever u and λ, so a strict preference there rejects the model. Section 4 develops the test and what it excludes.

[3] The information term admits a control interpretation. Read backward: it measures trace. Read forward, it measures how strongly the action controls the distribution of visible consequences. We return to this in Section 6.

2 The model

Fix a finite set O of *payoff-bearing outcomes*, with $|O| \geq 2$. Let A be a nonempty finite set of *actions* and R a nonempty finite set of *explicit visible records*. A joint channel is a stochastic kernel

$$K : A \longrightarrow \Delta(O \times R).$$

For each finite joint channel K, the primitive behavioural datum is a binary relation

$$\succeq_K \quad \text{on} \quad \Delta(A).$$

The interpretation is that q is a lottery over actions used by the agent and $q \succeq_K p$ means that, in the fixed environment K, q is evaluated at least as highly as p. Within that environment, the lottery is the procedure being evaluated and the channel is fixed.[4]

Axioms are imposed on the family $\{\succeq_K\}_K$. The first two axioms are standard, and Axiom (A3) is the usual nondegeneracy condition on material outcomes. Axioms (A4), (A6), and (A7) carry the preference for trace. Axiom (A5) puts the separate within-channel rankings on a common scale, while Axiom (A8) imposes consistency under compounding; neither fixes the sign of the trace term.

(A1) Weak order. For every finite joint channel K, $\succeq_K$ is complete and transitive on $\Delta(A)$.

(A2) Continuity. For each fixed triple of alphabets (A, O, R), the set

$$\{(K, q, p) : q \succeq_K p\} \subseteq \Delta(O \times R)^A \times \Delta(A)^2$$

is closed.

(A3) Material relevance. There are distinct outcomes $o^+, o^- \in O$ and a two-action channel K with a one-point record set in which a^+ delivers o^+ for sure and a^- delivers o^- for sure. For this channel,

$$\delta_{a^+} \succ_K \delta_{a^-}.$$

The agent strictly ranks at least two outcomes when each is delivered for sure and the two actions leave the same record.

(A4) Trace relevance. There is an outcome $o_* \in O$ and a four-action channel K in which every action delivers o_* for sure. Actions a_1 and a_2 leave distinct records r_1 and r_2, respectively, while each of a_3 and a_4 leaves either record with equal probability. For this channel,

$$\frac{1}{2}(\delta_{a_1} + \delta_{a_2}) \succ_K \frac{1}{2}(\delta_{a_3} + \delta_{a_4}).$$

Both are fair lotteries and produce the same distribution of visible consequences. The axiom requires a strict preference for the first, whose record identifies the realised action, over the second, whose record does not.

[4] Why not a single preference over lotteries on $A \times O \times R$? Every joint law is $q \otimes K$ for some pair, so the domain is rich enough. But in any one choice situation the channel is fixed: the agent moves the marginal on A, never the coupling. A preference over all joint laws would rank couplings, and no choice reveals those rankings.

Comparison environments. Some axioms compare lotteries carried by different channels. Each relation $\succeq_K$ ranks lotteries within one channel, so such comparisons need a domain in which two channels appear in a single preference statement. A larger channel that contains both as labelled blocks provides it.

For $K : A \to \Delta(O \times R)$ and $L : B \to \Delta(O \times R_L)$, let $K \sqcup L$ be the block-diagonal joint channel—the *block environment*—with action set $(A \times \{0\}) \cup (B \times \{1\})$, payoff set O, and record set $(R \times \{0\}) \cup (R_L \times \{1\})$: each block operates its own channel, and all cross-block probabilities are zero. For $q \in \Delta(A)$ and $p \in \Delta(B)$, let q^0 and p^1 denote their block-supported copies, defined on the tagged action set by

$$\begin{aligned} q^0(a,0) &:= q(a), \quad & q^0(b,1) &:= 0, \\ p^1(a,0) &:= 0, & p^1(b,1) &:= p(b), \end{aligned} \qquad (a \in A,\ b \in B).$$

Under q^0, the visible consequence is generated by K after the action is drawn according to q; under p^1, it is generated by L after the action is drawn according to p. So a choice between them is also a choice between the two channels, made inside a single environment. Write

$$(q, K) \succeq (p, L) \quad \iff \quad q^0 \succeq_{K \sqcup L} p^1.$$

Each instance of this shorthand is a single comparison inside the two-block environment $K \sqcup L$, which is built from the two channels; $\succ$ and $\sim$ denote the asymmetric and symmetric parts of this shorthand relation. The pair (q, K) has no meaning in isolation.[5] The construction extends to any finite family $\{K_i : A_i \to \Delta(O \times R_i)\}_{i \in I}$: write $\bigsqcup_{i \in I} K_i$ for the analogous block environment and q_i^i for the lottery that assigns $q_i(a)$ to (a, i) and zero to every other block.

(A5) Block-comparison coherence. *Duplication:* for every $K : A \to \Delta(O \times R)$ and all $q, p \in \Delta(A)$,

$$q \succeq_K p \quad \iff \quad (q, K) \succeq (p, K).$$

Irrelevant blocks: for every finite block environment $\bigsqcup_{k \in I} K_k$, every ordered pair of distinct blocks $i, j \in I$, and all $q_i \in \Delta(A_i)$ and $q_j \in \Delta(A_j)$,

$$q_i^i \succeq_{\bigsqcup_{k \in I} K_k} q_j^j \quad \iff \quad (q_i, K_i) \succeq (q_j, K_j).$$

Comparing q and p across two tagged copies of K is just comparing them in K; adding blocks not involved in a comparison leaves it unchanged.

Processing. The next two axioms use two ways of transforming an environment. A *record processor* is a stochastic kernel $T : O \times R \to \Delta(O \times R')$ that garbles the record without changing the payoff: $T(o', r' \mid o, r) = 0$ whenever $o' \neq o$. The garbled channel is simply the matrix product $KT : A \to \Delta(O \times R')$,

$$(KT)(o', r' \mid a) := \sum_{o,r} K(o, r \mid a) T(o', r' \mid o, r).$$

An *action processor* is a stochastic kernel $S : A \to \Delta(B)$ that replaces the action variable A by a processed action variable B. Given $q \in \Delta(A)$, it induces the lottery $qS \in \Delta(B)$,

$$(qS)(b) := \sum_a q(a) S(b \mid a).$$

[5] Formally, $(K \sqcup L)(o, (r, 0) \mid (a, 0)) = K(o, r \mid a)$ and $(K \sqcup L)(o, (r_L, 1) \mid (b, 1)) = L(o, r_L \mid b)$, with every other entry zero. In the binary case, with $A = \{a_1, a_2\}$ and $B = \{b_1, b_2\}$, the block environment has the four tagged actions $(a_1, 0)$, $(a_2, 0)$, $(b_1, 1)$, $(b_2, 1)$; a lottery $q = (q(a_1), q(a_2))$ becomes $q^0 = (q(a_1), q(a_2); 0, 0)$ and $p = (p(b_1), p(b_2))$ becomes $p^1 = (0, 0; p(b_1), p(b_2))$, the semicolon separating the blocks. Then $(q, K) \succeq (p, L)$ abbreviates $q^0 \succeq_{K \sqcup L} p^1$: one ordinary comparison of two lotteries in this one four-action environment.

A *Bayesian pushforward* of K by S at q is any channel $\widehat{K} : B \to \Delta(O \times R)$ satisfying[6]

$$(qS)(b)\,\widehat{K}(o,r \mid b) = \sum_a q(a)S(b \mid a)K(o,r \mid a) \quad \text{for every } (b,o,r). \tag{1}$$

In the processed environment, the alternatives are the noisy names b, while the visible consequences remain (O,R). Choosing qS under $\widehat{K}$ replaces dependence on the original action with dependence on its noisy name, without changing the marginal distribution of visible consequences.

(A6) Record data processing. For every $K : A \to \Delta(O \times R)$, every record processor T, and every $q \in \Delta(A)$,

$$(q,K) \succeq (q,KT).$$

The agent never gains when her record is garbled.

(A7) Action data processing. For every $K : A \to \Delta(O \times R)$, every action processor $S : A \to \Delta(B)$, every $q \in \Delta(A)$, and every Bayesian pushforward $\widehat{K}$ of K by S at q,

$$(q,K) \succeq (qS,\widehat{K}).$$

Nor does she gain when the action variable is processed, holding the distribution of visible consequences fixed. Special cases include swapping action labels, adding or deleting unused actions, splitting an action into identical copies, merging actions, and noisy recoding.

Compounding. For $k \geq 1$, let $P : A \to \Delta(\{1,\ldots,k\})$ be a first-stage record channel and let $(K_1,\ldots,K_k)$ be an ordered list of continuation channels $K_i : A \to \Delta(O \times R)$. Their *compound environment* is

$$(P \triangleright (K_1,\ldots,K_k))(o,(i,r) \mid a) := P(i \mid a)K_i(o,r \mid a), \qquad i = 1,\ldots,k. \tag{2}$$

For expositional purposes, consider the following two-stage description of the compound environment.[7] First, the action a generates only an interim record i through P; no payoff is delivered at this stage. Record i selects the continuation channel K_i, through which the same action generates the payoff o and a further record r. The agent does not act again. The interim record remains visible, so the final record is (i,r). Under q and P, record i is *reached* if $\sum_a q(a)P(i \mid a) > 0$. For every reached record, let $q_i := q(a)P(i \mid a)/\sum_{a'} q(a')P(i \mid a')$ denote the posterior distribution of the action.

(A8) Recordwise sure-thing principle. For every $q \in \Delta(A)$, every binary first-stage record channel $P : A \to \Delta(\{1,2\})$ for which record 1 is reached, and every three continuation channels $K,L,M : A \to \Delta(O \times R)$,

$$(q_1,K) \succeq (q_1,L) \quad \Longleftrightarrow \quad (q, P \triangleright (K,M)) \succeq (q, P \triangleright (L,M)). \tag{3}$$

Fix q and a binary first-stage record channel P, and suppose record 1 is reached, inducing posterior q_1. The two compound environments share the continuation M after record 2 and differ only after record 1, where one continues with K and the other with L. The axiom says that the compounds are ranked exactly as the lottery q_1 in channel K and the same lottery in channel L are ranked in their block environment. The common continuation on the other branch cannot affect the comparison.[8]

[6] If $(qS)(b) = 0$, the equation leaves that row unrestricted; every statement below applies to every consistent completion, making unreached processed actions irrelevant.

[7] The stage structure merely factorises the compound channel, which is a one-shot experiment. Preferences over the timing of uncertainty resolution are therefore abstracted from.

[8] This is a version of Savage's sure-thing principle with the event given by an interim record [35]. Unlike the exogenous event in Savage's formulation, reaching the record can be informative here: $P(1 \mid a)$ may vary with the action. This is why the conditional comparison is evaluated at q_1. If record 1 is never reached, the axiom is silent, as on Savage's null events. When P is an action-independent public coin, $q_1 = q$, and Axiom (A8) reduces to ordinary mixture independence with the coin retained in the record.

3 The representation theorem

For $q \in \Delta(A)$ and $K : A \to \Delta(O \times R)$, let

$$p_{qK}(o,r) := \sum_a q(a)K(o,r \mid a)$$

be the induced distribution of visible consequences, and define

$$I_{qK}(A;O,R) := \sum_{a,o,r} q(a)K(o,r \mid a) \log \frac{K(o,r \mid a)}{p_{qK}(o,r)}. \tag{4}$$

Summands with $q(a)K(o,r \mid a) = 0$ are set to zero, the logarithm has any fixed base greater than one, and expectations below are taken under the joint distribution $q(a)K(o,r \mid a)$.

Theorem 1. *For a family $\{\succeq_K\}_K$ as described above, the following are equivalent.*

(i) *The family satisfies Axioms* (A1)–(A8).

(ii) *There are a nonconstant function $u : O \to \mathbb{R}$ and a number $\lambda > 0$ such that, for every nonempty finite A, R, every $K : A \to \Delta(O \times R)$, and every $q, p \in \Delta(A)$,*

$$q \succeq_K p \quad \Longleftrightarrow \quad V(q,K) \geq V(p,K), \qquad V(q,K) := \mathbb{E}_{qK}[u(O)] + \lambda I_{qK}(A;O,R). \tag{5}$$

Moreover, under these equivalent conditions, block-supported cross-channel comparisons are represented on the same scale: for every finite block environment $\bigsqcup_{k\in I} K_k$, every two distinct blocks $i, j \in I$, and every $q_i \in \Delta(A_i)$ and $q_j \in \Delta(A_j)$,

$$q_i^i \succeq_{\bigsqcup_{k\in I} K_k} q_j^j \quad \Longleftrightarrow \quad V(q_i, K_i) \geq V(q_j, K_j).$$

We call a family satisfying (A1)–(A8) *trace-tempered*, and likewise the criterion that represents it. Its central restriction is that payoff and trace enter separately. The agent may value good outcomes and may value evidence about her action, but she cannot place an extra value on a good payoff being the revealing consequence. Thus, if under the same lottery two channels give the same expected utility from payoffs and induce the same distribution of posteriors over actions, the agent is indifferent between them, regardless of how payoffs and records are paired (Corollary 29, Appendix B.1). This separation is derived, not assumed.

3.1 Proof sketch

That the representation satisfies the axioms is a direct verification. The converse has four steps. It first establishes the common affine machinery, then restricts it to the constant-payoff domain and proves a pure-trace lemma, returns to general payoffs to align the material and trace scales, and finally assembles arbitrary channels branch by branch.

Step 1: reduction to terminal laws. Fix a full-support q. If two channels induce the same joint distribution of the realised payoff and the posterior over actions, each is a record processing of the other. Block coherence and record processing therefore make them indifferent. Action processing makes action relabellings immaterial and, for an arbitrary q, permits null actions to be deleted. Thus (q, K) is identified by its *terminal law*, the joint distribution of the realised payoff and posterior (Lemma 11(a) and Lemma 10(a)–(b)).

Step 2: cardinalisation. Let a public coin select between two channels. Because the coin is independent of the action, the posterior in either branch remains q. The recordwise sure-thing principle therefore gives mixture independence over terminal laws. Together with the weak order and continuity, the mixture-space theorem gives an affine representation on each full-support fibre. The cardinal structure enters here. The rest of the proof aligns and calibrates these fibrewise scales (Lemma 11(b)).

Step 3: payoff and trace. On singleton action sets there is no trace, so the affine representation reduces to expected utility. Material relevance makes its index u nonconstant, while action processing and block coherence make the same payoff scale apply throughout. At the payoff o_* that witnesses trace relevance, the payoff term is constant. The constant-payoff phase of Appendix A shows that the remaining order is represented by mutual information. The finite-branch form of Axiom (A8) supplies the ordinal continuation comparisons; fibrewise affinity and scale coherence convert them into an exact chain rule. Together with action relabelling and support transport, that rule gives Faddeev's grouping property, while Axiom (A2) supplies continuity and block coherence aligns comparisons across priors. Hence the pure-trace order is a scalar multiple of Shannon entropy [20]. Record data processing makes the scalar nonnegative, trace relevance makes it strictly positive, and material calibration determines one coefficient λ relative to u (Lemmas 25 and 26).

Step 4: assembly. The finite-branch sure-thing result makes insertion on a reached branch an order embedding. For a fixed background, the marked terminal law is affine in the inserted continuation; the affine fibre representation therefore makes this embedding affine. Calibration on the constant-o_* face and the mutual-information chain rule identify its coefficient as the branch probability; crossed public mixtures show that this coefficient is independent of the other branches. Begin with a compound that pays o_* after every first-stage record. If record y occurs with probability m_y, changing only the payoff after y from o_* to o changes value by $m_y[u(o) - u(o_*)]$. This baseline compound has value

$$u(o_*) + \lambda I_q(A; Y).$$

Summing across branches gives the representation when the payoff is determined by the first-stage record. For an arbitrary channel, factor it through the interim record (o, r) and a deterministic continuation whose payoff is o. The resulting compound channel and the original channel are record processings of one another and hence are indifferent. The representation follows (Lemmas 27 and 28).

3.2 Uniqueness

A real-valued index W on all finite lottery–channel pairs is a *common-scale representation* if, for every channel $K : A \to \Delta(O \times R)$ and every $q, p \in \Delta(A)$,

$$q \succeq_K p \quad \Longleftrightarrow \quad W(q, K) \geq W(p, K),$$

and, for every finite block environment $\bigsqcup_{k \in I} K_k$, every two distinct blocks $i, j \in I$, and all $q_i \in \Delta(A_i)$ and $q_j \in \Delta(A_j)$,

$$q_i^i \succeq_{\bigsqcup_{k \in I} K_k} q_j^j \quad \Longleftrightarrow \quad W(q_i, K_i) \geq W(q_j, K_j).$$

A common-scale representation W is *branch-additive* if, for every $q \in \Delta(A)$, every $P : A \to \Delta(\{1, \dots, k\})$, and every two ordered lists of continuation channels $(K_1, \dots, K_k)$ and $(L_1, \dots, L_k)$, with $K_i, L_i : A \to \Delta(O \times R)$, writing $m_i := \sum_a q(a) P(i \mid a)$,

$$W(q, P \triangleright (K_1, \dots, K_k)) - W(q, P \triangleright (L_1, \dots, L_k)) = \sum_{i : m_i > 0} m_i [W(q_i, K_i) - W(q_i, L_i)]. \quad (6)$$

In words, a continuation change worth x after record i changes ex-ante value by $m_i x$, and changes on different branches add. Axiom (A8) is ordinal: it says that a common continuation on the other branch cannot affect a conditional comparison. By Lemma 12, this is equivalent—given the other structural axioms—to the requirement that branchwise improvements cannot be reversed by finite compounding. Branch additivity is the corresponding requirement on numerical differences.

Proposition 2 (Uniqueness)**.** *Suppose* (A1)–(A8) *hold, and fix a representation V, with parameters (u, λ), as in Theorem 1. For any real-valued index W on all finite lottery–channel pairs:*

(i) *W is a common-scale representation if and only if $W = \varphi \circ V$ for a single strictly increasing $\varphi : [\underline{u}, \infty) \to \mathbb{R}$, where $\underline{u} := \min_{o \in O} u(o)$.*

(ii) *A common-scale representation W is branch-additive if and only if $W = \alpha V + \beta$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$.*

The proof is in Appendix B.1.

Every trace-tempered representation is a common-scale representation and, by the expected-utility and mutual-information chain rules, branch-additive. Part (ii) therefore implies that if (u, λ) and $(\widetilde{u}, \widetilde{\lambda})$ are two such representations, then

$$\widetilde{u} = \alpha u + \beta, \qquad \widetilde{\lambda} = \alpha \lambda$$

for some $\alpha > 0$ and $\beta \in \mathbb{R}$. Once two non-indifferent payoff outcomes are assigned numerical utilities, λ is unique for the chosen logarithm base. In the constant-payoff phase of Appendix A, multiplying the pure-trace index by a positive constant has no behavioural content. Here the payoff normalisation fixes that scale, and λ measures the trade-off between payoff and trace.

3.3 Sign of the trace weight

Uniqueness concerns scale; a complementary observation concerns sign. Axioms (A4), (A6), and (A7) orient the trace component. Write (A4$^-$), (A6$^-$), and (A7$^-$) for the clauses obtained by reversing their displayed comparisons, and write (A4^0), (A6^0), and (A7^0) for the corresponding indifference clauses.

Corollary 3 (Sign trichotomy)**.** *Suppose Axioms* (A1)–(A3), (A5), *and* (A8) *hold. Then:*

(i) *under* (A4), (A6), *and* (A7), *the family has the representation in Theorem 1, with $\lambda > 0$;*

(ii) *under* (A4$^-$), (A6$^-$), *and* (A7$^-$), *it has the same representation with $\lambda < 0$;*

(iii) *under* (A4^0), (A6^0), *and* (A7^0), *it is represented by expected utility, equivalently by the same formula with $\lambda = 0$.*

The proof is in Appendix B.1. The directions of the three orientation clauses cannot be chosen independently. For example, under the corollary's background axioms and any orientation of Axiom (A7), Axiom (A4) is inconsistent with (A6$^-$). Restrict the two lotteries in Axiom (A4) to their supports and relabel the actions. The second channel is then a total garbling of the first. Axiom (A6$^-$) weakly prefers the second, whereas Axiom (A4) strictly prefers the first.

