Lossless Projection

Factorization, Fibers, and Recoverability

Flyxion
Independent Researcher

November 2026

Abstract

A projection from a set of states to a set of descriptions may identify many states and still preserve everything needed to answer a given question. This paper makes that observation exact. For maps $p : X \to Y$ and $q : X \to Z$, the projection p is called lossless relative to q when q factors through p. The paper proves that this holds precisely when the kernel of p refines the kernel of q, equivalently when every fiber of p lies inside a single fiber of q, and that the coarsest description sufficient for q is the quotient of X by the kernel of q. The lattice of partitions of X organizes the consequences. The descriptions sufficient for a question form a principal ideal, a family of questions is answered jointly exactly when the kernel of the description refines the meet of their kernels, and the content determined by every member of a family is governed by the join.

Further results treat composition of projections, where loss is shown to be irreversible and its first stage is located exactly; impossibility for deterministic and randomized postprocessing; universal losslessness, which is shown to coincide with injectivity already when only two-valued questions are considered; an exact formula for the optimal worst-case recovery error in terms of Chebyshev radii of fiber images; measure-theoretic and entropic formulations; and a stability theorem that determines exactly which refinements of a question a given description can still answer.

Keywords: projection, factorization, quotient, partition lattice, recoverability, sufficiency

1 Introduction

A description of a system is useful because it discards detail, and it is dangerous for the same reason. Whether the discarded detail matters depends on what is later asked of the description. The sum of the coordinates of a point in the plane determines whether the point lies on any given line $x_1 + x_2 = c$ and determines nothing about the distance of the point from the origin. The sorted version of a list determines its sum, its median, and its length, and it does not determine which element came first. In each case the projection identifies many states and so cannot be inverted, yet it may be entirely adequate for a particular purpose. The aim of this paper is to state, in the simplest setting in which the matter makes sense, exactly when this happens.

Several established notions approach the matter from other directions. Injectivity, the requirement that a description lose nothing at all, is the strongest condition and is rarely needed. Sufficient statistics in the sense of Fisher [4] and of Halmos and Savage [6] preserve what a sample carries about the parameter of a family of distributions, which presupposes a probabilistic model, and the minimal sufficient statistics of Lehmann and Scheffé [9], treated at textbook length by Lehmann and Casella [8], are quotients of the sample space of the same general kind as the ones studied here. The information bottleneck of Tishby, Pereira, and Bialek [13] treats the tradeoff between compression of a description and its relevance to a target variable quantitatively, and Blackwell's comparison of experiments [2] ranks descriptions by the decisions they can support. Nerode's theorem [11] identifies the states of a minimal automaton with the classes of words that admit the same continuations, which is an instance of the construction below. What these have in common is a target, a retained description, and a recovery step. What the present paper isolates is the deterministic and set-theoretic core that remains when no probability measure, loss function, or computational model is assumed.

The core is a factorization theorem. Given a description p and a question q, both defined on the same set of states, the answer can be computed from the description alone exactly when $q = r \circ p$ for some function r, and this holds exactly when p never identifies two states that q distinguishes (Theorem 2.2). The condition is stated entirely in terms of the equivalence relations the two maps induce. This makes the lattice of partitions of the state set the natural home of the subject, and Section 3 develops the consequences in that setting: the coarsest sufficient description of a question, the treatment of families of questions through meets and joins, and a lower bound on the size of any sufficient description. Section 4 studies composition of projections and shows that loss along a pipeline of descriptions is irreversible and can be located at a definite stage. Section 5 establishes impossibility for any postprocessing and characterizes the descriptions that are lossless for every question. Section 6 separates recovery of an answer from reconstruction of a state. Section 7 treats approximate recovery in a metric answer space and gives an exact expression for the optimal worst-case error. Section 8 gives the measure-theoretic and entropic versions. Section 9 proves the stability theorem, which describes exactly which refinements of a question a description remains adequate for. Section 10 records what the theory does not claim and some computability remarks. Each individual result is an elementary consequence of quotient constructions, lattice theory, or standard probability and computability, and none is offered as a new theorem in isolation. The contribution is a single formal language in which these results, and in particular the interval characterization of Section 9, can be stated and used together.