The negative case is quiet in unconstrained choice over lotteries. With $\lambda < 0$, every payoff-maximising pure action is optimal. More generally, q is optimal in a fixed channel if and only if it is supported on payoff-maximising actions and $I_{qK}(A; O, R) = 0$. To identify $|\lambda|$ from a chosen optimum, one must therefore constrain the lottery or compare channels. Under an imposed fair lottery over two actions, the agent is willing to give up as much as $|\lambda| \log 2$ of expected utility from payoffs to replace full revelation with silence.

3.4 Independence of the axioms

Proposition 4 (Independence of the axioms)**.** *For each $i \in \{1, \ldots, 8\}$ there is a family satisfying the other seven axioms and violating Axiom* (Ai).

The eight families are constructed in Appendix B.2. Three mark the substantive frontier of the system. A conditional-entropy penalty satisfies every axiom except record data processing and values the removal of extraneous record noise. A Gini posterior-reduction criterion satisfies every axiom except action data processing. The satiation criterion $\mathbb{E}[u(O)] + \min\{I(A; O, R), c\}$, for $c > 0$, satisfies every axiom except the recordwise sure-thing principle and ceases to value trace once c units of information are secured.

4 Choice implications

This section works out what an agent with the representation of Theorem 1 does: which tests falsify the model directly (§4.1), when she deliberately randomises (§4.2), what she will pay for changes to the record technology (§4.3), and how that price can be measured (§4.4). Throughout, after fixing the affine normalisation of u and the logarithm base, λ is measured in payoff units per unit of information. Proofs for this section, together with full statements of the supporting results that the text presents in summary form, are collected in Appendix B.3.

4.1 The trace test

The criterion V reads the joint distribution of action and visible consequence, rather than the consequence distribution or action lottery alone. The following two-treatment design separates these objects. In both treatments, the lottery must be knowingly chosen and implemented by the subject, since the primitive ranks her own randomisations.[9]

Proposition 5 (The trace test)**.** *Four actions pay the same outcome o_0. Two of them leave distinct marks: a_1 always produces record r_1, and a_2 always produces the distinct record r_2. The other two are indistinguishable: a_3 and a_4 each produce r_1 or r_2 with equal probability. Call this channel K, and let q' be the fair coin over $\{a_1, a_2\}$ and q'' the fair coin over $\{a_3, a_4\}$.*

(i) *The two coins induce the same distribution of visible consequences, $p_{q'K} = p_{q''K}$; yet*

$$V(q', K) - V(q'', K) = \lambda \log 2 > 0, \qquad \text{so} \qquad q' \succ_K q''.$$

(ii) *Let K° be the channel in which every action pays o_0 and produces r_1 or r_2 with equal probability. Then $q' \sim_{K^\circ} q''$.*

This pattern detects a taste for action–consequence dependence; by itself it does not identify mutual information among all dependence indices.

Part (i) rules out any criterion that depends only on the distribution of visible consequences. The two coins induce the same distribution over payoff–record pairs, so any such criterion is indifferent between them. However, a theory that also reads the lottery can match it. Existing theories of deliberate randomisation are of this kind: in the hedging model of Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella [14], and in the preference for randomisation documented by Agranov and Ortoleva [2], the value of mixing depends only on the induced distribution and the lottery. Call an index W *channel-blind* if it reads only these two objects: there is a function F, fixed across channels, such that

$$W(q, K) = F(p_{qK}, q).$$

Part (ii) rules these out. Nothing a channel-blind index reads changes between the treatments: in all four lottery–channel combinations the distribution of visible consequences is uniform on $\{(o_0, r_1), (o_0, r_2)\}$, and the lotteries are the same two lotteries. A channel-blind index

[9] A direct value for information about which action was realised is observationally equivalent on this domain. It can be separated by assigning the same lottery externally: the information structure remains, but the subject no longer chooses the lottery.

therefore separates the coins in both treatments or in neither, so none reproduces the pattern of Proposition 5: strict preference under K and indifference under K°.[10]

The first treatment also identifies the orientation of taste: it selects the revealing coin when $\lambda > 0$, the noisy coin when $\lambda < 0$, and indifference when $\lambda = 0$. A strict preference between the coins in the second treatment rejects all three cases of Corollary 3, and with them the model.

4.2 Deliberate randomisation

Consider a decision maker free to choose lotteries among two actions, and let a pay weakly more. Playing a for sure maximises payoff and leaves no trace; mixing in b costs payoff but gives the record something to reveal. Write $K_a := K(\cdot \mid a)$ for the distribution of visible consequences under action a, $\bar{u}(a) := \mathbb{E}_{K_a}[u(O)]$ for its expected utility, and $\Delta := \bar{u}(a) - \bar{u}(b) \geq 0$ for the payoff gap. Each unit of probability moved from a to b costs Δ. What it buys is evidence. A consequence testifies about the act that produced it through a likelihood ratio: how much more probable it is given the realised act than given only the mixture she plays. With all probability on a, consequences are distributed exactly as K_a: every draw has ratio one and testifies to nothing. Now move a marginal unit of probability to b. The mixture's consequences are still, to first order, distributed as K_a; but when b realises, its consequence is drawn from K_b, so the draw carries $\log\big(K_b(o,r)/K_a(o,r)\big)$. Its expected value over b's draws is, by definition, the divergence $D_{\mathrm{KL}}(K_b \,\|\, K_a)$.[11] Because information is concave in the lottery, whether to mix at all is settled at the margin (Proposition 30, Appendix B.3.2):

$$\delta_a \text{ is optimal} \quad \iff \quad \Delta \geq \lambda\, D_{\mathrm{KL}}(K_b \,\|\, K_a).$$

When the payoff gap is below $\lambda D_{\mathrm{KL}}(K_b \,\|\, K_a)$, the optimal lottery is unique and mixes the two actions, with weakly more probability on b as λ rises—strictly when $\Delta > 0$—and less as Δ rises.

With more than two actions, each action is compared with the distribution of visible consequences generated by the lottery as a whole, p_{qK}. An action is credited with its expected utility and with how far its visible consequences depart from those generated on average. At an optimum this total is the same for every action in use; an unused action cannot offer more.

Proposition 6 (Optimality condition)**.** *Let $\lambda > 0$. A lottery q^* is optimal if and only if there is a constant c such that*

$$\bar{u}(a) + \lambda D_{\mathrm{KL}}(K_a \,\|\, p_{q^*K}) = c \quad \textit{for } a \in \operatorname{supp}(q^*), \qquad \leq c \quad \textit{for } a \notin \operatorname{supp}(q^*).$$

The condition has a sharp implication: every visible consequence that can arise under some action occurs with positive probability under any optimal lottery (Proposition 31, Appendix B.3.2). Hence, if each action can produce some consequence that no other action can produce, every optimal lottery uses every action. At the other extreme, suppose every action produces the same distribution of payoffs and records. The consequence then reveals nothing about the action, and all actions have the same expected utility. Hence the agent is indifferent among all lotteries.

Corollary 7 (Multinomial logit)**.** *If the row supports are pairwise disjoint, the visible consequence identifies the action, and $I_{qK}(A; O, R) = H(q)$. The unique optimum is then multinomial logit,*

$$q^*(a) = \frac{\exp\big(\bar{u}(a)/\lambda\big)}{\sum_{a'} \exp\big(\bar{u}(a')/\lambda\big)},$$

with $1/\lambda$ as the payoff sensitivity when information is measured in natural logarithms.[12]

[10] If $W(q,K) = F(p_{qK}, q)$ is channel-blind and $F(p,\cdot)$ is invariant under permutations of action labels—an entropy bonus, say—the first treatment is enough: $p_{q'K} = p_{q''K}$ and q'' is a permutation of q', so $W(q',K) = W(q'',K)$. The second treatment is needed for label-sensitive indices.

[11] For distributions P and Q, $D_{\mathrm{KL}}(P \,\|\, Q) := \sum_{o,r} P(o,r) \log\big(P(o,r)/Q(o,r)\big)$, with zero-mass summands set to zero. The divergence is $+\infty$ if P assigns positive probability to a consequence and Q assigns zero.

[12] The same formula arises in rational inattention, where, in contrast, q is fixed and K is chosen [29].

Debreu's red-bus/blue-bus conclusion depends on what is recorded. If blue is an exact copy of red, only their combined probability is pinned down. Any split of the red bus passengers between the two buses is optimal (Proposition 32, Appendix B.3.2). If colour is recorded, however, red and blue are distinct traces. With equal expected utilities, the corollary assigns probability 1/3 to each alternative. The model is invariant to duplicate names, not to distinct recorded traces.

With overlapping supports, the odds between two actions can depend on the rest of the menu. The optimality condition sets each action's probability against the average consequence distribution, and an added action shifts that average toward its own consequences, reducing the measured distinctiveness of the actions it resembles more than of the others; the odds then move. Pairwise-disjoint supports therefore imply independence of irrelevant alternatives, while overlap permits, but does not require, its failure: in a three-action example, adding the third action moves the odds between the first two from 1 to 3/4 (Remark 33, Appendix B.3.2).

4.3 The value of being on record

The representation prices changes to the record in payoff units. Suppose the record is garbled while the payoffs remain untouched. What does the agent lose?

Proposition 8 (Cost of a record garbling)**.** *Let $K : A \to \Delta(O \times R)$, let T be a record processor, and write $L := KT$. Fix $q \in \Delta(A)$ and let (A, O, R, R') carry the joint law $q(a)K(o, r \mid a)T(o, r' \mid o, r)$. Then*

$$V(q, K) - V(q, L) = \lambda\, I(A; R \mid O, R') \geq 0,$$

with equality if and only if, after (O, R') has been observed, the original record R contains no further information about the action.

Rewriting the record costs λ times the information about the action that the original record carried beyond the rewritten consequence (O, R'). A garbling is costless exactly when, conditional on (O, R'), it pools records that induce the same posterior about the action.[13] A case in point: let a coin, drawn independently of the action, decide which of two channels operates. Write Y for its outcome and R_0 for the remaining record, so the full record is (Y, R_0). Erasing Y is a garbling, and it costs $\lambda\, I(A; Y \mid O, R_0)$ (Corollary 34, Appendix B.3.3). The draw says nothing about the action by itself, but knowing which channel operated can sharpen what the consequences say.

Garbling makes a channel uniformly worse; does uniformly worse mean garbled? Suppose two channels K and L generate the same payoff distribution after every action, and every lottery is worth at least as much under K as under L. Garbling would explain this ranking. But the ranking proves less: it says only that K carries at least as much mutual information at every lottery, and two constant-payoff binary channels can satisfy that at every prior although no record processor maps one into the other (Proposition 35, Appendix B.3.3).

4.4 Measuring the trace weight

Under the stated normalization, the coefficient λ is the price of one unit of information in expected-utility units. Any indifference between two environments with different amounts of trace reveals it: writing $E_{qK} := \mathbb{E}_{qK}[u(O)]$, if $(q, K) \sim (p, L)$ and $I_{qK} \neq I_{pL}$, the representation rearranges to

$$\frac{E_{pL} - E_{qK}}{I_{qK} - I_{pL}} = \lambda,$$

and λ is the only number consistent with every such indifference (Proposition 36, Appendix B.3.4).

[13]The same calculation applies to added information. If K^{+S} appends a signal S drawn conditional on (a, o, r), then $V(q, K^{+S}) - V(q, K) = \lambda\, I(A; S \mid O, R)$. Under the deniability case of Corollary 3, the sign is reversed.

Plot expected payoff utility against information: indifference curves are parallel lines of slope $-\lambda$. The slope does not depend on the prior, the payoff level, the action alphabet, or the amount of information already present. This constancy is the observable restriction that Axiom (A8) adds to the rest of the system. A taste for trace that satiates satisfies every axiom except (A8) (Proposition 4) and has a slope that falls to zero beyond the satiation point.

The identity presupposes the utility function; the design in the next proposition does not, because its only payoffs are the two normalised outcomes and a baseline that cancels. At a constant payoff, revelation of the realised action is worth $\lambda H(q)$ more than silence, so compensating the silent side until the agent is indifferent measures her willingness to pay for trace.

Proposition 9 (Eliciting the trace weight)**.** *Normalise $u(o^+) = 1$ and $u(o^-) = 0$. Fix $q \in \Delta(A)$ with $H(q) > 0$, a constant payoff, and two channels, one whose record reveals the realised action and one whose record is silent. Dilute both by the same public side branch, reached with probability $\varepsilon \in (0,1)$: on the revealing side the branch pays o^- for sure; on the silent side it pays o^+ with probability β and o^- otherwise. If $\lambda H(q) \leq \varepsilon/(1-\varepsilon)$, a unique $\beta \in [0,1]$ produces indifference, and*

$$\lambda = \frac{\varepsilon}{1-\varepsilon}\,\frac{\beta}{H(q)};$$

otherwise the revealing side is strictly preferred at every β.

The common baseline payoff's utility cancels, so it need not be known; the information gap $H(q)$ is set by the design, so no information quantity need be estimated. The proof, the general compensation identity, and the experimental controls are in Appendix B.3.4 (Proposition 37 and Remark 38).

The design also tests the model. Run it twice with the same ε, at $q = (1/2, 1/2)$ and at $q = (1/4, 3/4)$. Two runs must return the same λ. Hence, the two indifference reports must satisfy

$$\frac{\beta_1}{\beta_2} = \frac{H(1/2,1/2)}{H(1/4,3/4)} = \frac{\log 2}{\frac{1}{4}\log 4 + \frac{3}{4}\log\frac{4}{3}} \approx 1.23.$$

Any other ratio is not consistent with the model. This number also separates Shannon entropy from the Gini-reduction rival $1 - \sum_a q(a)^2$, which instead commits to $4/3$.

5 Related literature

Grant, Kajii, and Polak [24] characterise an intrinsic preference for information independent of its use in planning; Caplin and Leahy [13] model anticipatory feelings generated by beliefs, while Ely, Frankel, and Kamenica [19] study preferences over paths of beliefs and the resulting value of suspense and surprise. These papers show that information need not be instrumental, but the relevant beliefs concern uncertainty faced by the agent, whereas the posterior here concerns her realised action. The axioms also restrict its value to expected entropy reduction and fix its rate of substitution with material utility.

Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella [14] study deliberate randomisation and characterise a model in which mixing hedges uncertainty about how to evaluate lotteries; Agranov and Ortoleva [2] document stochastic choice under announced repetition and evidence that much of it is deliberate. Trace-tempered choice can also make a nondegenerate lottery optimal. The test in Proposition 5, however, holds fixed the action lotteries and their distributions of visible consequences while changing whether the consequence identifies the realised action; its second treatment removes that dependence. Together the two treatments rule out every channel-blind criterion that reads only the lottery and its consequence distribution. Randomisation is an implication of traceable agency, not its primitive motive.

In signaling models following Spence [38], actions convey information about hidden characteristics. Glazer and Konrad [23] study charitable giving as a signal of wealth, and Andreoni and Bernheim [3] study social image and audience effects. Bénabou and Tirole [8] allow concerns for social reputation or self-respect. Their earlier paper [7] studies self-serving beliefs about ability, while Bernheim [9] shows how status concerns can induce agents with heterogeneous preferences to choose the same action, making behaviour less informative about predispositions—an instrumental counterpart to the deniability case in Corollary 3. In these models, inference matters through esteem, sanction, reputation, or self-regard. Here no such response enters utility. Beliefs serve only as the statistical yardstick for how clearly the consequence identifies the act.

The proof also relates to axiomatic characterisations of entropy. Shannon [36] characterised entropy from continuity, monotonicity on uniform distributions, and a grouping recursion; Faddeev [20] reduced these to symmetry, continuity of the binary entropy, and the recursion itself. In the present proof, Axiom (A6) supplies nonnegativity and Axiom (A4) rules out the zero multiple. The proof uses the strong-additivity form of Faddeev's theorem given by Baez, Fritz, and Leinster [6]; other formulations and extensions include Khinchin [26], Rényi [33], Aczél and Daróczy [1], and Csiszár [16]. In those characterisations, a numerical information functional is primitive and structural conditions are imposed directly on it; here the primitive is an ordinal family of weak orders over action lotteries. The information chain rule and Faddeev's recursion are consequences of behavioural axioms, not postulates on a given function. The theorem also calibrates the information term against material utility. The result is not another characterisation of entropy; it derives entropy reduction as the representation of a preference over trace.

Blackwell [10] compares experiments by their value in all downstream decision problems; Gilboa and Lehrer [22] characterise which functions on partitions can arise as values of information in Bayesian decision problems, and Azrieli and Lehrer [4] extend the analysis to stochastic information structures. Cabrales, Gossner, and Serrano [11] obtain entropy reduction from the decisions of ruin-averse investors. For Blackwell's decision-maker, information is an input: she can use a finer signal or ignore it, so more information cannot hurt. Here information is an output: the record is evidence about the agent, and no decision follows it. Uniform preference between two record technologies identifies the more-capable order, not Blackwell garbling (Proposition 35, Appendix B.3.3). Whether a finer record is desirable is therefore a matter of taste—positive for the trace-seeker, negative under the deniability dual of Corollary 3, and zero under expected utility. Being better informed is a theorem; being better known is a preference.

The closest economic comparison is with rational inattention and information costs. Sims [37] imposes a finite Shannon-capacity constraint on information flow, while Matějka and McKay [29] show that a cost proportional to mutual information yields a generalized multinomial-logit rule. Rational inattention fixes a prior over states and allows the agent to choose how much to learn before acting, whereas here the channel from actions to consequences is fixed and the agent chooses a distribution over actions. For fixed K, the same optimisation is familiar from information theory: q is the channel's input distribution, visible consequences are its output, and an action's shortfall from the highest expected payoff acts as its input cost weighted by $1/\lambda$. When visible consequences reveal the realised action, the resulting choice probabilities coincide with the rational-inattention formula: both are multinomial logit.

Caplin, Dean, and Leahy [12] characterise Shannon rational inattention through Invariance under Compression and develop posterior-separable generalisations; Denti [17] gives testable conditions under which state-dependent stochastic choice is rationalised by posterior-separable information costs. Frankel and Kamenica [21] characterise measures of information that arise as the value of news in a decision problem and show that the corresponding expected reduction in uncertainty equals expected information. Here the primitive is instead an ordinal family of preferences over action lotteries, indexed by action-to-consequence channels, and the axioms derive a common scale for payoff and trace.

Pomatto, Strack, and Tamuz [32] take a cardinal, prior-independent cost functional on Blackwell experiments and show that monotonicity, continuity, product additivity, and constant marginal cost characterise weighted sums of Kullback–Leibler divergences between state-conditional signal distributions. Their costs and the trace value studied here are not nested. The trace component depends on the prior and is not a prior-free additive cost of an experiment. Full revelation has finite trace value $H(q)$; under every nonzero cost in their class it is excluded from the finite-cost domain and would have infinite cost. Record and action data processing are behavioural assumptions about whether the agent values evidence of her action; reversing Axioms (A4), (A6), and (A7) together yields a preference for deniability, not an information-acquisition cost. The information-cost literature prices learning about the world; the present paper prices being learned about.

6 Interpretation and scope

Evidence, not authorship or control. The trace term prices evidence of agency: not authorship, and not control in its everyday sense. A pure action fixes the act in advance, although its consequence may remain stochastic; it has zero trace value even when the consequence names the act perfectly, because an act settled in advance leaves no uncertainty for the consequence to resolve. A coin the agent flips herself can do better. The model does not say that surrendering a choice to chance creates more agency; it says that an unsettled act can leave more evidence than a predetermined one. The lottery must be hers—knowingly chosen and implemented, as in the test of §4.1. This is the Underground Man's problem: the proof he wants cannot come from an act fixed in advance. A reading of control or authorship survives if what is valued is its display rather than its possession; but the display of which act was taken is exactly what the trace term measures, so the surviving reading is trace under another name.

Relation to empowerment. The channel maps actions into distributions over consequences, and the information term can be read from either end. Read backward, from consequence to action, it measures trace: how much the visible consequence reveals about the realised action. Read forward, from action to consequence, it measures how strongly the action shapes the distribution of visible consequences. Mutual information is symmetric, so the two readings share one index. No choice among lotteries separates them.

The forward reading has an exact counterpart. Suppose $\mathbb{E}[u(O) \mid a] = \bar{u}$ for every action. Since $\lambda > 0$, maximising V over q gives

$$\max_{q\in\Delta(A)} V(q, K) = \bar{u} + \lambda C(K), \qquad C(K) := \max_{q\in\Delta(A)} I_{qK}(A; O, R),$$

the Shannon capacity of the channel. When the visible consequence is read as the agent's sensor state, $C(K)$ is the one-step empowerment index of Klyubin, Polani, and Nehaniv [27]; see also [34]. Under this supposition, the trace-tempered agent solves exactly the problem that defines one-step empowerment: her optimal lotteries are the capacity-achieving action distributions, and the size of λ plays no role in the ranking—the price is visible only against payoffs. Empowerment summarises the channel by the optimised value $C(K)$; the representation recovers the ranking beneath it, $q \mapsto I_{qK}(A; O, R)$, from preferences and prices it against material utility. Empowerment is the optimised information index when expected payoffs are equal, and its contribution to indirect utility is $\lambda C(K)$; once payoffs differ across actions, the ranking also reflects material value that $C(K)$ does not encode.

What the axioms preclude. A principal substantive exclusion is selective legibility. Mutual information responds to an act's probability and to the distinctiveness of its consequences, not to what the act is. An agent who wants her generous acts recognised and her mean acts hidden may therefore prefer a channel to its garbling in one case and the garbling to its source in another. Such sign-reversing preferences violate the global data-processing requirement in Axiom (A6) and fall outside all three cases of the sign trichotomy. Such preferences are not irrational; they are tastes for something other than trace. Wanting particular acts known points to an audience, outside the agent or within her, for whom what is revealed matters. The taste axiomatised here does not depend on one: it prices how much the consequence reveals, not what. A model of selective legibility would have to supply the audience, making the value of evidence depend on its content and not only on the amount of uncertainty it removes.

The payoff–record distinction. Although the information term treats O and R jointly, the axioms cannot. Axiom (A6) permits the record to be garbled while the payoff is held fixed; this separates a change in evidence from a change in material consequences. Axiom (A8) permits records to arrive in stages while the payoff is delivered once. Its recordwise sure-thing restriction makes branchwise comparisons commensurable and supplies the cardinal structure used to place payoff utility and information on a common scale. Thus either O or R may reveal the action, but only O enters material utility. The record is also assumed to have no utility beyond what it reveals: a record that is itself pleasant or shameful belongs in O, or the domain must be enlarged. When that hedonic content is measurable from (O, R), recoding of this kind moves content into payoffs and leaves the trace term alone: the display of a shameful act is penalised through u, yet what it reveals is still priced, content-neutrally, through λ. Shame is a payoff; it cannot be counted again as negative trace. Once u is normalised, willingness to sacrifice payoff utility for additional mutual information identifies λ.

One shot, with no audience. The model describes one episode. It specifies neither an audience nor any response to the beliefs induced by the record; the posterior is a statistic of the implemented lottery and the channel, not anyone's belief. The stages in Axiom (A8) factor a single one-shot interaction: no action follows the record, and no relationship extends beyond the episode. Because audience access is not a primitive, the representation predicts no change when an audience is added while the action lottery and channel are held fixed. If choices change, the model is missing the value of being observed by others.

Proofs

A Proof of the representation theorem

This appendix gives one forward proof of Theorem 1. Its dependency spine is simple. We first derive the affine and behavioural tools used throughout. We then restrict attention to constant payoffs, align the comparison scales across priors and support faces, and identify the resulting trace value as mutual information. Returning to the full payoff–record domain, we put material utility and trace on one scale, prove the exact reached-branch identity, and assemble an arbitrary channel by a finite telescope and sequentialisation.

The constant-payoff phase deliberately precedes material calibration. In particular, Proposition 15 retains its weaker assumptions and does not use material relevance. Given marked-law sufficiency, cardinalisation occurs only once: Axioms (A1), (A2), and the public-mixture case of (A8) yield an affine representative on each full-support fibre. Axioms (A5)–(A7) supply the transport and cross-environment comparisons; the product and branch gauges, using the finite-branch form of (A8), align those fibrewise scales and produce the exact chain rule. Faddeev's entropy theorem then completes the constant-payoff identification.

A.1 Shared affine preliminaries

We first record the support, relabelling, and replacement facts that let us move between comparison environments. They require no continuity or compounding axiom.

Recall that $(q, K) \succeq (p, L)$ denotes the paper's canonical two-block comparison. For channels E, G at a common prior q, write $E \succeq_q G$ for $(q, E) \succeq (q, G)$. A *finite experiment at q* is a finite payoff–record channel paired with that prior; its marked terminal law is denoted by η_E.

Lemma 10 (Blocks, supports, and dummies)**.** *Suppose Axioms* (A1) *and* (A5)–(A7) *hold.*

(a) *After identifying alternatives related by invertible action or record relabellings, the derived comparison of pairs is complete and transitive. If either member of a pair comparison is replaced by an indifferent alternative, the comparison is unchanged.*

(b) *For every finite experiment (q, K), if $S = \operatorname{supp} q$, then (q, K) is indifferent to the restriction of (q, K) to S.*

(c) *For every finite experiment (q, K), every nonempty finite B, and every $s \in \Delta(B)$, if $K^{\uparrow}(o, r \mid a, b) := K(o, r \mid a)$, then*

$$(q \otimes s, K^{\uparrow}) \sim (q, K). \tag{7}$$

Proof. Place any finite collection of alternatives in one finite block environment. Axiom (A5) removes unused blocks, and Axiom (A1) supplies completeness and transitivity there. Invertible action and record relabellings are neutral by applying Axioms (A6) and (A7) in both directions. In particular, swapping both tags identifies the two orientations of a two-block comparison. This proves part (a); a three-block environment gives transitivity and a four-block environment gives replacement.

For part (b), process an action in S by its inclusion in A. In the other direction, output $a \in S$ when the realised action is a, and choose an arbitrary element of S off the support. Both processors satisfy (1); rows attached to null processed actions may be completed arbitrarily. Axiom (A7) gives the two comparisons.

For part (c), first project (a, b) to a. Conversely, given a, draw b independently from s and output (a, b). Again the two joint-law identities in (1) are exact, so Axiom (A7) applies in both directions. Part (a) completes both arguments. $\square$

Fix for the moment a full-support prior $q \in \Delta(A)$. For a joint channel K, write

$$p_K(o,r) := \sum_a q(a)K(o,r \mid a), \qquad \pi^K_{or}(a) := \frac{q(a)K(o,r \mid a)}{p_K(o,r)}$$

when $p_K(o,r) > 0$, and define its *marked terminal law*

$$\eta_K := \sum_{(o,r):p_K(o,r)>0} p_K(o,r)\, \delta_{(o,\pi^K_{or})}. \tag{8}$$

It records the terminal payoff and the posterior about the realised action. Write $I_q(E)$ for the mutual information between the action and the complete visible output of E under q. When the output is a first-stage record Y, we also write this quantity as $I_q(A;Y)$.

Lemma 11 (Terminal laws and affine fibres)**.** *Under Axioms* (A1)*,* (A2)*, and* (A5)–(A8)*, the following statements hold.*

(a) *If q has full support and K, L are finite experiments at q with $\eta_K = \eta_L$, then $(q,K) \sim (q,L)$.*

(b) *At every full-support q, the quotient of finite experiments by their marked terminal laws has an affine representation W_q.*

Proof. We give the finite Blackwell argument behind part (a). For a payoff o and posterior π, let

$$M_K(o,\pi) := \sum_{\substack{r:\, p_K(o,r)>0 \\ \pi^K_{or}=\pi}} p_K(o,r).$$

Equality of marked laws gives $M_K(o,\pi) = M_L(o,\pi) =: M(o,\pi)$. For $p_K(o,r) > 0$, draw a record s of L according to

$$T(s \mid o,r) := \mathbf{1}_{\{p_L(o,s)>0,\, \pi^L_{os}=\pi^K_{or}\}} \frac{p_L(o,s)}{M(o,\pi^K_{or})}. \tag{9}$$

Choose an arbitrary probability row when $p_K(o,r) = 0$. This is a payoff-preserving record processor. Bayes' formula gives

$$q(a)K(o,r \mid a) = p_K(o,r)\pi^K_{or}(a).$$

For a reached target record s, with posterior $\pi = \pi^L_{os}$, it follows that

$$\sum_r K(o,r \mid a)T(s \mid o,r) = \frac{p_L(o,s)}{M(o,\pi)} \frac{\pi(a)}{q(a)} M(o,\pi) = L(o,s \mid a).$$

If s is unreached, full support makes both sides zero. Thus postprocessing K by T gives L. The symmetric construction gives a processor from L to K. Axiom (A6) and Lemma 10(a) give indifference.

Boundary priors are handled later by first restricting to their positive support using Lemma 10(b).

For part (b), take the quotient of actual finite marked experiments by equality of (8). If E and G are two such experiments, their public mixture first draws a coin independent of the action, runs the selected experiment, and retains the coin in the record. Its marked law is

$$\eta_{tE\oplus(1-t)G} = t\eta_E + (1-t)\eta_G. \tag{10}$$

Axiom (A8) gives, for $0 < t < 1$ and any third experiment L,

$$E \succeq_q L \quad \Longleftrightarrow \quad tE \oplus (1-t)G \succeq_q tL \oplus (1-t)G. \tag{11}$$

This is the action-independent special case of the axiom: the public coin is the binary first-stage record and G is the common continuation. If record alphabets differ, put them in a common disjoint union before invoking the axiom; part (a) identifies the result with (10).

It remains only to check the Archimedean step. A closed segment between two experiments, and any fixed comparator, can be realised on one fixed union of their finite record alphabets. Its channel entries vary continuously with $t \in [0,1]$. Axiom (A2) therefore makes both contour sets along the segment closed. Completeness says that they cover $[0,1]$; when a target lies between the two endpoints, the sets contain opposite endpoints and connectedness makes them intersect. Thus every such target is indifferent to a point of the segment. The mixture-space theorem now supplies an affine representative W_q. This quotient is closed under every mixture and embedding used below. ☐

A.2 From the axioms to their working forms

The axioms were stated in their weakest forms: a binary sure-thing principle and fixed-channel relevance comparisons. The proof uses their working forms: finite-branch continuation and environment-comparison relevance. This subsection records the two equivalences and makes explicit where the other axioms enter them. Only the statements are used below.

Lemma 12 (Finite-branch extension)**.** *Suppose Axioms* (A1) *and* (A5)–(A7) *hold. Then Axiom* (A8) *is equivalent to the following property. Given $k \geq 1$, a first-stage record channel $P : A \to \Delta(\{1,\dots,k\})$, a prior $q \in \Delta(A)$, and two ordered lists of continuation channels $(K_1,\dots,K_k)$ and $(L_1,\dots,L_k)$, with $K_i, L_i : A \to \Delta(O \times R)$, define for every reached i*

$$m_i := \sum_a q(a)P(i \mid a), \qquad q_i(a) := \frac{q(a)P(i \mid a)}{m_i}.$$

If every reached i satisfies

$$(q_i, K_i) \succeq (q_i, L_i),$$

then

$$(q, P \triangleright (K_1,\dots,K_k)) \succeq (q, P \triangleright (L_1,\dots,L_k)), \tag{12}$$

strictly if one reached branch comparison is strict.

Proof. Assume first Axiom (A8). Put

$$m_i := \sum_a q(a)P(i \mid a), \qquad I^+ := \{i \in \{1,\dots,k\} : m_i > 0\}.$$

If $m_i = 0$, then $P(i \mid a) = 0$ for every $a \in \operatorname{supp} q$. Consequently, changing the continuation after such a record changes the compound channel only on action rows outside $\operatorname{supp} q$. Apply Axiom (A7) in both directions using the identity action processor: its Bayesian pushforwards agree on supported rows and allow arbitrary completions elsewhere. Continuations following unreached records are therefore immaterial. We may hold them fixed and enumerate the reached records as $I^+ = \{i_1,\dots,i_n\}$.

If $n = 1$, the first-stage record is constant on $\operatorname{supp} q$. Deleting that constant label and adjoining it again are payoff-preserving record processings in both directions on every reached row; the off-support rows are immaterial by the preceding paragraph. Axiom (A6) therefore identifies the compound comparison with the sole branch comparison. Suppose henceforth that $n \geq 2$.

For $j = 0,\dots,n$, define hybrid continuation channels by

$$H_i^j := \begin{cases} K_i, & i \in \{i_1,\dots,i_j\}, \\ L_i, & i \in I^+ \setminus \{i_1,\dots,i_j\}, \end{cases}$$

and use a common arbitrary continuation on $\{1,\dots,k\}\setminus I^+$. Write $E_j := P \triangleright (H_1^j,\dots,H_k^j)$. Fix $j \ge 1$ and collapse the first-stage record to the binary channel

$$B_j(1 \mid a) := P(i_j \mid a), \qquad B_j(2 \mid a) := \sum_{i \ne i_j} P(i \mid a).$$

Both binary records are reached. Put all binary continuations on the common record set $\{1,\dots,k\} \times R$. Pad the two continuations that may differ by

$$\widetilde{K}_j(o,(i,r) \mid a) := \mathbf{1}_{\{i=i_j\}} K_{i_j}(o,r \mid a), \qquad \widetilde{L}_j(o,(i,r) \mid a) := \mathbf{1}_{\{i=i_j\}} L_{i_j}(o,r \mid a).$$

On record 2, collect all unchanged branches in the continuation channel

$$M_j(o,(i,r) \mid a) := \frac{P(i \mid a)}{B_j(2 \mid a)} H_i^j(o,r \mid a), \qquad i \ne i_j, \tag{13}$$

with zero probability on records (i_j,r), whenever $B_j(2 \mid a) > 0$, and complete the other rows arbitrarily. Since $H_i^j = H_i^{j-1}$ for $i \ne i_j$, this continuation is common to the two comparisons. Binary record 1 has probability m_{i_j} and carries the conditional action lottery q_{i_j}; record 2 carries the complementary conditional lottery, and their mixture is q. On the active record support, deleting or adjoining the constant label i_j gives payoff-preserving record processings in both directions. Axiom (A6) therefore identifies the comparison of $\widetilde{K}_j$ and $\widetilde{L}_j$ with that of K_{i_j} and L_{i_j}. Hence Axiom (A8) gives

$$(q, B_j \triangleright (\widetilde{K}_j, M_j)) \succeq (q, B_j \triangleright (\widetilde{L}_j, M_j)),$$

strictly whenever the comparison at q_{i_j} is strict.

On the active record support, the nested records are in bijection with the flat records through

$$(1,(i_j,r)) \longmapsto (i_j,r), \qquad (2,(i,r)) \longmapsto (i,r) \quad (i \ne i_j).$$

Extend these maps arbitrarily over records of zero probability. Together with the preceding removal of off-support rows, they give payoff-preserving record processings in both directions. Axiom (A6) therefore makes the two binary compounds indifferent to E_j and E_{j-1}, respectively. Thus $(q,E_j) \succeq (q,E_{j-1})$, strictly at any strict replacement. Place the finite collection $E_0,\dots,E_n$ in one block environment. Axiom (A5) transports each pairwise comparison to that environment, and transitivity in Axiom (A1) yields $(q,E_n) \succeq (q,E_0)$, strictly if at least one step is strict. Restoring the immaterial null branches proves (12).

Conversely, assume the displayed finite-branch property, including its strict clause. In the binary environment of Axiom (A8), reflexivity and duplication coherence make the common continuation weakly comparable to itself, so the finite-branch property gives the forward implication in (3). If the aggregate comparison in that equation held while $(q_1,K) \not\succeq (q_1,L)$, completeness would give $(q_1,L) \succ (q_1,K)$. Applying the strict finite-branch property to the reversed lists would give the strict reverse aggregate comparison, a contradiction. This proves the converse implication and hence Axiom (A8). $\square$

Corollary 13 (One-branch insertion)**.** *Suppose Axioms* (A1) *and* (A5)–(A8) *hold. Fix $q \in \Delta(A)$, a first-stage record channel $P : A \to \Delta(\{1,\dots,k\})$, a reached record i^* with posterior q_{i^*}, and continuation channels $K_1,\dots,K_k,L_1,\dots,L_k : A \to \Delta(O \times R)$ such that $K_i = L_i$ for every $i \ne i^*$. Then*

$$(q_{i^*}, K_{i^*}) \succeq (q_{i^*}, L_{i^*}) \iff (q, P \triangleright (K_1,\dots,K_k)) \succeq (q, P \triangleright (L_1,\dots,L_k)),$$

and the same equivalence holds for strict preference.

Proof. The forward implication follows from Lemma 12, using duplication coherence on the common branches. If the aggregate comparison held but the branch comparison failed, completeness would make the reversed branch comparison strict; the strict part of the finite-branch extension would then give a strict reverse aggregate comparison. The strict equivalence follows by applying the weak equivalence in both directions. □

In the appendices, references to Axiom (A8) with more than two first-stage records mean its finite-branch extension in Lemma 12; when only one branch changes, we cite Corollary 13.

The relevance axioms in the main text are stated inside fixed channels. The proofs below use their equivalent comparison-environment forms.

For $o \in O$, write $K_o : \{*\} \to \Delta(O \times \{*\})$ for the channel that delivers o for sure. On a finite action set A, write $K_o^{\mathrm{id}} : A \to \Delta(O \times A)$ for the channel that delivers o and reports the action, and $K_o^0 : A \to \Delta(O \times \{*\})$ for the channel that delivers o and no record.

Lemma 14 (Fixed-channel and environment relevance)**.** *Given Axioms* (A1)*,* (A5)*,* (A6)*,* (A7)*, and* (A8)*, the following equivalences hold.*

(a) *Axiom* (A3) *is equivalent to the existence of $o^+, o^- \in O$ such that*

$$(\delta_*, K_{o^+}) \succ (\delta_*, K_{o^-}).$$

(b) *Axiom* (A4) *is equivalent to the following condition: there is an outcome $o_* \in O$ such that, for every finite A with $|A| \geq 2$ and every full-support $q \in \Delta(A)$,*

$$(q, K_{o_*}^{\mathrm{id}}) \succ (q, K_{o_*}^0).$$

Proof. For part (a), action processing in both directions identifies each pure lottery in Axiom (A3) with the corresponding singleton experiment; the unrestricted row at the unreached action supplies the reverse pushforward. Replacement transports the strict comparison both ways.

For part (b), restrict the witnessing lotteries to their supports. Two-way action processing identifies them with fair binary full revelation and a channel with two identical rows; two-way record processing identifies the latter with silence. Thus Axiom (A4) is equivalent to

$$((\tfrac{1}{2}, \tfrac{1}{2}), K_{o_*}^{\mathrm{id}}) \succ ((\tfrac{1}{2}, \tfrac{1}{2}), K_{o_*}^0) \tag{14}$$

on a binary action set.

For any full-support binary p, choose $0 < \varepsilon \leq 2 \min_i p_i$ and a record y with $P(y \mid i) = \varepsilon/(2p_i)$. It has mass ε and fair posterior. Reveal only after y in one continuation plan and remain silent everywhere in the other. Equation (14) and Corollary 13 rank the compounds strictly. Full revelation garbles to the first and the second garbles to silence, so $(p, K_{o_*}^{\mathrm{id}}) \succ (p, K_{o_*}^0)$.

For general full-support q on $|A| \geq 2$, report a two-cell partition. Two-way action processing identifies this with binary full revelation and identifies silence with binary silence. Hence the cell report is strictly better than silence; full revelation weakly dominates the cell report by record processing. The converse takes the fair binary prior and reverses the same support identifications. □

A.3 The pure-trace core

Whenever Axiom (A4) is imposed below, fix one of its witnessing outcomes and denote it by o_*. Every channel between nonempty finite sets $P : A \to \Delta(X)$ has the constant-payoff lift

$$K_P(o, x \mid a) := \mathbf{1}_{\{o = o_*\}} P(x \mid a). \tag{15}$$

Throughout this constant-payoff phase, comparisons of pure record channels mean comparisons of these lifts in the model already defined in the main text:

$$q \succeq_P q' \quad :\Longleftrightarrow \quad q \succeq_{K_P} q'.$$

Likewise, $(q,P) \succeq (r,Q)$ denotes the main-text two-block comparison of the lifts, and $P \sqcup Q$, processing, continuation lists, and $P_1 \triangleright (Q_1, \ldots, Q_k)$ denote the lifted main-text constructions. The lift commutes with them up to tagged or singleton-record bijections, neutral by Axiom (A6) in both directions. These are only abbreviations, not a new primitive relation or behavioural condition.