2 Descriptions, Questions, and the Factorization Theorem

Throughout, X is a nonempty set of states. A *description* is a function $p : X \to Y$ and a *question* is a function $q : X \to Z$, where Y is a set of descriptions and Z is a set of possible answers. The words projection and description are used interchangeably for an arbitrary function from states to descriptions; neither implies idempotence, a linear structure, or a coordinate projection. Recovery maps are defined on the image $p(X)$ unless stated otherwise, as in Section 8, where measurability calls for maps defined on all of Y and values off the image are immaterial. For a function f with domain X write

$$\ker f = \{(x, x') \in X \times X : f(x) = f(x')\}$$

for the equivalence relation it induces, and write $f^{-1}(y)$ for the fiber over y. The set $f(X)$ is the image of f.

Definition 2.1. The description p is *lossless relative to* the question q if there exists a function $r : p(X) \to Z$ such that

$$q = r \circ p.$$

Such an r is called a *recovery map.*

The definition makes losslessness a property of a pair. Nothing is asserted about whether p preserves the state, only about whether it preserves the answer to q. The following theorem characterizes the property in four equivalent ways.

Theorem 2.2 (Factorization)**.** *For functions $p : X \to Y$ and $q : X \to Z$ the following conditions are equivalent.*

(i) *There exists $r : p(X) \to Z$ with $q = r \circ p$.*

(ii) *For all $x, x' \in X$, $p(x) = p(x') \implies q(x) = q(x')$.*

(iii) *Every fiber of p is contained in a single fiber of q.*

(iv) $\ker p \subseteq \ker q$.

When these hold, the recovery map r is unique.

Proof. (i)⇒(ii). If $p(x) = p(x')$ then $q(x) = r(p(x)) = r(p(x')) = q(x')$.

(ii)⇒(iii). Let $y \in p(X)$ and fix $x_0 \in p^{-1}(y)$. By (ii) every $x \in p^{-1}(y)$ satisfies $q(x) = q(x_0)$, so $p^{-1}(y) \subseteq q^{-1}(q(x_0))$.

(iii)⇒(iv). If $(x, x') \in \ker p$ then x and x' lie in the same fiber of p, hence in the same fiber of q, so $(x, x') \in \ker q$.

(iv)⇒(i). For $y \in p(X)$ define $r(y) = q(x)$ for any $x \in p^{-1}(y)$. By (iv) the value does not depend on the choice of x, so r is a well-defined function, and $r(p(x)) = q(x)$ for every x by construction.

For uniqueness, if r_1 and r_2 both satisfy (i), then for any $y \in p(X)$ choose x with $p(x) = y$ and observe $r_1(y) = q(x) = r_2(y)$. □

Remark 2.3. The construction of r in the last step selects no representative of any fiber, because every representative gives the same value. The recovery map is therefore the graph $\{(y, z) : \exists x,\ p(x) = y,\ q(x) = z\}$, which is a function exactly when condition (iv) holds, and no appeal to the axiom of choice is involved. In categorical language the theorem is the universal property of the quotient map $X \to X/\ker p$ in the category of sets: a map out of X factors through the quotient exactly when it is constant on the classes [10].

Example 2.4. Let $X = \mathbb{R}^2$ and $p(x_1, x_2) = x_1 + x_2$. The question $q(x) = \mathbf{1}[x_1 + x_2 = 1]$ is answered by p with $r(t) = \mathbf{1}[t = 1]$. The question $q(x) = x_1^2 + x_2^2$ is not, because $(2, 0)$ and $(1, 1)$ have the same sum and different squared norms.

Let $X = \mathbb{R}^n$ and let p send a tuple to its sorted rearrangement. Then p is lossless relative to the sum, the maximum, and the median of the entries, and it is not lossless relative to the first entry, since $(1, 2)$ and $(2, 1)$ have the same image.

3 The Lattice of Partitions

Condition (iv) of Theorem 2.2 compares two equivalence relations by inclusion, which suggests organizing the subject by the order structure of the set of all equivalence relations on X. Write $\mathrm{Part}(X)$ for this set, with its elements identified with partitions of X, and order it by inclusion of relations. Write $\pi \sqsubseteq \sigma$ when $\pi \subseteq \sigma$, which holds exactly when every block of π lies inside a block of σ. In this order finer partitions are smaller.

Proposition 3.1 (Ore [12]; see also Birkhoff [1])**.** $(\mathrm{Part}(X), \sqsubseteq)$ *is a complete lattice. The meet of a family is the intersection of the relations, that is, the common refinement. The join of a family is the equivalence relation generated by their union, that is, the transitive closure of the union. The least element is the equality relation and the greatest element is* $X \times X$.