Two pieces of notation will be used repeatedly. If $S : A \to \Delta(B)$ is an action processor, let $S^q P : B \to \Delta(X)$ denote any Bayesian pushforward satisfying

$$(qS)(b)(S^q P)(x \mid b) = \sum_a q(a) S(b \mid a) P(x \mid a).$$

Its rows at zero-mass reports are arbitrary, exactly as in (1). For channels $P : A \to \Delta(X)$ and $Q : B \to \Delta(Y)$, their independent product is

$$(P \otimes Q)((x,y) \mid (a,b)) := P(x \mid a) Q(y \mid b).$$

For $q \in \Delta(A)$ and $P : A \to \Delta(X)$, write

$$m_{qP}(x) := \sum_a q(a) P(x \mid a), \qquad r_x^{qP}(a) := \frac{q(a)P(x \mid a)}{m_{qP}(x)} \quad \text{when } m_{qP}(x) > 0,$$

and define the finite posterior law

$$\mu_{qP} := \sum_{x : m_{qP}(x) > 0} m_{qP}(x)\, \delta_{r_x^{qP}}.$$

Let Id_A be full revelation and U_A the one-record uninformative channel. With logarithms in the fixed base used in the main text, put

$$I_{qP}(A;X) := H(m_{qP}) - \sum_a q(a) H(P(\cdot \mid a)) = H(q) - \int_{\Delta(A)} H(r)\, d\mu_{qP}(r).$$

Proposition 15 (Pure-trace lemma)**.** *Suppose the preference family satisfies main-text Axioms* (A1)*,* (A2)*, and* (A4)–(A8)*. Then, for every record channel between nonempty finite sets* $P : A \to \Delta(X)$ *and every* $q, q' \in \Delta(A)$*,*

$$q \succeq_P q' \quad \Longleftrightarrow \quad I_{qP}(A;X) \geq I_{q'P}(A;X).$$

The same numerical scale represents every block-supported comparison: for every nonempty finite family $\{P_i : A_i \to \Delta(X_i)\}_{i \in J}$ *of channels between nonempty finite sets, distinct* $i, j \in J$*, and priors* $q_i \in \Delta(A_i)$ *and* $q_j \in \Delta(A_j)$*,*

$$q_i^i \succeq_{\bigsqcup_{k \in J} P_k} q_j^j \quad \Longleftrightarrow \quad I_{q_i P_i}(A_i; X_i) \geq I_{q_j P_j}(A_j; X_j).$$

Assume Proposition 15's hypotheses. Every working property below is the corresponding Axiom (A1), (A2), or (A4)–(A8) applied to constant-payoff lifts.

Proof strategy. First, marked terminal laws give affine fixed-prior values, made exact under relabelling and support transport. Second, product and reached-branch calibrations meet; their two-groupings argument eliminates the possible product interaction and aligns every scale. Third, the exact chain rule produces a symmetric, continuous, strongly additive uncertainty functional, identified as Shannon entropy by Faddeev's theorem. The product-coefficient and branch tangent-space lemmas may be read as technical interfaces.

A.3.1 Fixed-prior affine values

For a finite action set A, let X_A be the set of finitely supported probability measures on $\Delta(A)$, and let

$$b(\mu) := \int_{\Delta(A)} s\, d\mu(s), \qquad \mathcal{M}_q := \{\mu \in \mathsf{X}_A : b(\mu) = q\}.$$

Thus $\mathcal{M}_q$ is the mixture set of finite Bayes-plausible posterior laws with prior q, and $\mu_{qP} \in \mathcal{M}_q$ for every finite channel P. On the constant-payoff face, the marked terminal law from (8) is simply

$$\eta_{K_P} = \delta_{o_*} \otimes \mu_{qP}. \tag{16}$$

Let $\delta_q \in \mathcal{M}_q$ be the posterior law of silence and let

$$\chi_q := \sum_{a\in A} q(a)\delta_{e_a} \in \mathcal{M}_q$$

be the posterior law of full revelation.

Lemma 16 (Pure affine fibres)**.** Global order facts. *Pure prior–channel pairs form a weak order with replacement, unaffected by adding or deleting other blocks. For every channel P and priors q, r on its action set,*

$$q \succeq_P r \iff (q,P) \succeq (r,P), \tag{17}$$

and on fixed alphabets the pair-order graph is closed under entrywise convergence of channels and priors. Fixed full-support fibre. *Fix a full-support $q \in \Delta(A)$. Then the following statements hold.*

(a) *If $\mu_{qP} = \mu_{qQ}$, record processors exist in both directions, so $(q,P) \sim (q,Q)$ and either channel may replace the other in a block comparison. Conversely, every $\mu = \sum_{k=1}^{n} \lambda_k \delta_{s_k} \in \mathcal{M}_q$, written with $\lambda_k > 0$, is implemented by*

$$C_\mu(k \mid a) := \frac{\lambda_k s_k(a)}{q(a)}, \qquad \mu_{qC_\mu} = \mu. \tag{18}$$

Hence the fixed-q order descends to all of $\mathcal{M}_q$.

(b) *For publicly tagged mixtures,*

$$\mu_{q,tP\oplus(1-t)R} = t\mu_{qP} + (1-t)\mu_{qR}, \tag{19}$$

and, for $0 < t < 1$,

$$(q,P) \succeq (q,Q) \iff (q, tP \oplus (1-t)R) \succeq (q, tQ \oplus (1-t)R).$$

(c) *There is an affine $F_q : \mathcal{M}_q \to \mathbb{R}$ representing the fixed-prior order:*

$$(q,P) \succeq (q,Q) \iff F_q(\mu_{qP}) \geq F_q(\mu_{qQ}). \tag{20}$$

It may be normalized by

$$F_q(\delta_q) = 0, \qquad F_q(\mu) \geq 0 \quad (\mu \in \mathcal{M}_q), \tag{21}$$

and, whenever $|A| \geq 2$,

$$F_q(\chi_q) = 1. \tag{22}$$

These two anchors select a unique affine representative on every nondegenerate full-support fibre. On a singleton fibre we retain $F_q \equiv 0$.

Proof. The order, replacement, and closed-graph assertions are respectively Lemma 10(a), the duplication clause of Axiom (A5), and Axiom (A2), all specialized to the lifts.

For part (a), equation (16) turns equality of posterior laws into equality of marked terminal laws. The matching processors in Lemma 11(a) therefore give record processors in both directions, including arbitrary completions on zero-mass rows. Replacement gives the quotient order. Conversely, Bayes plausibility gives $\sum_k \lambda_k s_k = q$, so the rows in (18) sum to one and Bayes' formula gives $\mu_{qC_\mu} = \mu$.

For part (b), Bayes' formula gives (19). On constant-payoff lifts this is the public-mixture identity (10), and the independence argument in Lemma 11(b) gives the biconditional. Tagged disjoint unions handle unequal record alphabets, with part (a) making the padding neutral.

For part (c), restrict the affine representative from Lemma 11(b) to laws $\delta_{o_*} \otimes \mu$ and use parts (a)–(b) and canonical implementability. If $|A| \geq 2$, Lemma 14(b) makes the order nonconstant, so positive-affine uniqueness applies. Deleting the record garbles every channel to silence; hence δ_q is a minimum. Translate its value to zero and divide by the positive value of χ_q. This gives the two anchors and nonnegativity. If $|A| = 1$, Bayes plausibility makes $\mathcal{M}_q = \{\delta_q\}$. □

The only continuity used here is the finite-alphabet segment continuity already verified in Lemma 11: implement the finitely many laws on one tagged record alphabet and apply Axiom (A2) along their mixtures. No topology on laws with growing supports, and no pointwise integrand on $\Delta(A)$, is being assumed.

A.3.2 Exact transport across priors and support faces

Action processing handles relabelling, undetectable distinctions, and restriction to the positive support in one step.

For a bijection $\pi : A \to A'$, write $q\pi$ for the transported prior and $\pi_\# \mu$ for the posterior law obtained by transporting every atom of μ. If $B = \operatorname{supp}(q)$, write q_B for the full-support restriction of q to B, P_B for the restricted channel, and $\pi_B \mu$ for the law obtained by restricting every posterior atom to B.

Lemma 17 (Relabelling, undetectable distinctions, and support transport)**.** *The canonically normalized fibre values have the following properties.*

(a) Exact relabelling. *If q has full support and $\pi : A \to A'$ is a bijection, then*

$$F_{q\pi}^{A'}(\pi_\# \mu) = F_q^A(\mu) \qquad (\mu \in \mathcal{M}_q). \tag{23}$$

(b) Undetectable distinctions. *Let $S : A \to B$ be an onto deterministic map, let q have full support, and suppose that $P : A \to \Delta(X)$ is constant on every fibre of S. Then*

$$(q, P) \sim (qS, S^q P). \tag{24}$$

In particular, if G_S reports only $S(a)$, then $(q, G_S) \sim (qS, \mathrm{Id}_B)$.

(c) Support restriction. *For an arbitrary prior $q \in \Delta(A)$ with nonempty support $B := \operatorname{supp}(q)$ and every pair of finite channels $P : A \to \Delta(X)$ and $Q : A \to \Delta(Y)$,*

$$(q, P) \sim (q_B, P_B). \tag{25}$$

Consequently comparisons at q depend only on the rows indexed by B, and

$$(q, P) \succeq (q, Q) \quad \iff \quad (q_B, P_B) \succeq (q_B, Q_B). \tag{26}$$

Every atom of a law $\mu \in \mathcal{M}_q$ is supported on B, and the definition

$$F_q^A(\mu) := F_{q_B}^B(\pi_B \mu) \tag{27}$$

represents all continuation comparisons at the boundary prior q.

Proof. Part (a). Axiom (A7) applied to π and π^{-1} identifies the two orders. Hence $\mu \mapsto F^{A'}_{q\pi}(\pi_{\#}\mu)$ and F^A_q are affine representatives of the same order. Relabelling preserves silence and full revelation, so the two anchors and affine uniqueness give (23); singleton values vanish.

Part (b). One direction of (24) is Axiom (A7). For the reverse, use the Bayesian refinement kernel

$$R(a \mid b) := \begin{cases} q(a)/(qS)(b), & S(a) = b, \\ 0, & S(a) \neq b. \end{cases}$$

It satisfies $(qS)R = q$, and fibre constancy means that pushing $S^q P$ back through R recovers P. A second application of Axiom (A7) gives the reverse comparison. Taking $P = G_S$ gives the final assertion.

Part (c). Lemma 10(b) gives (25), and replacement gives (26). If $\mu = \sum_k \lambda_k \delta_{s_k} \in \mathcal{M}_q$ with $\lambda_k > 0$, then $0 = q(a) = \sum_k \lambda_k s_k(a)$ off B; nonnegativity puts every s_k on $\Delta(B)$. Thus $\pi_B \mu \in \mathcal{M}_{q_B}$, and the full-support representation on B proves (27). Compatibility with relabelling follows from part (a). $\square$

Henceforth action-set superscripts are suppressed, and boundary subscripts mean transport from the positive support.

Normalization ledger. F_q has the silence/full-revelation normalization; product calibration gives $G_q = c_q F_q$ with a possible interaction κ; branch calibration gives $V_q = G_q/a_q$. Compatibility will make a_q universal and $\kappa = 0$.

A.3.3 Scale coherence

Product calibration and a common comparison scale. Write $P \succeq_q Q$ for $(q, P) \succeq (q, Q)$, and similarly for strict preference. For a channel $P : A \to \Delta(X)$, let $\widetilde{P}$ denote its lift to $A \times B$, constant in the B-coordinate; for $R : B \to \Delta(Y)$, let $\widetilde{R}$ denote the symmetric lift, constant in the A-coordinate. By Lemma 16, null records and record permutations are neutral.

Lemma 18 (Cancellation of a common independent background)**.** *Let $p \in \Delta(A)$ and $r \in \Delta(B)$ have full support. For channels P, Q on A and a channel R on B,*

$$P \otimes R \succeq_{p\otimes r} Q \otimes R \quad \Longleftrightarrow \quad P \succeq_p Q. \tag{BI}$$

The symmetric assertion holds with the two coordinates interchanged. The equivalence, including its strict form, remains valid when a posterior reached in the background coordinate lies on a proper support face.

Proof. Pad P and Q by null records so that they have the same record alphabet. Regard the lift $\widetilde{R}$ of R as a first-stage experiment on $A \times B$, followed on every reached branch by $\widetilde{P}$, respectively $\widetilde{Q}$. Up to the harmless permutation of the two record coordinates, the resulting compounds are $P \otimes R$ and $Q \otimes R$.

If y is reached under R, its posterior on $A \times B$ is

$$p \otimes r_y, \qquad r_y := r_y^{r,R}.$$

The coordinate projection $A \times \mathrm{supp}(r_y) \to A$ and Lemma 17(b)–(c) give

$$(p \otimes r_y, \widetilde{P}) \sim (p, P), \qquad (p \otimes r_y, \widetilde{Q}) \sim (p, Q). \tag{BI$'$}$$

Thus every reached branch has the same weak comparison as P versus Q at p. If $P \succeq_p Q$, Lemma 12 applied to these branches proves the forward implication in (BI).

For the converse, suppose that $P \not\succeq_p Q$. Completeness gives $Q \succ_p P$. By (BI$'$), the reversed comparison is strict on every reached branch; at least one such branch has positive mass. The strict part of Lemma 12 therefore gives $Q \otimes R \succ_{p\otimes r} P \otimes R$, contradicting $P \otimes R \succeq_{p\otimes r} Q \otimes R$. This proves the converse and also records why weak branch monotonicity alone would not suffice. Interchange the coordinates for the symmetric assertion. □

For finite posterior laws define

$$\left(\sum_i m_i \delta_{u_i}\right) \otimes \left(\sum_j n_j \delta_{v_j}\right) := \sum_{i,j} m_i n_j \delta_{u_i \otimes v_j}.$$

Bayes' rule gives

$$\mu_{p\otimes r, P\otimes R} = \mu_{pP} \otimes \mu_{rR}. \tag{PL}$$

Lemma 19 (Product gauge)**.** *There are positive rescalings $G_q = c_q F_q$ of the zero-normalised affine representatives and a single number $\kappa \in \mathbb{R}$ such that*

$$G_{p\otimes r}(\mu \otimes \nu) = G_p(\mu) + G_r(\nu) + \kappa G_p(\mu) G_r(\nu) \tag{QA}$$

for all full-support priors p, r and all $\mu \in \mathcal{M}_p$, $\nu \in \mathcal{M}_r$. These rescalings can be chosen exactly invariant under bijections. If a factor is a singleton, its representative is zero and the same formula holds. Finally,

$$1 + \kappa G_r(\nu) > 0 \tag{POS}$$

whenever the r-fibre is nonconstant.

Proof. We give the algebra because both the common value of κ and the strict inequality (POS) are used later. Initially retain the notation F_q. For nondegenerate p, r, put

$$L_{p,r}(\mu, \nu) := F_{p\otimes r}(\mu \otimes \nu), \qquad x := F_p(\mu), \quad y := F_r(\nu).$$

The map $L_{p,r}$ is separately affine, and Lemma 18 says that each slice represents the appropriate factor order. We record the short uniqueness argument. Fixing ν gives $L_{p,r} = \alpha(\nu)x + \gamma(\nu)$ with $\alpha(\nu) > 0$. At $\mu = \delta_p$, affine uniqueness gives $\gamma(\nu) = B_{p,r} y$ for some $B_{p,r} > 0$. At any $\bar{\mu}$ with $\bar{x} := F_p(\bar{\mu}) \neq 0$, the other slice has the form $A_{p,r}\bar{x} + D_{p,r} y$ with $A_{p,r} > 0$. Comparing the two expressions gives $\alpha(\nu) = A_{p,r} + C_{p,r} y$, where $C_{p,r} := (D_{p,r} - B_{p,r})/\bar{x}$. Hence

$$L_{p,r}(\mu, \nu) = A_{p,r} x + B_{p,r} y + C_{p,r} xy. \tag{28}$$

The symmetric calculation and strict slice cancellation give

$$A_{p,r} + C_{p,r} y > 0, \qquad B_{p,r} + C_{p,r} x > 0 \tag{29}$$

throughout the value ranges.

Exact relabelling makes the evaluations under the two associations of a triple product equal. Comparing the coefficients of x, y, z gives

$$A_{p\otimes r,s} A_{p,r} = A_{p,r\otimes s}, \tag{30}$$

$$A_{p\otimes r,s} B_{p,r} = B_{p,r\otimes s} A_{r,s}, \tag{31}$$

$$B_{p\otimes r,s} = B_{p,r\otimes s} B_{r,s}. \tag{32}$$

The coordinate swap similarly gives

$$B_{r,p} = A_{p,r}, \qquad A_{r,p} = B_{p,r}, \qquad C_{r,p} = C_{p,r}. \tag{SW}$$

This coefficient comparison is legitimate: the laws $(1-t)\delta_p + t\chi_p$ make x range over a nondegenerate interval, and the same holds for y and z. Let $\rho(p, r) := B_{p,r}/A_{p,r}$. Dividing (31) by (30) yields

$$\rho(p, r) = \rho(p, r \otimes s) A_{r,s}.$$

Fix a nondegenerate reference prior q_0 and set $g(r) := \rho(q_0, r)$. Then

$$A_{r,s} = \frac{g(r)}{g(r \otimes s)}. \tag{33}$$

The uniqueness of the coefficients in (28), applied after product relabellings, shows that g is invariant under bijections. Consequently the positive rescaling $G_p := g(p)F_p$ preserves exact relabelling. Equation (33) makes the new first coefficient equal to one, and (SW) makes the second coefficient equal to one as well. We henceforth use G for these gauged representatives.

Write the remaining coefficient as $\kappa_{p,r}$. Associativity of

$$x \oplus_{p,r} y := x + y + \kappa_{p,r} xy$$

and comparison of the coefficients of xy, xz, yz give

$$\kappa_{p,r} = \kappa_{p,r\otimes s}, \qquad \kappa_{p\otimes r,s} = \kappa_{p,r\otimes s}, \qquad \kappa_{p\otimes r,s} = \kappa_{r,s}. \tag{KA}$$

(The xyz coefficient gives the product of the corresponding two sides and adds no restriction.) Taking $s = q_0$ in (KA) shows that $\kappa_{p,r} = \kappa_{r,q_0}$, so it is independent of p; the swap in (SW) makes it symmetric, hence independent of r. Call the common value κ. After the gauge, (29) is exactly $1 + \kappa G_r(\nu) > 0$ and its symmetric counterpart.

For a singleton prior set $G \equiv 0$; the canonical bijections $A \cong A \times \{*\}$ give (QA). For a boundary prior, define G_r by exact transport from its positive-probability support, as in Lemma 17. This convention is relabelling invariant. □

Lemma 20 (Cross-prior block representation). *For arbitrary finite priors $q \in \Delta(A)$ and $r \in \Delta(B)$ and channels P, Q on the respective action sets,*

$$q^0 \succeq_{P\sqcup Q} r^1 \iff G_q(\mu_{qP}) \geq G_r(\mu_{rQ}). \tag{BC}$$

Proof. First suppose that q, r have full support. By (QA) and $G_q(\delta_q) = G_r(\delta_r) = 0$,

$$G_{q\otimes r}(\mu_{qP} \otimes \delta_r) = G_q(\mu_{qP}), \qquad G_{q\otimes r}(\delta_q \otimes \mu_{rQ}) = G_r(\mu_{rQ}).$$

The two laws on the left lie in the same fibre and are implemented by $P \otimes U_B$ and $U_A \otimes Q$, so the affine fibre representation compares them by the displayed numbers.

Lemma 10(c) removes the independent dummy coordinates, giving $(q \otimes r, P \otimes U_B) \sim (q, P)$ and $(q \otimes r, U_A \otimes Q) \sim (r, Q)$. Replacement proves (BC). Arbitrary priors follow by support restriction; singleton supports use the zero convention. □

Branch calibration and coherent face scales. Changing a continuation on one reached branch changes the posterior law by a signed measure with zero total mass and zero barycentre. The following space is exactly the space of such feasible directions. For a nonempty finite set A, let

$$W_A := \left\{ \eta : \begin{array}{l} \eta \text{ is a finitely supported signed measure on } \Delta(A), \\ \eta(\Delta(A)) = 0, \quad \int p \, d\eta(p) = 0 \end{array} \right\}.$$

If $\iota : B \hookrightarrow A$ is an injection, let $i_\iota : W_B \to W_A$ push every posterior forward along ι; for a literal subset $B \subseteq A$, write i_B for the inclusion case. Let L_q^A be the linear part of the affine functional G_q on the affine hull of $\mathcal{M}_q$.

Lemma 21 (Branch aggregation). *Let q have full support on A. For every $r \in \Delta(A)$ with $B := \operatorname{supp}(r)$ and $|B| \geq 2$, there is a unique $\beta_A(q, r) > 0$ satisfying*

$$L_q^A \circ i_B = \beta_A(q, r) L_{r_B}^B \quad \text{on } W_B. \tag{34}$$

It depends only on q, r and the face inclusion. If $k \geq 1$ and $P_1 : A \to \Delta(\{1, \dots, k\})$, define

$$m_i := \sum_a q(a)P_1(i \mid a), \qquad r_i(a) := \frac{q(a)P_1(i \mid a)}{m_i} \quad (m_i > 0). \tag{35}$$

Then every list of finite continuation channels $Q_1, \dots, Q_k$ on A satisfies

$$G_q(\mu_{q,P_1 \triangleright (Q_1,\dots,Q_k)}) = G_q(\mu_{qP_1}) + \sum_{i:m_i>0} m_i \beta_A(q, r_i) G_{r_i}(\mu_{r_i,Q_i}). \tag{BA}$$

For singleton r_i, the continuation value is zero and the corresponding coefficient may be fixed by any positive convention.

Proof. We first record the common tangent space. If q has full support, then

$$\operatorname{span}\{\mu - \nu : \mu, \nu \in \mathcal{M}_q\} = W_A. \tag{36}$$

Only the reverse inclusion needs proof. For $0 \neq \eta \in W_A$, write its Jordan decomposition as $\eta = \eta^+ - \eta^-$, where both positive parts have the same mass $c > 0$ and, after division by c, the same barycentre b. Choose $\varepsilon > 0$ small enough that $\tau = (q - \varepsilon b)/(1 - \varepsilon)$ belongs to $\Delta(A)$. Then

$$\mu^\pm := \frac{\varepsilon}{c}\eta^\pm + (1 - \varepsilon)\delta_\tau \in \mathcal{M}_q, \qquad \eta = \frac{c}{\varepsilon}(\mu^+ - \mu^-),$$

which proves (36). The same argument intrinsically on a face gives the space W_B.

Fix a nondegenerate r supported on B. It is reachable from q: choose

$$0 < \varepsilon < \min\left\{1, \min_{b \in B} \frac{q(b)}{r(b)}\right\}$$

and use a two-record experiment whose first record has conditional probability $\varepsilon r(a)/q(a)$. That record has mass ε and posterior r.

Implement $\nu, \nu' \in \mathcal{M}_{r_B}$ by channels on B, extend their off-support rows arbitrarily to A, and pad their record alphabets if necessary. Vary only the continuation on the branch reaching r from ν to ν'. The aggregate posterior law changes by $\varepsilon i_B(\nu - \nu')$. Corollary 13 and support transport give

$$\operatorname{sgn} L_q^A i_B(\nu - \nu') = \operatorname{sgn} L_{r_B}^B(\nu - \nu').$$

Indeed, strict preference handles the positive case, swapping continuations handles the negative case, and indifference gives both inequalities in the zero case. By (36), this sign identity extends by positive scaling to W_B. Positive trace orientation makes

$$L_{r_B}^B(\chi_{r_B} - \delta_{r_B}) = G_{r_B}(\chi_{r_B}) > 0,$$

so $L_{r_B}^B \neq 0$. Two nonzero linear functionals with the same positive half-space have the same kernel and are positive scalar multiples; this gives the unique coefficient in (34). Since both linear functionals in that equation are fixed before a reaching experiment is chosen, the coefficient is path-independent.

For the aggregation formula, put $B_i := \operatorname{supp}(r_i)$ and

$$\nu_i := \pi_{B_i}\mu_{r_i,Q_i} \in \mathcal{M}_{(r_i)_{B_i}}.$$

The compound law differs from the first-stage law by

$$\mu_{q,P_1 \triangleright (Q_1,\dots,Q_k)} - \mu_{qP_1} = \sum_{i:m_i>0} m_i \, i_{B_i}(\nu_i - \delta_{(r_i)_{B_i}}).$$

Apply L_q^A, use (34), and use $G_{r_i}(\delta_{r_i}) = 0$. Singleton summands vanish. This is (BA). $\square$

For full-support $q \in \Delta(A)$, an injection $\iota : B \hookrightarrow A$, and full-support $r \in \Delta(B)$, define

$$\beta_\iota(q, r) := \beta_A(q, \iota_\# r).$$

The image posterior $\iota_\# r$ is reachable by the two-record construction in the proof, and its support inclusion is precisely i_ι. Thus

$$L_q^A i_\iota = \beta_\iota(q, r) L_r^B \quad \text{on } W_B.$$

Lemma 22 (Cocycle, face coherence, and the chain gauge)**.** *There is a relabelling-invariant family of positive scales*

$$\{a_q^A : A \text{ finite}, \ |A| \geq 2, \ q \in \Delta(A) \text{ full support}\}$$

for the coefficients of Lemma 21, such that

$$\beta_\iota(q, r) = \frac{a_q^A}{a_r^B} \tag{FS}$$

for every injection $\iota : B \hookrightarrow A$ with $|A|, |B| \geq 2$ and every full-support $q \in \Delta(A)$ and $r \in \Delta(B)$; here $\beta_\iota(q, r)$ is the coefficient for reaching the posterior $\iota_\# r$ on the face $\iota(B)$. For each prior in this family, put

$$V_q := \frac{G_q}{a_q}$$

. These representatives satisfy the following exact chain rule. For every finite A with $|A| \geq 2$, every full-support $q \in \Delta(A)$, every $k \geq 1$, every $P_1 : A \to \Delta(\{1, \ldots, k\})$, and every list of finite continuation channels $Q_1, \ldots, Q_k$ on A, use the branch quantities in (35). Then

$$V_q(\mu_{q, P_1 \triangleright (Q_1, \ldots, Q_k)}) = V_q(\mu_{q P_1}) + \sum_{i : m_i > 0} m_i V_{r_i}(\mu_{r_i, Q_i}). \tag{CR}$$

Singleton continuation terms are zero.

Proof. There are two normalization tasks: first solve the branch cocycle within each fixed action-set dimension, then remove the residual scalar attached to embedding an m-point face in an n-point simplex.

We first prove the cocycle before choosing scales. For nested injections $C \overset{\jmath}{\hookrightarrow} B \overset{\iota}{\hookrightarrow} A$ and full-support priors q, r, s on A, B, C, respectively, the defining identities on W_C are

$$L_q^A i_{\iota \circ \jmath} = \beta_\iota(q, r) L_r^B i_\jmath = \beta_\iota(q, r) \beta_\jmath(r, s) L_s^C.$$

The direct coefficient from q to s is unique because $L_s^C \neq 0$ by Axiom (A4). Therefore

$$\beta_{\iota \circ \jmath}(q, s) = \beta_\iota(q, r) \beta_\jmath(r, s). \tag{37}$$

This single identity covers full-support paths, full-to-face paths, and nested support faces.

For a fixed nondegenerate A, apply (37) to identity embeddings. Fix a full-support basepoint q_A and put

$$a_q^A := \beta_{\mathrm{id}_A}(q, q_A).$$

Then

$$\beta_{\mathrm{id}_A}(q, r) = \frac{a_q^A}{a_r^A}. \tag{WS}$$

Choosing q_A to be uniform and using exact relabelling of G makes these initial scales invariant under bijections. They remain free up to one positive multiplier for each cardinality.

For an injection $\iota : B \hookrightarrow A$, define its residual defect by

$$K(\iota) := \beta_\iota(q,r)\frac{a_r^B}{a_q^A}.$$

Equation (WS) shows that this number is independent of q and r: changing either prior merely inserts and cancels the corresponding within-set coefficient. Equation (37) gives

$$K(\iota \circ \jmath) = K(\iota)K(\jmath).$$

Exact relabelling implies that $K(\iota)$ depends only on $n = |A|$ and $m = |B|$; write it $K_{n,m}$. Thus

$$K_{n,\ell} = K_{n,m}K_{m,\ell}, \qquad K_{n,n} = 1.$$

Set $t_n := K_{n,2}$ and $t_2 := 1$. Then $K_{n,m} = t_n/t_m$. Replacing every a_q^A on an n-point set by $t_n a_q^A$ changes the defect to $K_{n,m}t_m/t_n = 1$, while preserving (WS) and relabel invariance. This proves (FS).

For a boundary prior, a_r means the scale on its support; the preceding construction says precisely that this agrees with the scale obtained by reaching that support as a face. Divide (BA) by a_q and use (FS). Every nondegenerate branch coefficient becomes

$$\frac{1}{a_q}\beta_A(q,r_i)G_{r_i} = \frac{1}{a_{r_i}}G_{r_i} = V_{r_i}.$$

Singleton values vanish, so their scales and coefficients may be fixed arbitrarily. This proves (CR). □

Compatibility of the product and branch calibrations. The product and sequential evaluations of full revelation now meet. Their compatibility first removes the provisional product interaction and then leaves a strongly additive uncertainty functional.

Lemma 23 (Interaction collapse and universal chain scale)**.** *The coefficient κ in (QA) is zero. Moreover there is a single $C > 0$ such that $a_q = C$ for every finite prior with support of size at least two, and the same value may be assigned on singleton supports. Consequently the final representatives $V_q = G_q/C$ satisfy, for all nonempty finite sets A, B, all priors $p \in \Delta(A)$ and $r \in \Delta(B)$, and all $\mu \in \mathcal{M}_p$ and $\nu \in \mathcal{M}_r$,*

$$V_{p\otimes r}(\mu \otimes \nu) = V_p(\mu) + V_r(\nu). \tag{PA}$$

The chain rule (CR) extends from full-support fibres to all finite priors by exact support transport. The cross-prior bridge of Lemma 20 is valid with V in place of G.

Proof. In this proof only, put

$$H_G(q) := G_q(\chi_q), \qquad Z(q) := 1 + \kappa H_G(q).$$

For nondegenerate q, Axiom (A4) and zero normalisation give $H_G(q) > 0$, while the strict slice-slope (POS) gives

$$Z(q) > 0. \tag{38}$$

For a singleton, $H_G = 0$ and $Z = 1$.

Product revelation versus sequential revelation. For nondegenerate full-support p, r, product quasi-additivity gives

$$H_G(p \otimes r) = H_G(p) + H_G(r) + \kappa H_G(p)H_G(r). \tag{QH}$$

Compute the same full revelation by first revealing the A-coordinate and then the B-coordinate. The first-stage law is $\chi_p \otimes \delta_r$ and has value $H_G(p)$ by (QA). After action a is revealed, the posterior is $e_a \otimes r$ on the face $\{a\} \times B$; support restriction and relabelling identify the continuation value with $H_G(r)$. Every such branch has coefficient $a_{p\otimes r}/a_r$ by (FS). Since the branch weights sum to one, Lemma 21 gives

$$H_G(p \otimes r) = H_G(p) + \frac{a_{p\otimes r}}{a_r} H_G(r). \tag{SA}$$

Comparing (QH) and (SA), and using $H_G(r) > 0$, yields

$$\frac{a_{p\otimes r}}{a_r} = Z(p). \tag{R1}$$

Revealing B first gives symmetrically

$$\frac{a_{p\otimes r}}{a_p} = Z(r). \tag{R2}$$

Thus $a_r Z(p) = a_p Z(r)$ for every pair of nondegenerate priors, even on different finite sets. By (38), there is a universal $C > 0$ such that

$$a_q = CZ(q). \tag{AS}$$

Assign $a = C$ on singleton supports.

The grouping weight. Let q have full support and let $\mathcal{P} = \{A_1, \ldots, A_m\}$ partition its action set into nonempty cells. Put

$$p_i = q(A_i), \qquad q^i = q(\,\cdot \mid A_i), \qquad p = (p_1, \ldots, p_m).$$

If $C_{\mathcal{P}}$ reports only the cell, undetectable-distinctions neutrality gives

$$(q, C_{\mathcal{P}}) \sim (p, \mathrm{Id}_{\{1,\ldots,m\}}).$$

The cross-prior bridge therefore identifies the coarse value as

$$G_q(\mu_{q,C_{\mathcal{P}}}) = H_G(p). \tag{BG}$$

Append full revelation within each cell. The final law is χ_q, and the branch formula together with (AS) gives

$$H_G(q) = H_G(p) + Z(q) \sum_i p_i \frac{H_G(q^i)}{Z(q^i)}. \tag{GR}$$

Here singleton cells contribute zero.

If $\kappa = 0$, then $Z \equiv 1$ and the conclusion follows. Otherwise set $w(q) := 1/Z(q) > 0$. Because $\kappa \neq 0$ in this case, $\kappa H_G(x) = Z(x) - 1$. Substitution into (GR) gives

$$Z(q) - 1 = Z(p) - 1 + Z(q) \sum_i p_i \left(1 - \frac{1}{Z(q^i)}\right).$$

Cancelling and inverting yields the positive grouping-weight equation

$$w(q) = w(p) \sum_i p_i w(q^i). \tag{W}$$

All distributions in what follows are read on their positive-probability supports. For $u \otimes v$, partition by the first coordinate: the coarse distribution is u and every conditional distribution is v. Thus (W) gives

$$w(u \otimes v) = w(u)w(v). \tag{M}$$

Two groupings force the weight to be constant. Take arbitrary finite probability vectors u, v and write $x = w(u)$ and $y = w(v)$. Let T be the law of a triple (ℓ, z_1, z_2): ℓ is uniform on $\{0, 1\}$, and conditional on $\ell = 0$ the pair (z_1, z_2) has law $u \otimes u$, while conditional on $\ell = 1$ it has law $v \otimes v$. In tagged notation,

$$T := \tfrac{1}{2}(u \otimes u)^0 \sqcup \tfrac{1}{2}(v \otimes v)^1.$$

Grouping first by ℓ and using (M) gives

$$w(T) = w(\tfrac{1}{2}, \tfrac{1}{2})\frac{x^2 + y^2}{2}. \tag{E1}$$

Instead group first by (ℓ, z_1). That marginal is $S = \frac{1}{2}u^0 \sqcup \frac{1}{2}v^1$, so

$$w(S) = w(\tfrac{1}{2}, \tfrac{1}{2})\frac{x + y}{2}.$$

The conditional law of z_2 is u when $\ell = 0$ and v when $\ell = 1$. A second use of (W) therefore gives

$$w(T) = w(S)\frac{x + y}{2} = w(\tfrac{1}{2}, \tfrac{1}{2})\left(\frac{x + y}{2}\right)^2. \tag{E2}$$

The common leading factor is positive. Comparing (E1) and (E2) yields $(x - y)^2 = 0$. Thus w is constant on all finite probability vectors. Its singleton value is one, so $w \equiv 1$.

It follows that $Z \equiv 1$, hence $\kappa H_G(q) = 0$ for every q. Axiom (A4) supplies a nondegenerate q with $H_G(q) > 0$, so $\kappa = 0$, contradicting the temporary alternative. Equation (AS) now gives $a_q = C$ universally. Dividing by C proves (PA) and (CR) at full support; exact support transport extends them to boundary priors. The common positive division also preserves the cross-prior bridge. □

A.3.4 Entropy identification

Lemma 24 (Entropy reduction and Faddeev identification)**.** *For every nonempty finite A and every $q \in \Delta(A)$, define the value of full revelation by*

$$\mathcal{H}(q) := V_q(\chi_q),$$

with boundary priors read on their support. For every nonempty finite X and every channel $P : A \to \Delta(X)$, put $m_x := \sum_a q(a)P(x \mid a)$ and, for each reached record x, $r_x(a) := q(a)P(x \mid a)/m_x$. Then

$$V_q(\mu_{qP}) = \mathcal{H}(q) - \sum_{x:m_x>0} m_x \mathcal{H}(r_x). \tag{ER}$$

Moreover, there is a single $\alpha > 0$, independent of all finite alphabets and priors, such that

$$\mathcal{H}(q) = \alpha H(q)$$

and consequently

$$V_q(\mu_{qP}) = \alpha I_{qP}(A; X). \tag{MI}$$

Proof. Entropy reduction. Assume first that q has full support. Append full revelation after P. The final posterior law is χ_q, while the continuation after record x has full-revelation law χ_{r_x} on the support of r_x. The exact chain rule (CR) therefore gives

$$\mathcal{H}(q) = V_q(\mu_{qP}) + \sum_{x:m_x>0} m_x \mathcal{H}(r_x),$$

which is (ER). A boundary prior is reduced to its support; null action rows and null records affect neither side.

Strong additivity. Let $\mathcal{P} = \{A_1, \dots, A_m\}$ be a partition of the support of q, and put

$$p_i = q(A_i), \qquad q^i = q(\,\cdot \mid A_i), \qquad p = (p_1, \dots, p_m).$$

In (GR), Lemma 23 gives $Z \equiv 1$ and $H_G = C\mathcal{H}$. Dividing that identity by C yields

$$\mathcal{H}(q) = \mathcal{H}(p) + \sum_i p_i \mathcal{H}(q^i). \tag{FA}$$

Relabelling invariance makes $\mathcal{H}$ symmetric, support restriction makes it expansive, and $\mathcal{H}$ vanishes on point masses. Record processing gives $\mathcal{H} \geq 0$, while Axiom (A4) gives strict positivity at nondegenerate full-support priors. Equation (FA) also allows zero outer weights by deleting and reinserting null cells.

Continuity: the binary case. Write $h(t) := \mathcal{H}(t, 1-t)$. We derive, rather than assume, its continuity from Axiom (A2). Fix $c \geq 0$. Choose a full-support binary prior θ with $\mathcal{H}(\theta) > 0$, as supplied by Axiom (A4). Choose $n \geq 1$ and $\lambda \in [0, 1]$ so that

$$c = \lambda n \mathcal{H}(\theta).$$

For $\theta_n := \theta^{\otimes n}$, product additivity (PA) gives $\mathcal{H}(\theta_n) = n\mathcal{H}(\theta)$. Let $E_{n,\lambda}$ reveal the n-bit action fully with probability λ and report a common erasure symbol otherwise. Its posterior law is $\lambda \chi_{\theta_n} + (1-\lambda)\delta_{\theta_n}$, so affinity gives

$$V_{\theta_n}(\mu_{\theta_n, E_{n,\lambda}}) = c.$$

Now fix the block environment

$$K_c := \mathrm{Id}_{\{0,1\}} \sqcup E_{n,\lambda}.$$

Let q_t^0 put the binary prior $(t, 1-t)$ on its first block and let θ_n^1 put θ_n on its second block. The cross-prior bridge says

$$\begin{aligned} h(t) \geq c &\iff q_t^0 \succeq_{K_c} \theta_n^1, \\ h(t) \leq c &\iff \theta_n^1 \succeq_{K_c} q_t^0. \end{aligned}$$

The environment is fixed and $t \mapsto q_t^0$ is continuous, so Axiom (A2) makes both level sets closed for every $c \geq 0$. For a threshold $c < 0$, the superlevel set is all of $[0, 1]$ and the sublevel set is empty because $h \geq 0$. Thus every upper and lower level set of h is closed, and h is continuous.