Every function f determines the partition $\ker f$, and every partition π arises as the kernel of the quotient map $X \to X/\pi$. Moreover f induces a bijection from $X/\ker f$ onto $f(X)$. A question considered up to renaming of its answers is therefore the same thing as a partition of X, and Theorem 2.2 can be restated as follows.

Proposition 3.2 (Sufficiency is the refinement order)**.** *The description p is lossless relative to q if and only if* $\ker p \sqsubseteq \ker q$. *Consequently:*

(a) *The descriptions lossless relative to q are exactly those whose kernels lie in the principal ideal $\{\pi : \pi \sqsubseteq \ker q\}$, which has greatest element* $\ker q$.

(b) *The questions answered by p are exactly those whose kernels lie in the principal filter $\{\kappa : \ker p \sqsubseteq \kappa\}$.*

(c) *The quotient map $X \to X/\ker q$, which is the question q itself up to renaming of answers, is the coarsest description lossless relative to q, and q is a function of every description lossless relative to q, that is, q factors through it.*

(d) *If p is lossless relative to q then $|p(X)| \geq |q(X)|$, and equality is attained by q.*

Proof. The first statement is condition (iv) of Theorem 2.2. Parts (a) and (b) are restatements of it. For (c), if p is lossless relative to q then $q = r \circ p$, so q is obtained from p by postprocessing. For (d), $q(X) = r(p(X))$, so $|q(X)| \leq |p(X)|$. □

Parts (a) and (b) are the two sections of the single relation $\sqsubseteq$: (a) varies the description at a fixed question and (b) varies the question at a fixed description, so the two are dual. The refinements of q that p can still answer form the set $\{\kappa : \ker p \sqsubseteq \kappa\} \cap \{\kappa : \kappa \sqsubseteq \ker q\}$, the intersection of a principal filter and a principal ideal, and Section 9 studies exactly this set.

Part (d) gives a floor on description length: a description lossless relative to a question with N distinct possible answers needs at least N distinct values, hence at least $\lceil \log_2 N \rceil$ bits under a fixed-length binary encoding, and the floor is reached by the question itself. Minimality here is minimality in the refinement order. It says nothing about computational cost, storage complexity, or expected code length. The lattice structure handles several questions at once.

Proposition 3.3 (Joint questions and common content)**.** *Let $(q_i)_{i \in I}$ be a family of questions on X with $\kappa_i = \ker q_i$, and let $Q = (q_i)_{i \in I}$ be the joint question with values in the product of the answer sets.*

(a) $\ker Q = \bigwedge_i \kappa_i$. *Hence p is lossless relative to Q if and only if it is lossless relative to every q_i, if and only if $\ker p \sqsubseteq \bigwedge_i \kappa_i$.*

(b) *A question f on X can be answered from each q_i separately if and only if $\bigvee_i \kappa_i \sqsubseteq \ker f$. The finest such question has kernel $\bigvee_i \kappa_i$.*

Proof. (a) Two states have equal joint answers exactly when they have equal answers to each q_i, so $\ker Q = \bigcap_i \ker q_i$, which is the meet by Proposition 3.1. The rest follows from Proposition 3.2. (b) By Proposition 3.2, f is answered from q_i exactly when $\kappa_i \sqsubseteq \ker f$. This holds for every i exactly when the least upper bound $\bigvee_i \kappa_i$ refines $\ker f$. □

Part (b) is the deterministic counterpart of the common part of two random variables in the sense of Gács and Körner [5]: what two questions determine in common is governed by the connected components of the relation generated by their identifications, and no more can be extracted from both. The comparison is an analogy only, since the construction of Gács and Körner is probabilistic and concerns a quantitative entropy, whereas the statement here is purely set-theoretic.

Example 3.4 (Divisibility)**.** Let $X = \mathbb{Z}$ and let $p_k(n) = n \bmod k$. Then $\ker p_k \sqsubseteq \ker p_m$ exactly when m divides k. A description of n modulo 12 is lossless relative to n mod 4 and to n mod 3, hence to the joint question, which is n mod 12 by the Chinese remainder theorem, and is not lossless relative to n mod 5. The divisibility order embeds in Part($\mathbb{Z}$) through this construction.