Continuity on every finite simplex. Proceed by induction on the number of coordinates. For $q = (q_1, (1-q_1)\bar{q})$ with $q_1 < 1$, recursion (FA) for the partition $\{\{1\}, \{2, \dots, n\}\}$ gives

$$\mathcal{H}(q) = h(q_1) + (1-q_1)\mathcal{H}(\bar{q}). \tag{39}$$

The right side is continuous away from $q_1 = 1$ by binary continuity and the induction hypothesis. At $q_1 = 0$, support restriction and $h(0) = 0$ give the same formula on the face. At $q_1 = 1$, the induction hypothesis makes $\mathcal{H}$ bounded on the compact $(n-1)$-simplex, so the second term in (39) tends to zero and the first tends to $h(1) = 0$. Relabelling handles the other faces. Hence $\mathcal{H}$ is continuous on each closed finite simplex.

We have now verified all hypotheses of the strong-additivity form of Faddeev's theorem. Axiom (A2) supplied continuity; Lemma 17 supplied relabelling invariance and expansibility; $\mathcal{H}$ vanishes on point masses; and Lemmas 22 and 23 supplied strong additivity (FA). Record processing has already established $\mathcal{H} \geq 0$. The finite entropy characterisation [6, Theorem 6] therefore gives

$$\mathcal{H}(q) = \alpha H(q)$$

for some $\alpha \geq 0$. Axiom (A4) forces $\alpha > 0$. Substitution in (ER) gives

$$V_q(\mu_{qP}) = \alpha \left(H(q) - \sum_x m_x H(r_x) \right) = \alpha I_{qP}(A; X),$$

where the sum is over positive-mass records. This proves (MI). □

Conclusion. The cross-prior bridge and Lemma 24 give, for arbitrary pairs (q, P) and (r, Q),

$$q^0 \succeq_{P \sqcup Q} r^1 \iff I_{qP} \geq I_{rQ}, \tag{BMI}$$

including boundary priors by support restriction. With $Q = P$, the duplication clause of Axiom (A5) gives the proposition's fixed-channel claim. The positive multiplier α cancels from these comparisons; its cardinal calibration against payoff utility is the next step. Thus the conclusion here is ordinal: it fixes the common ranking of pure-trace pairs, not their exchange rate against material utility. For distinct $i, j \in J$, the irrelevant-block clause of Axiom (A5) reduces an arbitrary block union to the canonical two-block comparison:

$$q_i^i \succeq_{\bigsqcup_{k \in J} P_k} q_j^j \iff q_i^0 \succeq_{P_i \sqcup P_j} q_j^1.$$

On the information side, the block tag is constant under a block-supported prior, so

$$I_{q_i^i, \bigsqcup_{k \in J} P_k} = I_{q_i P_i},$$

and similarly for block j. Applying (BMI) to the right-hand side proves the block claim and completes Proposition 15.

A.4 The two common scales

Lemma 25 (Material utility and a common affine scale)**.** *Under Axioms* (A1)–(A3) *and* (A5)–(A8)*, there are outcomes o^+, o^- and a nonconstant function $u : O \to \mathbb{R}$ such that*

$$u(o^+) = 1, \qquad u(o^-) = 0.$$

For every payoff lottery $\ell \in \Delta(O)$, the payoff-lottery order is represented by $\sum_o \ell(o)u(o)$. For every full-support prior q, the representatives in Lemma 11 may be normalised so that

$$W_q(\ell) = \sum_o \ell(o)u(o). \tag{40}$$

With this normalisation, for every full-support $q \in \Delta(A)$, every nonempty finite B, every full-support $s \in \Delta(B)$, and every finite experiment E at q,

$$W_{q \otimes s}(E^{\uparrow}) = W_q(E),$$

where $E^{\uparrow}$ ignores the dummy coordinate. Moreover, for every pair of nonempty finite action sets, every two full-support priors q, q' on those sets, and every two finite experiments E, G on the respective sets,

$$(q, E) \succeq (q', G) \iff W_q(E) \geq W_{q'}(G). \tag{41}$$

Proof. On a singleton action set, terminal-law sufficiency identifies an experiment with its payoff lottery. Public-mixture independence and continuity, proved as in Lemma 11, give an affine utility on $\Delta(O)$. By Lemma 14, Axiom (A3) supplies $o^+ \succ o^-$. Normalise their values to one and zero and let $u(o)$ be the value of the degenerate lottery at o. Affinity on the finite simplex gives expected utility.

An action processor may ignore its input and draw an arbitrary target prior. Axiom (A7), applied in both directions, therefore makes this payoff lottery order independent of the prior and action alphabet. Normalise each W_q at the same two deterministic payoffs. The payoff-lottery embedding is affine and order-reflecting, so positive-affine uniqueness gives (40).

Dummy lifting between full-support fibres is likewise an affine order embedding. Positive-affine uniqueness initially permits

$$W_{q \otimes s}(E^{\uparrow}) = cW_q(E) + d, \qquad c > 0.$$

The two payoff anchors have values zero and one on both sides; hence $c = 1$ and $d = 0$. To compare E at q with G at q', lift each to the common product prior $q \otimes q'$, using the other prior as the dummy. Equation (7), the replacement part of Lemma 10, and the common product representative give (41). $\square$

Retain the outcome o_* fixed in the preceding pure-trace subsection to witness Axiom (A4). Proposition 15 has established that mutual information represents every within-channel and block-supported comparison on this constant-payoff face. The normalization $u(o^+) = 1$ and $u(o^-) = 0$ fixes only the material origin and unit. The trace anchor o_* need not be either material anchor, and its intercept $u(o_*)$ is retained. It remains to identify the cardinal coefficient of mutual information relative to the material unit and to show that this coefficient is common to all priors and finite alphabets.

Lemma 26 (The trace scale)**.** *Under Axioms* (A1)–(A8)*, there is one number $\lambda > 0$, independent of the prior and all finite alphabets, such that, for every nonempty finite A with $|A| \geq 2$, every full-support $q \in \Delta(A)$, every nonempty finite record set R, and every experiment $E : A \to \Delta(O \times R)$ that pays o_* surely,*

$$W_q(E) = u(o_*) + \lambda I_q(E). \tag{42}$$

Proof. Fix a nontrivial full-support q. On the constant-o_* face, W_q is affine by Lemma 11. For any experiment E, the finite entropy identity

$$I_q(E) = H(q) - \sum_{(o,\pi)\in \operatorname{supp} \eta_E} \eta_E(o,\pi) H(\pi). \tag{43}$$

shows that mutual information depends only on the marked terminal law and is affine under public mixtures. Proposition 15 says that W_q and I_q represent the same nonconstant order on the constant-o_* face. Positive-affine uniqueness therefore gives

$$W_q(E) = \lambda_q I_q(E) + b_q, \qquad \lambda_q > 0.$$

The uninformative experiment has mutual information zero and, by (40), value $u(o_*)$. Thus $b_q = u(o_*)$.

The coefficient does not depend on q. Let q and q' be nontrivial full-support priors. Lift full revelation at q to $q \otimes q'$ by adding an independent dummy coordinate. Normalised value and mutual information are both unchanged. Since the common information value is $H(q) > 0$, cancellation gives $\lambda_{q\otimes q'} = \lambda_q$. Lifting full revelation from the other coordinate gives $\lambda_{q\otimes q'} = \lambda_{q'}$. Hence all such coefficients agree. Equivalently, fix the uniform prior on a two-point dummy alphabet and call its coefficient λ. Axiom (A4) fixes the strict orientation, so $\lambda > 0$.

Finally, any experiment that pays o_* surely is a pure record experiment after the constant payoff coordinate is deleted; restoring that coordinate changes neither its marked terminal law nor its mutual information. This proves (42). $\square$

A.5 Exact branch weights

Let A and Y be arbitrary nonempty finite sets, let $q \in \Delta(A)$ have full support with $|A| \geq 2$, and let $P : A \to \Delta(Y)$ be a first-stage record channel. For every $y \in Y$, put

$$m_y := \sum_{a\in A} q(a) P(y \mid a).$$

Whenever $m_y > 0$, define the posterior on A and its support by

$$\widehat{r}_y(a) := \frac{q(a)P(y \mid a)}{m_y}, \qquad S_y := \operatorname{supp} \widehat{r}_y,$$

and write $r_y \in \Delta(S_y)$ for the full-support restriction of $\widehat{r}_y$ to S_y.

We need notation for changing a continuation without fixing what happens on the other branches. For a finite continuation experiment $E : S_y \to \Delta(O \times R_E)$, choose a retraction $\rho_y : A \to S_y$ whose restriction to S_y is the identity and set $E^{\uparrow}(\cdot \mid a) := E(\cdot \mid \rho_y(a))$. Given an arbitrary background $\mathbf{H} = (H_z)_{z \neq y}$ of continuation channels on A, let

$$K_z^E := \begin{cases} E^{\uparrow}, & z = y, \\ H_z, & z \neq y, \end{cases} \qquad J_{y,\mathbf{H}}(E) := P \triangleright (K_z^E)_{z \in Y}.$$

Here and below, fix any ordering of Y when writing the compound. Before taking the compound, all continuation record alphabets are embedded as tagged copies in one finite disjoint union. Payoff-preserving record isomorphisms make the choice of tags immaterial. Also, $a \notin S_y$ implies $P(y \mid a) = 0$, because q has full support; hence the rows of $E^{\uparrow}$ outside S_y are never used on branch y. The compound channel itself is therefore independent of the chosen retraction.

Lemma 27 (Exact reached-branch weight)**.** *Under Axioms* (A1)–(A8)*, for every nonempty finite A, Y, every full-support $q \in \Delta(A)$ with $|A| \geq 2$, every $P : A \to \Delta(Y)$, every reached $y \in Y$, every two finite continuation experiments E, G on S_y, and every finite background $\mathbf{H} = (H_z)_{z \neq y}$,*

$$W_q(J_{y,\mathbf{H}}(E)) - W_q(J_{y,\mathbf{H}}(G)) = m_y\big[W_{r_y}(E) - W_{r_y}(G)\big]. \tag{44}$$

Proof. At the branch posterior $\widehat{r}_y$, support restriction identifies $(\widehat{r}_y, E^{\uparrow})$ with (r_y, E), and likewise for G. Corollary 13 therefore gives, for every E and G,

$$E \succeq_{r_y} G \quad \Longleftrightarrow \quad J_{y,\mathbf{H}}(E) \succeq_q J_{y,\mathbf{H}}(G),$$

with the same equivalence for strict preference. This statement includes continuations with different record alphabets: the tagged disjoint union puts them in the common comparison environment required by the corollary.

To see the affine structure explicitly, let $\iota_y : \Delta(S_y) \to \Delta(A)$ pad a distribution by zeros outside S_y. There is a fixed subprobability measure $\beta_{y,\mathbf{H}}$, of mass $1 - m_y$, containing the terminal atoms from all branches other than y, such that

$$\eta_{J_{y,\mathbf{H}}(E)} = \beta_{y,\mathbf{H}} + m_y(\mathrm{id}_O, \iota_y)_{\#}\eta_E.$$

Thus insertion is affine on marked terminal laws. If S_y is nontrivial, the local order is nonconstant: at payoff o_*, full revelation and silence have, by Lemma 26, value difference $\lambda H(r_y) > 0$. Affine uniqueness consequently gives, for every fixed background $\mathbf{H}$, constants $c_{y,\mathbf{H}} > 0$ and $d_{y,\mathbf{H}} \in \mathbb{R}$ such that

$$W_q(J_{y,\mathbf{H}}(E)) = c_{y,\mathbf{H}} W_{r_y}(E) + d_{y,\mathbf{H}}. \tag{45}$$

It remains to identify the coefficient and to show that the background cannot alter it. Let $\mathbf{H}^*$ pay o_* for sure and reveal only a singleton record after every $z \neq y$. For every continuation E, the mutual-information chain rule gives

$$I_q(J_{y,\mathbf{H}^*}(E)) = I_q(A; Y) + m_y I_{r_y}(E). \tag{46}$$

Take for E full revelation, E^{id}, and for G silence, E^0, both at the constant payoff o_*. By Lemma 26, their local value difference is $\lambda H(r_y)$, while (46) makes their aggregate value difference $\lambda m_y H(r_y)$. Comparing these differences in (45), and using $\lambda H(r_y) > 0$, yields $c_{y,\mathbf{H}^*} = m_y$.

Now fix arbitrary E, G, and an arbitrary background $\mathbf{H}$. The crossed public half-mixtures

$$\tfrac{1}{2}J_{y,\mathbf{H}}(E) \oplus \tfrac{1}{2}J_{y,\mathbf{H}^*}(G) \quad \text{and} \quad \tfrac{1}{2}J_{y,\mathbf{H}}(G) \oplus \tfrac{1}{2}J_{y,\mathbf{H}^*}(E)$$

have the same marked terminal law. Indeed, each contains one half of each background contribution and, on branch y, one half of each of $m_y\eta_E$ and $m_y\eta_G$ (with posterior supports padded by ι_y). Lemma 11(a) makes the mixtures indifferent. Affinity of W_q and cancellation therefore give

$$W_q(J_{y,\mathbf{H}}(E)) - W_q(J_{y,\mathbf{H}}(G)) = W_q(J_{y,\mathbf{H}^*}(E)) - W_q(J_{y,\mathbf{H}^*}(G)).$$

Applying (45) to the right-hand side with $c_{y,\mathbf{H}^*} = m_y$ proves (44) whenever $|S_y| \geq 2$.

If S_y is a singleton, let s be a full-support prior on a two-point set B. Adjoin B independently to the global prior, lift the first-stage channel by $P^{\uparrow}(y \mid a, b) = P(y \mid a)$, and lift every local and background continuation so that it ignores b. The reached posterior is $r_y \otimes s$, its support is nontrivial, and its branch mass remains m_y. The preceding argument applies to this common dummy extension. Lemma 25 says that adjoining or deleting the independent dummy preserves both aggregate values and both local values. Removing it gives (44) for the singleton case. ☐

For a deterministic payoff profile $g : Y \to O$, let $C(P, g)$ denote the compound that first runs P and then pays $g(y)$ with no further record. Write g_y^o for the profile that pays o at y and o_* elsewhere.

Lemma 28 (Deterministic branch assembly). *Under Axioms* (A1)–(A8)*, for every nonempty finite A, Y, every full-support $q \in \Delta(A)$ with $|A| \geq 2$, every $P : A \to \Delta(Y)$, and every deterministic payoff profile $g : Y \to O$,*

$$W_q(C(P, g)) = \lambda I_q(A; Y) + \sum_{y \in Y} m_y u(g(y)). \tag{47}$$

Proof. Fix $y \in Y$ and $o \in O$. If $m_y = 0$, changing the payoff after y leaves the marked terminal law unchanged. If $m_y > 0$, apply Lemma 27 with deterministic local payoffs o and o_* and with every other branch paying o_*. Equation (40) gives the local value difference, and hence

$$W_q(C(P, g_y^o)) - W_q(C(P, o_*)) = m_y[u(o) - u(o_*)]. \tag{48}$$

The arbitrary-background clause of Lemma 27 gives the same increment when the payoffs on the other branches have already been changed. Enumerate the reached branches and change their payoffs from o_* to $g(y)$ one at a time; unreached branches have no effect. The resulting telescope is

$$W_q(C(P, g)) - W_q(C(P, o_*)) = \sum_{y \in Y} m_y[u(g(y)) - u(o_*)]. \tag{49}$$

The baseline compound pays o_* surely and its complete record is Y. Lemma 26 therefore gives $W_q(C(P, o_*)) = u(o_*) + \lambda I_q(A; Y)$. Substitution in (49), followed by $\sum_{y \in Y} m_y = 1$, proves the formula. ☐

A.6 Proof of the theorem

Proof of Theorem 1. Assume first Axioms (A1)–(A8). Let q have full support on a nontrivial action set, and let $K : A \to \Delta(O \times R)$ be arbitrary. Put $Y = O \times R$ and take as first stage

$$P_K(y \mid a) := K(y \mid a), \qquad y = (o, r).$$

After $y = (o, r)$, pay o for sure and reveal nothing further. Call the resulting compound K^{seq}. Conditional on its final record $((o, r), *)$, the posterior over actions is exactly the posterior conditional on (o, r) under K, and both records carry payoff o. Thus K and K^{seq} have the same marked terminal law. By Lemma 11(a), $(q, K) \sim (q, K^{\text{seq}})$.

Apply Lemma 28 with $g(o, r) = o$. The two identities needed are

$$I_q(A; Y) = I_{qK}(A; O, R)$$

and

$$\sum_{o,r} p_{qK}(o,r)u(o) = \mathbb{E}_{qK}[u(O)],$$

respectively. Together with sequentialisation, they give

$$W_q(K) = \mathbb{E}_{qK}[u(O)] + \lambda I_{qK}(A;O,R). \tag{50}$$

If A is a singleton, mutual information is zero and the marked terminal law is that of the induced payoff lottery. Equation (40) gives (50) directly. If q is on the boundary, put $S = \operatorname{supp} q$ and write (q^S, K^S) for the restriction. Then

$$(q,K) \sim (q^S, K^S), \qquad V(q,K) = V(q^S, K^S). \tag{51}$$

The first identity is Lemma 10(b); the second follows directly from the finite sums defining expected utility and mutual information. Thus no representative W_q is assigned to a boundary fibre.

For arbitrary finite experiments (q,K) and (p,L), restrict both priors to their supports, apply (41) and (50) there, and use (51) to transport the comparison back. This gives

$$(q,K) \succeq (p,L) \quad \Longleftrightarrow \quad V(q,K) \geq V(p,L), \tag{52}$$

with the same u and λ for every pair of finite alphabets. The duplication clause of Axiom (A5) turns this pair formula into the within-channel representation. Its irrelevant-block clause applies the same formula to any two block-supported priors in any finite block environment. This is exactly the theorem's same-scale moreover clause. The construction made u nonconstant and $\lambda > 0$, proving the forward implication.

Conversely, suppose that a nonconstant $u : O \to \mathbb{R}$ and $\lambda > 0$ represent every finite environment by V. Completeness and transitivity of the order on $\mathbb{R}$ give Axiom (A1). On every fixed finite (A,O,R), expected utility and mutual information are jointly continuous in (K,q); for mutual information, use its finite entropy formula with $0 \log 0 = 0$. Therefore

$$\{(K,q,p) : V(q,K) \geq V(p,K)\}$$

is closed, which is Axiom (A2).

For a duplicated channel, a prior supported on either tagged copy has the same payoff law and information as its untagged prior. This proves both directions of the duplication clause in Axiom (A5). More generally, under a prior supported on block i of $\bigsqcup_{k\in I} K_k$, the block label is constant and

$$V\left(q_i^i, \bigsqcup_{k\in I} K_k\right) = V(q_i, K_i).$$

The same identity for block j proves both directions of the irrelevant-block clause, for every finite I and every ordered pair $i \neq j$. Thus Axiom (A5) holds.

A payoff-preserving record processor leaves the payoff marginal unchanged and, by the Markov chain $A - (O,R) - (O,R')$, weakly lowers mutual information. This is Axiom (A6). For every action processor $S : A \to \Delta(B)$ and every Bayesian pushforward in (1), the payoff–record marginal is again unchanged and the undivided joint law forms the Markov chain

$$(O,R) - A - B.$$

Hence $I(B;O,R) \leq I(A;O,R)$, proving Axiom (A7). A completion on a processed action with $(qS)(b) = 0$ has zero joint mass, so this argument covers every completion allowed by the axiom.

For Axiom (A8), fix every binary first-stage channel and continuation triple covered by that axiom, and put

$$\alpha := \sum_{a\in A} q(a)P(1 \mid a) > 0.$$

The expected-utility and mutual-information chain rules give the exact difference

$$\begin{aligned} &V(q, P \triangleright (K, M)) - V(q, P \triangleright (L, M)) \\ &\quad = \alpha\big[V(q_1, K) - V(q_1, L)\big]. \end{aligned}$$

The first-stage information term and the common continuation after record 2 cancel. Since $\alpha > 0$, the aggregate difference is nonnegative exactly when the conditional difference is nonnegative, and it is positive exactly when the conditional difference is positive. This proves the weak biconditional in Axiom (A8) and the strict biconditional.