Example 3.5 (Continuations of a language)**.** Let Σ^* be the set of finite words over a finite alphabet and let $L \subseteq \Sigma^*$ be a language. For each word z define the question $q_z(w) = \mathbf{1}[wz \in L]$. The meet $\bigwedge_z \ker q_z$ is the relation that identifies w and w' exactly when $wz \in L \iff w'z \in L$ for every z, which is the Nerode equivalence [11]. By Proposition 3.3, a description of words is lossless relative to every future continuation question exactly when its kernel refines the Nerode equivalence, and the coarsest such description is $\Sigma^*/\sim_L$. Its cardinality, when finite, is the number of states of the minimal deterministic automaton for L.

4 Composition: Loss Along a Pipeline

Descriptions are frequently processed further, and a later stage can destroy an answer that an earlier stage preserved. The precise statement involves the question induced on the image of the first stage. The second-stage map is written s, to distinguish it from the recovery map r.

Proposition 4.1 (Composition)**.** *Let $p : X \to Y$ be lossless relative to q with recovery map $r : p(X) \to Z$, and let $s : Y \to W$. Then $s \circ p$ is lossless relative to q if and only if the restriction $s|_{p(X)}$ is lossless relative to r, regarded as a question on the set $p(X)$. Equivalently,*

$$s(y) = s(y') \implies r(y) = r(y') \qquad \text{for all } y, y' \in p(X).$$

Proof. By Theorem 2.2, $s \circ p$ is lossless relative to q if and only if $s(p(x)) = s(p(x'))$ implies $q(x) = q(x')$ for all $x, x' \in X$. Since $q = r \circ p$ and p maps X onto $p(X)$, this is the stated condition on $y = p(x)$ and $y' = p(x')$. □

Corollary 4.2 (Loss is irreversible)**.** *Let $p_0 = p$ and $p_k = s_k \circ p_{k-1}$ for $k = 1, \ldots, n$. Then $\ker p_0 \sqsubseteq \ker p_1 \sqsubseteq \cdots \sqsubseteq \ker p_n$. Consequently, if p_k is not lossless relative to q then no p_m with $m \geq k$ is lossless relative to q, and the first stage at which losslessness fails is the first k for which $\ker p_k \not\sqsubseteq \ker q$.*

Proof. Each p_k is a function of p_{k-1}, so $p_{k-1}(x) = p_{k-1}(x')$ implies $p_k(x) = p_k(x')$, which gives the chain of kernels. If $\ker p_k \not\sqsubseteq \ker q$ then $\ker p_m \not\sqsubseteq \ker q$ for $m \geq k$, because $\ker p_k \sqsubseteq \ker p_m$. □

Example 4.3. Let $X = \mathbb{R}^2$, let $p(x) = x_1 + x_2$, and let $s = \lfloor \cdot \rfloor$. The description p is lossless relative to both $q_1(x) = \mathbf{1}[x_1 + x_2 \geq 0]$ and $q_2(x) = \mathbf{1}[x_1 + x_2 \geq \frac{1}{2}]$. The composite $s \circ p$ remains lossless relative to q_1, since $\lfloor t \rfloor \geq 0$ if and only if $t \geq 0$, and it is not lossless relative to q_2, since sums 0.25 and 0.75 have the same floor and different answers.

Remark 4.4. The corollary concerns pipelines in which each stage receives only the output of its predecessor. A later stage supplied with an independent observation $a(x)$ of the original state may well recover $q(x)$, but that is a different description, namely $x \mapsto (p(x), a(x))$, to which the theory applies afresh. It is outside the hypothesis of the corollary and not a counterexample to it.

5 Impossibility and Universal Losslessness

When the condition of Theorem 2.2 fails, the failure cannot be repaired by further computation on the description. Corollary 4.2 already shows this along a chain of stages; the next result makes the point for a single postprocessing step, deterministic or randomized.

Proposition 5.1 (Impossibility)**.** *Suppose $p(x) = p(x')$ and $q(x) \neq q(x')$. Then for every function $t : Y \to Z$, the value $t(p(x))$ differs from at least one of $q(x)$ and $q(x')$. Moreover, if Z has measurable singletons and $y \mapsto \nu_y$ is any randomized recovery assigning a probability measure ν_y on Z to each description, then*

$$\min\{\nu_{p(x)}(\{q(x)\}),\ \nu_{p(x)}(\{q(x')\})\} \leq \tfrac{1}{2}.$$

Proof. For the first claim, $t(p(x)) = t(p(x'))$ is a single point of Z, and it cannot equal both $q(x)$ and $q(x')$. For the second claim, set $y = p(x) = p(x')$. The singletons $\{q(x)\}$ and $\{q(x')\}$ are disjoint, so $\nu_y(\{q(x)\}) + \nu_y(\{q(x')\}) \leq 1$, and the smaller of the two terms is at most $\frac{1}{2}$. □

Thus no postprocessing of the description, however elaborate, deterministic or randomized, and however much computation it uses, can resolve an ambiguity that the projection itself introduced. Thus, on such a pair of states, no randomized recovery succeeds with probability above one half at both states simultaneously, and the worse of the two success probabilities is at most one half. The bound sharpens as the ambiguity grows.