Finally, nonconstancy of u supplies distinct $o^+, o^- \in O$ with $u(o^+) > u(o^-)$. In the two-action channel specified by Axiom (A3), the two pure lotteries have those values and zero mutual information, proving material relevance. For any $o \in O$, the two lotteries in the four-action channel specified by Axiom (A4) both pay o surely, while their information values are $\log 2$ and zero. Because $\lambda > 0$, the required strict comparison holds (in fact at every o). This verifies Axioms (A3) and (A4) and completes the converse. The displayed block identity also proves the theorem's moreover clause with the same u and λ. $\square$

B Further results and choice implications

This appendix follows the order of Sections 3 and 4. It first proves the representation's immediate consequences and the independence result, then turns to the trace test, deliberate randomisation, record value, and measurement. None of these results is used to prove Theorem 1; Appendix A contains the complete representation argument.

For $q \in \Delta(A)$ and $K : A \to \Delta(O \times R)$, write $E_{qK} := \mathbb{E}_{qK}[u(O)]$ and, whenever $p_{qK}(o,r) > 0$,

$$\pi^{qK}_{or}(a) := \frac{q(a)K(o,r \mid a)}{p_{qK}(o,r)}, \qquad \mu_{qK} := \sum_{o,r} p_{qK}(o,r)\, \delta_{\pi^{qK}_{or}},$$

where zero-probability terms are omitted.

B.1 Consequences of the representation

Corollary 29 (No payoff–trace complementarity)**.** *Suppose Axioms* (A1)–(A8) *hold. Fix A and $q \in \Delta(A)$. If channels K and L satisfy $E_{qK} = E_{qL}$ and $\mu_{qK} = \mu_{qL}$, then $(q,K) \sim (q,L)$.*

Proof. The result follows from Theorem 1 and $I_{qK}(A;O,R) = H(q) - \int H(\pi) d\mu_{qK}(\pi)$. □

Proof of Proposition 2. Range. Choose $o_* \in O$ with $u(o_*) = \underline{u}$. The range of V is $[\underline{u}, \infty)$. The lower bound is immediate. Let a constant-o_* channel reveal the realised action with probability t and otherwise report $*$:

$$E_t(o_*, a' \mid a) = t\mathbf{1}_{a'=a}, \qquad E_t(o_*, * \mid a) = 1 - t.$$

At the uniform lottery on n actions, $V(q, E_t) = \underline{u} + \lambda t \log n$ fills $[\underline{u}, \underline{u} + \lambda \log n]$; these intervals are unbounded.

Ordinal uniqueness. Let W be common-scale and define $\varphi(v) := W(q,K)$ for any pair with $V(q,K) = v$. This is well defined: equality of the two V-values gives indifference in their two-block environment by Theorem 1, and the common-scale property gives equality of their W-values; strict inequalities are preserved in the same way. Thus $W = \varphi \circ V$ for a strictly increasing φ. The converse follows from Theorem 1, proving part (i).

Affine uniqueness. The two chain rules give, for every $P : A \to \Delta(\{1,\ldots,k\})$ and every $K_1,\ldots,K_k, L_1,\ldots,L_k : A \to \Delta(O \times R)$, writing $m_i := \sum_a q(a)P(i \mid a)$,

$$V(q, P \triangleright (K_1,\ldots,K_k)) - V(q, P \triangleright (L_1,\ldots,L_k)) = \sum_{i:m_i>0} m_i [V(q_i,K_i) - V(q_i,L_i)].$$

Hence every $\alpha V + \beta$, $\alpha > 0$, is common-scale and branch-additive.

Conversely, write a branch-additive common-scale representation as $W = \varphi \circ V$ and put

$$\psi(x) := \varphi(\underline{u} + x) - \varphi(\underline{u}), \qquad x \geq 0.$$

For $x \in [0, \lambda \log n]$, use the range construction after the first branch of a public coin with probabilities θ and $1-\theta$, and the uninformative constant-o_* channel elsewhere. Comparing this compound with the uninformative compound in (6) gives

$$\psi(\theta x) = \theta \psi(x) \qquad (0 \leq \theta \leq 1).$$

Taking $x = \lambda \log n$ shows that ψ is linear on $[0, \lambda \log n]$. These intervals are nested and unbounded, so $\psi(x) = \alpha x$ on $[0,\infty)$ for one $\alpha > 0$. Therefore $\varphi(v) = \alpha v + \beta$, proving part (ii).

Parameters. Any second trace-tempered index $\widetilde{V}$ is common-scale by block invariance and branch-additive by the same chain rules. Part (ii), restricted first to deterministic payoffs and then to constant-payoff record channels, gives $\widetilde{u} = \alpha u + \beta$ and $\widetilde{\lambda} = \alpha\lambda$. □

Proof of Corollary 3. Part (i) is Theorem 1. For part (ii), reverse the primitive relation in every environment. The reversed family satisfies Axioms (A1), (A2), (A5), and (A8); Axiom (A3) holds with its two witnessing outcomes interchanged. It also satisfies the forward versions of Axioms (A4), (A6), and (A7). Theorem 1 therefore represents the reversed family by $\mathbb{E}[\widehat{u}(O)] + \widehat{\lambda} I(A; O, R)$ with $\widehat{\lambda} > 0$. Negating that index represents the original family, so it has the asserted form with $u = -\widehat{u}$ and $\lambda = -\widehat{\lambda} < 0$.

For part (iii), collapse the action alphabet of any pair (q, K) to a singleton. Axiom (A7^0) makes the pair indifferent to the resulting one-action channel, whose output law is p_{qK}. Axiom (A6^0) then erases its record without changing value. Thus every pair is indifferent to the one-action channel that delivers precisely its payoff marginal. Axiom (A5) makes these reductions coherent across blocks. On payoff lotteries, Axiom (A8) applied to an action-independent public coin gives mixture independence, Axiom (A2) gives continuity, and Axiom (A3) gives non-triviality. The mixture-space theorem therefore yields a nonconstant expected-utility index. Conversely, expected utility is unaffected by record or action processing and satisfies the three indifference clauses. The positive and negative formulae satisfy their respective clauses by data processing and the two chain rules. □

B.2 Independence of the axioms

Proof of Proposition 4. Fix nonconstant u and write $U(q, K) := \mathbb{E}_{qK}[u(O)]$, $Z := (O, R)$, and $I(q, K) := I_{qK}(A; Z)$. All scores below use the same formula in block and compound environments. Both coordinates are block invariant and processing monotone; under a common-branch replacement their differences are multiplied by the reached-branch probability. Moreover, U distinguishes the material test and I takes values $\log 2$ and zero in the trace test. For witnesses built only from (U, I), these facts verify every unmentioned axiom below.

Failure of Axiom (A1)*.* Use the Pareto relation on the two coordinates (U, I):

$$(q, K) \succeq (p, L) \quad \Longleftrightarrow \quad U(q, K) \geq U(p, L) \text{ and } I(q, K) \geq I(p, L).$$

It is transitive with closed graph, but incomplete whenever payoff and information rank two alternatives oppositely.

Failure of Axiom (A2)*.* Order alternatives lexicographically by (U, I), with payoff first. This is a weak order but is not continuous even within one fixed channel. Let a three-action channel reveal the action fully and let only action 2 pay the high outcome. Put

$$p = (1/4, 1/2, 1/4), \qquad q = (0.49, 0.50, 0.01), \qquad q_n = (0.49 - \varepsilon_n, 0.50 + \varepsilon_n, 0.01),$$

where $\varepsilon_n \downarrow 0$. The lotteries p and q have the same expected utility but $H(q) < H(p)$, so $p \succ_K q$. Every q_n has strictly higher expected utility than p, so $q_n \succ_K p$. Thus the graph in Axiom (A2) is not closed.

Failure of Axioms (A3) *and* (A4)*.* For Axiom (A3), use $V := I$: pure lotteries carry no information, although the trace test has values $\log 2$ and zero. For Axiom (A4), use $V := U$: suitable outcomes have different utility, but the two trace-test lotteries have the same payoff.

Failure of Axiom (A5)*.* The idea is to add a linear-in-the-prior bonus that vanishes in every canonical two-block comparison but becomes active in a three-block environment. Define the weights directly from the literal labels of each finite action alphabet A. If every element of A is literally a pair (x, i) and the set of outermost second coordinates is $\{0, \ldots, m-1\}$ for some $m \geq 3$, put

$$w_A(x, i) := \mathbf{1}_{\{i=0\}}.$$

Otherwise put $w_A = 0$. In a nested tagged sum this convention reads only the literal outermost coordinate, so it is a single-valued function of the actual action set and not of its construction history. For $\varepsilon > 0$, use within each environment

$$V_\varepsilon(q, K) := U(q, K) + I(q, K) + \varepsilon \sum_a q(a) w_A(a).$$

This is a continuous weak order. The material- and trace-relevance tests can use action sets on which $w_A = 0$, so Axioms (A3) and (A4) retain their usual witnesses. Every comparison in Axioms (A6)–(A8) is evaluated in a canonical outer two-block environment; there the outermost tag set is $\{0, 1\}$, even when a component alphabet is itself tagged. Hence $w_A = 0$ throughout those comparisons and the score reduces to $U + I$. For three identical one-action, constant-payoff, silent channels, however, blocks 0 and 1 are indifferent in their canonical two-block comparison but differ by ε in the three-block environment. Thus Axiom (A5) fails.

Failure of Axiom (A6). For $\varepsilon > 0$, use

$$V(q, K) := U(q, K) + I(q, K) - \varepsilon H_{qK}(Z \mid A). \tag{53}$$

Conditional entropy is continuous and block invariant, and mutual information and conditional entropy both obey the branch chain rule. Under action processing $A \to B$, the original-minus-processed difference is

$$\big[I(A; Z) - I(B; Z)\big] + \varepsilon\big[H(Z \mid B) - H(Z \mid A)\big] = (1 + \varepsilon)I(A; Z \mid B) \geq 0,$$

so Axiom (A7) holds. In the trace test the two pairs $(I(q, K), H(Z \mid A))$ are $(\log 2, 0)$ and $(0, \log 2)$. But with constant payoff and a fair record independent of A, the score is $-\varepsilon \log 2$; deleting that noise raises it to zero, violating Axiom (A6).

Failure of Axiom (A7). Let $G(q) := 1 - \sum_a q(a)^2$ be Gini entropy and define its posterior reduction

$$J_G(q, K) := G(q) - \int G(\pi)\, d\mu_{qK}(\pi), \qquad V := U + J_G.$$

The score J_G is continuous and block invariant. Concavity of G gives record processing, and the potential-difference form gives the branch chain rule; the trace-test values are $1/2$ and zero. Action processing can nevertheless increase J_G. Take $q = (1/8, 1/8, 3/4)$, a constant payoff, and a binary record with

$$\Pr(0 \mid a_1) = \Pr(0 \mid a_2) = 0, \qquad \Pr(0 \mid a_3) = 1/4,$$

and merge a_1, a_2. Direct calculation gives $J_G = 9/416$ before the merge and $J_G = 3/104$ afterwards, violating Axiom (A7).

Failure of Axiom (A8). The idea is that a cap preserves processing monotonicity but makes a branch's value depend on how much information has already accumulated elsewhere. Use $V(q, K) := U(q, K) + \min\{I(q, K), 1\}$ in bits. The cap preserves continuity, block coherence, processing, and both relevance tests. Let q be uniform on four actions and let a first-stage record reveal which of two pairs contains the action. Keep payoffs constant and one continuation silent; after the other record, compare a silent continuation with one that reveals the action within its pair. The latter is strictly preferred at the branch posterior, but the silent and revealing compounds have one and three-halves bits, respectively; both are capped at one and are therefore indifferent. Taking K silent and L revealing in (3), the compound comparison is true by indifference whereas $(q_1, K) \not\succsim (q_1, L)$; the biconditional fails. □

B.3 Choice implications

The next four blocks mirror Section 4. Write $Z := O \times R$; when $K : A \to \Delta(Z)$ is fixed, put $K_a := K(\cdot \mid a)$ and $\bar{u}(a) := \mathbb{E}_{K_a}[u(O)]$; its output law under q is p_{qK}.

B.3.1 The trace test

Proof of Proposition 5. (i) Under either lottery, each visible pair (o_0, r_j), $j = 1, 2$, has probability $\frac{1}{2}$: under q' because each a_j produces (o_0, r_j) for sure, under q'' because each of a_3, a_4 produces either pair with equal probability. Hence $p_{q'K} = p_{q''K}$. On $\mathrm{supp}(q')$ the rows are degenerate and distinct, so the visible pair reveals the action and $I_{q'K} = H(q') = \log 2$; on $\mathrm{supp}(q'')$ the rows are identical, so the visible pair is independent of the action and $I_{q''K} = 0$. The payoff terms agree because the payoff marginals do, giving the displayed gap. (ii) All rows of K° are equal, so $I_{qK^\circ} = 0$ for every q; since the payoff is constantly o_0, the representation gives indifference. $\square$

B.3.2 Deliberate randomisation

Proposition 30 (Optimal mixing)**.** *Let $\lambda > 0$ and $A = \{a, b\}$ with $K_a \neq K_b$ and $\Delta := \bar{u}(a) - \bar{u}(b) \geq 0$. For $x \in [0,1]$ let q_x put mass x on b, so that $p_{q_x K} = (1-x)K_a + xK_b =: p_x$, and for $x \in (0,1)$ define*

$$g(x) := D_{\mathrm{KL}}(K_b \,\|\, p_x) - D_{\mathrm{KL}}(K_a \,\|\, p_x)\,.$$

(i) *g is continuous and strictly decreasing, with $g(0^+) = D_{\mathrm{KL}}(K_b \,\|\, K_a)$ and $g(1^-) = -D_{\mathrm{KL}}(K_a \,\|\, K_b)$ as extended-real limits. The pure lottery δ_a is optimal if and only if $\Delta \geq \lambda D_{\mathrm{KL}}(K_b \,\|\, K_a)$. Below this threshold the optimum is unique and interior, $q^* = q_{x^*}$ with x^* the unique solution of $\Delta = \lambda g(x)$.*

(ii) *The optimal mass $x^*(\Delta, \lambda)$ on the inferior action is nonincreasing in Δ and nondecreasing in λ. While the optimum is interior, it is strictly decreasing in Δ and, when $\Delta > 0$, strictly increasing in λ. At $\Delta = 0$ it is the capacity-achieving input of the two-row channel, independently of λ.*

Proof. Write

$$f(x) := V(q_x, K) = \bar{u}(a) - x\Delta + \lambda I(x), \qquad I(x) := H(p_x) - (1-x)H(K_a) - xH(K_b).$$

For $x \in (0,1)$, differentiation gives $I'(x) = g(x)$ and

$$g'(x) = -c_{\log} \sum_{z \in Z} \frac{(K_b(z) - K_a(z))^2}{p_x(z)} < 0,$$

where $c_{\log} = 1/\ln b_0 > 0$ for the fixed logarithm base b_0 and zero summands are omitted. Thus $f'(x) = -\Delta + \lambda g(x)$ is strictly decreasing. As $x \downarrow 0$, finite summands converge to those in $D_{\mathrm{KL}}(K_b \| K_a)$, while any z with $K_b(z) > 0 = K_a(z)$ makes the limit $+\infty$; exchanging a and b gives the limit at one. Consequently $x = 0$ is optimal exactly when $\Delta \geq \lambda D_{\mathrm{KL}}(K_b \| K_a)$. Otherwise $f'(0^+) > 0$ and $f'(1^-) = -\Delta - \lambda D_{\mathrm{KL}}(K_a \| K_b) < 0$, so the unique optimum is the root $\Delta = \lambda g(x)$. At an interior optimum, $x^* = g^{-1}(\Delta/\lambda)$; since g^{-1} is strictly decreasing, this gives strict decrease in Δ and, when $\Delta > 0$, strict increase in λ, together with the stated weak comparative statics at the boundary. At $\Delta = 0$, $g(x) = 0$ is the usual capacity condition and is independent of λ. $\square$

General optimality. For every q, the entropy identity gives

$$\begin{aligned} V(q, K) &= \lambda H(p_{qK}) + \sum_a q(a)\left[\bar{u}(a) - \lambda H(K_a)\right] \\ &= \sum_{a \in \mathrm{supp}(q)} q(a)\left[\bar{u}(a) + \lambda D_{\mathrm{KL}}(K_a \,\|\, p_{qK})\right]. \end{aligned} \tag{54}$$

Define

$$\Phi_a(p) := \bar{u}(a) + \lambda D_{\mathrm{KL}}(K_a \,\|\, p)\,.$$

For $q_t := (1-t)q + t\delta_a$, direct differentiation in the extended-real sense gives

$$\left.\frac{d}{dt} V(q_t, K)\right|_{t=0^+} = \Phi_a(p_{qK}) - V(q, K). \tag{55}$$

Proof of Proposition 6. Expected utility is linear in q, while mutual information is concave in the input distribution [15]; continuity on the compact simplex gives existence. At an optimum q^*, every vertex direction in (55) is nonpositive, so $\Phi_a(p_{q^*K}) \le V(q^*, K)$ for all a. Equation (54) says their q^*-weighted average is $V(q^*, K)$; hence equality holds on $\operatorname{supp}(q^*)$.

Conversely, the proposition's equalities and inequalities, averaged under q^*, give $c = V(q^*, K)$. The derivative toward any lottery r is

$$\sum_a r(a)\,[\Phi_a(p_{q^*K}) - c] \le 0.$$

Concavity makes this condition sufficient for global optimality. At constant u it is the classical capacity condition. □

Proposition 31 (Support and pure choice)**.** *Let $\lambda > 0$ and let $Z_K := \{z \in Z : K_a(z) > 0$ for some $a \in A\}$.*

(i) *Every optimal lottery q^* satisfies $p_{q^*K}(z) > 0$ for every $z \in Z_K$. Consequently, if every action a has a private consequence—some z_a with $K_a(z_a) > 0$ and $K_b(z_a) = 0$ for all $b \ne a$—then every optimal lottery has full support.*

(ii) *The pure lottery δ_a is optimal if and only if*

$$\bar{u}(b) + \lambda D_{\mathrm{KL}}(K_b \,\|\, K_a) \le \bar{u}(a) \qquad \text{for every } b \in A,$$

and uniquely optimal if these inequalities are strict for $b \ne a$. In particular, if some rival reaches a consequence outside the support of K_a, then δ_a is not optimal, however large its payoff advantage.

Proof. (i) Put $p^* := p_{q^*K}$. If $p^*(z) = 0 < K_a(z)$, then $D_{\mathrm{KL}}(K_a \| p^*) = +\infty$, contradicting either the on-support equality or the off-support inequality in Proposition 6. If action a has a private consequence but $q^*(a) = 0$, that consequence has zero probability under p^*, contradicting the first claim.

(ii) At $q = \delta_a$ the output law is K_a and the on-support value is $\bar{u}(a)$. Proposition 6 therefore gives exactly the displayed inequalities. If they are strict off a, every nonzero direction from δ_a has strictly negative directional derivative, so concavity makes δ_a the unique maximiser. If K_b is not supported by K_a, then $D_{\mathrm{KL}}(K_b \| K_a) = +\infty$, so the displayed condition fails. □

Proof of Corollary 7. Pairwise-disjoint row supports make the visible consequence identify the action, so $I_{qK}(A; Z) = H(q)$; they also give every action a private consequence, so Proposition 31(i) gives full support at the optimum. On $\operatorname{supp} K_a$, $p_{qK}(z) = q(a)K_a(z)$. Hence $D_{\mathrm{KL}}(K_a \| p_{qK}) = -\log q(a)$. The equalities in Proposition 6 give $\bar{u}(a) - \lambda \log q^*(a) = c$. In natural-log units, solving and normalising yields the displayed logit formula. Strict concavity of $\sum_a q(a)\bar{u}(a) + \lambda H(q)$ gives uniqueness; changing the logarithm base merely rescales λ. □

Proposition 32 (Optimal marginal and multiplicity)**.** *Let $\lambda > 0$ and write $C^*(K) := \max_{q \in \Delta(A)} V(q, K)$. There is a unique $p^* \in \Delta(Z)$ such that $p_{q^*K} = p^*$ for every optimal lottery q^*. With*

$$\mathcal{A}^* := \{a \in A : \bar{u}(a) + \lambda D_{\mathrm{KL}}(K_a \,\|\, p^*) = C^*(K)\},$$

the set of optimal lotteries is exactly

$$\{q \in \Delta(A) : p_{qK} = p^*,\ \operatorname{supp}(q) \subseteq \mathcal{A}^*\}.$$

The optimum is unique whenever the rows $\{K_a : a \in \mathcal{A}^*\}$ are affinely independent. If actions are partitioned into classes with identical rows, $V(q, K)$ depends on q within each class only through the class's total mass. The optimal lotteries are exactly the lifts of the optimal lotteries of the collapsed channel, and each class's mass may be distributed arbitrarily among its members.

Proof of Proposition 32. If optimal q_0, q_1 induced different output laws, the first line of (54) would be strictly concave along their segment while its remaining term is linear, producing a value above $C^*(K)$. Thus every optimum induces the same p^*. Proposition 6 and (54) then show that every active score equals $C^*(K)$, so $\mathrm{supp}(q^*) \subseteq \mathcal{A}^*$. Conversely, if $p_{qK} = p^*$ and $\mathrm{supp}(q) \subseteq \mathcal{A}^*$, then $V(q, K) = \sum_a q(a)\Phi_a(p^*) = C^*(K)$. Affine independence makes the representation $p^* = \sum_a q(a)K_a$ with $\mathrm{supp}(q) \subseteq \mathcal{A}^*$ unique. Identical rows share $\bar{u}(a)$ and $H(K_a)$, so both p_{qK} and (54) are unchanged by redistribution within a class; collapsing the classes and lifting aggregate masses proves the last claim. □

Remark 33 (Overlapping traces and IIA)**.** Let three actions have equal expected utility and rows

$$K_1 = (1, 0, 0), \qquad K_2 = (0, 1, 0), \qquad K_3 = (1/2, 0, 1/2).$$

With only the first two actions, the optimal lottery is $(1/2, 1/2)$. Adding the third gives

$$q^* = (1/3, 4/9, 2/9), \qquad p_{q^*K} = (4/9, 4/9, 1/9).$$

Each active row has divergence $\log(9/4)$ from p_{q^*K}, so Proposition 6 verifies optimality. The ratio of the first two choice probabilities falls from 1 to 3/4. Thus overlapping rows can generate the red-bus/blue-bus pattern even when expected utilities are equal.

B.3.3 The value of being on record

Proof of Proposition 8. The processor leaves the payoff coordinate unchanged, so the expected-utility terms coincide and the difference is $\lambda[I_{qK}(A; O, R) - I_{qL}(A; O, R')]$. Write $Z := (O, R)$ and $Z' := (O, R')$. By construction $A \to Z \to Z'$ is a Markov chain, so expanding $I(A; Z, Z')$ in both orders gives

$$I(A; Z) - I(A; Z') = I(A; Z \mid Z') = I(A; R \mid O, R'),$$

the last equality because O is part of Z'. Conditional mutual information is nonnegative and vanishes exactly when A and (O, R) are conditionally independent given (O, R') [15]. This is the zero-price condition stated in the proposition. Under a garbling, it is also equivalent to equality of the posterior laws, $\mu_{qK} = \mu_{qL}$. Indeed, the Markov property gives $\Pr(A = \cdot \mid Z, Z') = \pi_Z^{qK}$ almost surely, while conditional independence gives $\Pr(A = \cdot \mid Z, Z') = \pi_{Z'}^{qL}$; the two random posteriors then coincide almost surely. Conversely, by the identity in the proof of Corollary 29, equality of the posterior laws gives equality of the information terms, and the payoff terms already agree. □

Corollary 34 (The recorded coin)**.** *Let $K, L : A \to \Delta(O \times R)$ and $t \in (0, 1)$. Let $tK \oplus (1-t)L$ be the public mixture that draws an action-independent coin Y with $\Pr(Y = 0) = t$, applies K if $Y = 0$ and L otherwise, and includes the coin in the record, so the full public-mixture record is (Y, R); let $M_t := tK + (1-t)L$ be the rowwise mixture, which hides it. Then, for every $q \in \Delta(A)$,*

(i) $V(q,\, tK \oplus (1-t)L) = t\, V(q, K) + (1-t)\, V(q, L)$*;*

(ii) $V(q,\, tK \oplus (1-t)L) - V(q, M_t) = \lambda\, I(A; Y \mid O, R) \geq 0$*, with equality if and only if A and Y are conditionally independent given (O, R).*

Proof of Corollary 34. Part (i) is the chain rule with an action-independent first stage: $I(A; Y) = 0$, and each reached branch reproduces the joint law of the corresponding constituent. Part (ii) follows from Proposition 8: deleting the coin coordinate is a deterministic payoff-preserving record processor carrying the public mixture to M_t. □

Proposition 35 (Uniform dominance is more-capable, not garbling)**.** *Let $K : A \to \Delta(O \times R)$ and $L : A \to \Delta(O \times R')$ have identical payoff kernels. Write $K \succeq_g L$ if $L = KT$ for some record processor T, and write $K \succeq_c L$ if $I_{qK}(A; O, R) \geq I_{qL}(A; O, R')$ for every $q \in \Delta(A)$. The latter is the* more-capable *order [28].*

(i) *$K \succeq_c L$ if and only if $V(q,K) \geq V(q,L)$ for every $q \in \Delta(A)$.*

(ii) *$K \succeq_g L \Rightarrow K \succeq_c L$, but the converse fails. There are channels for which $K \succeq_c L$ although $L \neq KT$ for every record processor T.*

Proof of Proposition 35. (i) Because the payoff kernels coincide, $V(q,K)-V(q,L) = \lambda[I_{qK}-I_{qL}]$ with $\lambda > 0$. Uniform value dominance and $K \succeq_c L$ are therefore the same statement.

(ii) If $L = KT$, Proposition 8 gives $I_{qK} \geq I_{qL}$ at every q, and hence $K \succeq_g L \Rightarrow K \succeq_c L$. For failure of the converse, let $A = \{0,1\}$, fix $\bar{o} \in O$, and let both channels pay $\bar{o}$ surely. Let K carry the binary erasure record with erasure probability $\varepsilon = \frac{1}{2}$ and L the binary symmetric record with crossover $p = \frac{1}{5}$:

$$K(\bar{o}, a \mid a) = 1-\varepsilon, \quad K(\bar{o}, \mathrm{e} \mid a) = \varepsilon, \qquad L(\bar{o}, a \mid a) = 1-p, \quad L(\bar{o}, 1-a \mid a) = p.$$

The payoff kernels agree row by row, so it suffices to compare informations, and the comparison is invariant to the base; compute in bits with h the binary entropy. For $q = (1-t, t)$, grouping records gives

$$I_{qK} = (1-\varepsilon)\, h(t), \qquad I_{qL} = h(p * t) - h(p), \qquad p * t := p(1-t) + (1-p)t.$$

By Mrs Gerber's Lemma [39], $v \mapsto h(p * h^{-1}(v))$ is convex on $[0,1]$, with h^{-1} valued in $[0, \frac{1}{2}]$; since $h(p*t)$ is invariant under $t \mapsto 1-t$, take $t \leq \frac{1}{2}$ and $v = h(t)$. The chord between $v = 0$ and $v = 1$ gives

$$h(p*t) \leq (1-v)\, h(p) + v, \qquad \text{hence} \qquad I_{qL} \leq (1-h(p))h(t) \leq (1-\varepsilon)\, h(t) = I_{qK},$$

the last step because $\varepsilon = \frac{1}{2} \leq h(\frac{1}{5})$. Thus $K \succeq_c L$. Now suppose $L = KT$ for a processor T, and write $t_r := T(0 \mid \bar{o}, r)$ for $r \in \{0, 1, \mathrm{e}\}$. Matching the probability of the processed record 0 in the two rows,

$$(1-\varepsilon)t_0 + \varepsilon t_{\mathrm{e}} = 1-p, \qquad (1-\varepsilon)t_1 + \varepsilon t_{\mathrm{e}} = p,$$

and subtracting gives $t_0 - t_1 = (1-2p)/(1-\varepsilon) = \frac{6}{5} > 1$, impossible. Compositions of record processors are record processors, so no chain of garblings produces L from K either. □

B.3.4 Measuring the trace weight

Proposition 36 (Single information price)**.** *If $(q,K) \sim (p,L)$ and $I_{qK} \neq I_{pL}$, then*

$$\frac{E_{pL} - E_{qK}}{I_{qK} - I_{pL}} = \lambda,$$

and λ is the only number satisfying this identity at every such indifference. More generally, writing

$$\phi(q,K) := \big(I_{qK}(A; O, R), E_{qK}\big), \qquad \underline{u} := \min_o u(o), \quad \overline{u} := \max_o u(o),$$

the image of ϕ over all finite pairs is $[0,\infty) \times [\underline{u}, \overline{u}]$. In these coordinates the block-comparison indifference classes are the nonempty intersections of the lines $E = c - \lambda I$ with that strip.

Proof. The ratio identity rearranges $E_{qK} + \lambda I_{qK} = E_{pL} + \lambda I_{pL}$. For the image claim, only the reverse inclusion needs proof. Given (t,w) in the strip, choose n with $\log n \geq t$ and, by continuity of entropy between a point mass and the uniform lottery, choose q_t on $\{1,\dots,n\}$ with $H(q_t) = t$. Choose $o^+, o^- \in O$ with $u(o^+) = \overline{u}$ and $u(o^-) = \underline{u}$. Let a channel reveal the action in its record and, independently of the action, pay o^+ with probability β and o^- otherwise, where $\beta u(o^+) + (1-\beta)u(o^-) = w$. It realises $\phi(q_t, K) = (t, w)$, and Theorem 1 gives the stated lines.

For uniqueness, if $\lambda' \neq \lambda$, put $t_0 := \frac{1}{2}\min\{1, (\overline{u} - \underline{u})/\lambda\} > 0$. The image claim supplies pairs with coordinates $(t_0, \overline{u} - \lambda t_0)$ and $(0, \overline{u})$. They are indifferent, whereas their $E + \lambda' I$ values differ by $(\lambda' - \lambda)t_0 \neq 0$. □

Proposition 37 (Willingness to pay for trace)**.** *Let o^+, o^- attain the maximum and minimum of u, and put*

$$\Delta u := u(o^+) - u(o^-) > 0, \qquad w(\beta) := \beta u(o^+) + (1-\beta)u(o^-).$$

For a channel $H : C \to \Delta(O \times T)$, zero-pad it to $T^* := T \sqcup \{*\}$. For $\varepsilon \in [0,1)$ and $\beta \in [0,1]$, let $P_\varepsilon : C \to \Delta(\{1,2\})$ satisfy $P_\varepsilon(2 \mid c) = \varepsilon$ for every c, and let B_β have constant rows

$$B_\beta(o^+, * \mid c) = \beta, \qquad B_\beta(o^-, * \mid c) = 1 - \beta.$$

Write $D_{\varepsilon,\beta}H := P_\varepsilon \triangleright (H, B_\beta)$. Fix $q \in \Delta(A)$, $p \in \Delta(B)$, and channels $K : A \to \Delta(O \times R)$ and $L : B \to \Delta(O \times S)$.

(i) *For every $\varepsilon \in [0,1)$ and $\beta \in [0,1]$,*

$$V(q,\, D_{\varepsilon,\beta}K) = (1-\varepsilon)\, V(q,K) + \varepsilon\, w(\beta).$$

(ii) *If $\varepsilon \in (0,1)$ and $\beta, \beta' \in [0,1]$ satisfy $(q, D_{\varepsilon,\beta}K) \sim (p, D_{\varepsilon,\beta'}L)$, then*

$$\frac{\varepsilon}{1-\varepsilon}\,[w(\beta) - w(\beta')] = V(p,L) - V(q,K).$$

When the material parts agree, $E_{qK} = E_{pL}$, the left-hand side equals $\lambda[I_{pL}(B;O,S) - I_{qK}(A;O,R)]$: information is priced linearly at rate λ.

(iii) *Such a pair (β, β') exists if and only if $|V(p,L) - V(q,K)| \le \frac{\varepsilon}{1-\varepsilon}\Delta u$; a boundary response therefore brackets the value difference instead of measuring it.*

Proof. (i) The material part of the compound is $(1-\varepsilon)\mathbb{E}_{qK}[u(O)] + \varepsilon w(\beta)$. For the trace part, the coin is action-independent, so both branch posteriors equal q and the chain rule for the compound gives an information term of $(1-\varepsilon)I_{qK}(A;O,R)$, the side-bet branch contributing zero. (ii) By the moreover clause of Theorem 1, the indifference holds if and only if $(1-\varepsilon)V(q,K) + \varepsilon w(\beta) = (1-\varepsilon)V(p,L) + \varepsilon w(\beta')$; rearrange, and substitute the representation when the material parts cancel. (iii) $w(\beta) - w(\beta')$ is continuous on $[0,1]^2$ with range $[-\Delta u, \Delta u]$; apply the intermediate value theorem to the equation in (ii). □

Proof of Proposition 9. Write K^{id} and K^0 for the revealing and the silent channel at the constant payoff, so that $V(q,K^{\mathrm{id}}) - V(q,K^0) = \lambda H(q)$. By Proposition 37(i), the two diluted values are $(1-\varepsilon)V(q,K^{\mathrm{id}}) + \varepsilon w(0)$ and $(1-\varepsilon)V(q,K^0) + \varepsilon w(\beta)$, and under the normalisation $w(\beta) = \beta$. Indifference therefore holds if and only if $\varepsilon\beta = (1-\varepsilon)\lambda H(q)$. The left side is continuous and strictly increasing in β with range $[0,\varepsilon]$, so a solution exists, and is then unique, exactly when $\lambda H(q) \le \varepsilon/(1-\varepsilon)$; when this fails, the revealing side's value exceeds the silent side's at every β. □

Remark 38 (Implementation and controls)**.** Normalise $u(o^+) = 1$ and $u(o^-) = 0$. All comparisons below are primitive block comparisons. For each $o \in O$, elicit β_o from $(\delta_*, K_o) \sim (\delta_*, B_{\beta_o})$, where K_o pays o surely; both one-action channels carry zero information, so $u(o) = \beta_o$. Next take the uniform q on $n \ge 2$ actions and the constant-payoff revealing and silent channels K_o^{id} and K_o^0, and elicit β from

$$(q, D_{\varepsilon,\beta}K_o^0) \sim (q, D_{\varepsilon,0}K_o^{\mathrm{id}}), \qquad \lambda = \frac{\varepsilon}{1-\varepsilon}\frac{\beta}{\log n}.$$

If even $\beta = 1$ does not produce indifference, the response gives the lower bound $\lambda > \varepsilon/((1-\varepsilon)\log n)$, and ε is raised.

Three controls are essential. *Orientation:* revelation must beat silence at every constant payoff. *Garbling:* in a separate comparison of the undiluted, equal-payoff channels K_o^{id} and K_o^0, total record erasure must make the two blocks indifferent; otherwise some display feature still traces the action. *Transport:* changing the prior, payoff, or action-set size must return the same λ by Proposition 36; a decline with $\log n$ instead signals the capped-information criterion.

References

[1] János Aczél and Zoltán Daróczy. *On Measures of Information and Their Characterizations.* Academic Press, New York, 1975.

[2] Marina Agranov and Pietro Ortoleva. Stochastic choice and preferences for randomization. *Journal of Political Economy*, 125(1):40–68, 2017.

[3] James Andreoni and B. Douglas Bernheim. Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636, 2009.

[4] Yaron Azrieli and Ehud Lehrer. The value of a stochastic information structure. *Games and Economic Behavior*, 63(2):679–693, 2008.

[5] Björn Bartling, Ernst Fehr, and Holger Herz. The intrinsic value of decision rights. *Econometrica*, 82(6):2005–2039, 2014.

[6] John C. Baez, Tobias Fritz, and Tom Leinster. A characterization of entropy in terms of information loss. *Entropy*, 13(11):1945–1957, 2011.

[7] Roland Bénabou and Jean Tirole. Self-confidence and personal motivation. *Quarterly Journal of Economics*, 117(3):871–915, 2002.

[8] Roland Bénabou and Jean Tirole. Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678, 2006.

[9] B. Douglas Bernheim. A theory of conformity. *Journal of Political Economy*, 102(5):841–877, 1994.

[10] David Blackwell. Equivalent comparisons of experiments. *Annals of Mathematical Statistics*, 24(2):265–272, 1953.

[11] Antonio Cabrales, Olivier Gossner, and Roberto Serrano. Entropy and the value of information for investors. *American Economic Review*, 103(1):360–377, 2013.

[12] Andrew Caplin, Mark Dean, and John Leahy. Rationally inattentive behavior: Characterizing and generalizing Shannon entropy. *Journal of Political Economy*, 130(6):1676–1715, 2022.

[13] Andrew Caplin and John Leahy. Psychological expected utility theory and anticipatory feelings. *Quarterly Journal of Economics*, 116(1):55–79, 2001.

[14] Simone Cerreia-Vioglio, David Dillenberger, Pietro Ortoleva, and Gil Riella. Deliberately stochastic. *American Economic Review*, 109(7):2425–2445, 2019.

[15] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley-Interscience, Hoboken, second edition, 2006.

[16] Imre Csiszár. Axiomatic characterizations of information measures. *Entropy*, 10(3):261–273, 2008.

[17] Tommaso Denti. Posterior separable cost of information. *American Economic Review*, 112(10):3215–3259, 2022.

[18] Fyodor Dostoevsky. *Notes from the Underground*, Part I. 1864. Translated by Constance Garnett, 1918.

[19] Jeffrey Ely, Alexander Frankel, and Emir Kamenica. Suspense and surprise. *Journal of Political Economy*, 123(1):215–260, 2015.

[20] D. K. Faddeev. On the concept of entropy of a finite probabilistic scheme. *Uspekhi Matematicheskikh Nauk*, 11(1):227–231, 1956. In Russian.

[21] Alexander Frankel and Emir Kamenica. Quantifying information and uncertainty. *American Economic Review*, 109(10):3650–3680, 2019.

[22] Itzhak Gilboa and Ehud Lehrer. The value of information—an axiomatic approach. *Journal of Mathematical Economics*, 20(5):443–459, 1991.

[23] Amihai Glazer and Kai A. Konrad. A signaling explanation for charity. *American Economic Review*, 86(4):1019–1028, 1996.

[24] Simon Grant, Atsushi Kajii, and Ben Polak. Intrinsic preference for information. *Journal of Economic Theory*, 83(2):233–259, 1998.

[25] I. N. Herstein and John Milnor. An axiomatic approach to measurable utility. *Econometrica*, 21(2):291–297, 1953.

[26] Aleksandr I. Khinchin. *Mathematical Foundations of Information Theory*. Dover, New York, 1957.

[27] Alexander S. Klyubin, Daniel Polani, and Chrystopher L. Nehaniv. Empowerment: A universal agent-centric measure of control. In *Proceedings of the 2005 IEEE Congress on Evolutionary Computation*, volume 1, pages 128–135. IEEE, 2005.

[28] János Körner and Katalin Marton. Comparison of two noisy channels. In Imre Csiszár and Peter Elias, editors, *Topics in Information Theory*, Colloquia Mathematica Societatis János Bolyai, volume 16, pages 411–423. North-Holland, Amsterdam, 1977.

[29] Filip Matějka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298, 2015.

[30] Prachi Mistry and Mimi Liljeholm. Instrumental divergence and the value of control. *Scientific Reports*, 6:36295, 2016.

[31] Kaitlyn G. Norton and Mimi Liljeholm. The rostrolateral prefrontal cortex mediates a preference for high-agency environments. *Journal of Neuroscience*, 40(22):4401–4409, 2020.

[32] Luciano Pomatto, Philipp Strack, and Omer Tamuz. The cost of information: The case of constant marginal costs. *American Economic Review*, 113(5):1360–1393, 2023.

[33] Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 547–561. University of California Press, 1961.

[34] Christoph Salge, Cornelius Glackin, and Daniel Polani. Empowerment—an introduction. In Mikhail Prokopenko, editor, *Guided Self-Organization: Inception*, pages 67–114. Springer, Berlin, 2014.

[35] Leonard J. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.

[36] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[37] Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003.

[38] Michael Spence. Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374, 1973.

[39] Aaron D. Wyner and Jacob Ziv. A theorem on the entropy of certain binary sequences and applications: Part I. *IEEE Transactions on Information Theory*, 19(6):769–772, 1973.