Corollary 5.2. *Let Z have measurable singletons, and suppose some fiber of p is mapped by q onto a set containing m distinct answers $z_1, \dots, z_m$, and let $y \mapsto \nu_y$ be a randomized recovery. Then $\min_j \nu_y(\{z_j\}) \le 1/m$. If the set contains infinitely many distinct answers, then the infimum over the states in that fiber of the success probability of the recovery is zero.*

Proof. The singletons $\{z_j\}$ are pairwise disjoint, so $\sum_{j=1}^{m} \nu_y(\{z_j\}) \le 1$ and the smallest term is at most $1/m$. For an infinite set apply this with every finite m. □

At the other extreme are descriptions that lose nothing for any question whatever.

Theorem 5.3 (Universal losslessness)**.** *For a description $p : X \to Y$ the following are equivalent.*

(i) *p is injective.*

(ii) *p is lossless relative to every question on X, with arbitrary answer set.*

(iii) *p is lossless relative to every two-valued question on X.*

(iv) *p is lossless relative to the identity question* id_X*.*

Proof. (i)⇒(ii). If p is injective then $\ker p$ is the equality relation, the least element of $\mathrm{Part}(X)$, so $\ker p \sqsubseteq \ker q$ for every q. (ii)⇒(iii) is immediate. (iii)⇒(i). Suppose $p(x) = p(x')$ with $x \neq x'$. The question $q = \mathbf{1}[\,\cdot = x\,]$ takes the value 1 at x and 0 at x', so $\ker p \not\sqsubseteq \ker q$, contradicting (iii). Finally (i)⇔(iv) because $\ker \mathrm{id}_X$ is the equality relation, so $\ker p \subseteq \ker \mathrm{id}_X$ holds exactly when p is injective. □

The implication (iii)⇒(i) shows that the boundary between purpose-relative preservation and unrestricted preservation is sharp: a description that fails to be injective fails already on some yes-or-no question. A finite count shows how much a non-injective description can still answer.

Proposition 5.4 (Counting answerable questions)**.** *Let X be finite with $|X| = n$ and let p have $k = |p(X)|$ values. The number of questions on X answered by p, counted up to renaming of answers, is the Bell number B_k, whereas the number of all questions on X counted the same way is B_n.*

Proof. By Proposition 3.2(b), the answered questions correspond to partitions coarser than $\ker p$. Such a partition is determined by a partition of the set of k blocks of $\ker p$, and these are counted by B_k. □

For $n = 10$ and $k = 5$ the count is $B_5 = 52$ against $B_{10} = 115975$. A description with five values on ten states still answers a small but nonzero fraction of all questions, counted as partitions. The ratio is a combinatorial fraction under uniform counting of partitions, and it should not be read as the proportion of practically relevant questions that remain answerable, since no distribution over questions has been supplied.

6 Recoverability Versus Reconstruction

Two claims are easily conflated. *State reconstruction* asks for a map d with $d \circ p = \mathrm{id}_X$. *Answer recovery* asks for a map r with $r \circ p = q$. The first holds if and only if p is injective; the second holds if and only if the fibers of p respect the distinctions made by q. State reconstruction is the special case of answer recovery in which the question is the identity (Theorem 5.3(iv)).

The relation between the two can be described without choosing the question. Factor p through its quotient as $p = \iota \circ \pi_p$, where $\pi_p : X \to X/\ker p$ is the quotient map and $\iota : X/\ker p \to Y$ is injective. Answer recovery for q is exactly the statement that q descends to a function $\bar{q} : X/\ker p \to Z$, and it makes no demand that the points of $X/\ker p$ be resolved into states. If $s : p(X) \to X$ is any section of p, which exists by choice, then $q = (q \circ s) \circ p$ whenever p is lossless relative to q, so a description answers q exactly when reconstructing an arbitrary compatible state and asking q of it gives a result that does not depend on which compatible state was chosen.

This distinction prevents a common equivocation between recovering the original object and recovering everything relevant to a specified purpose. It also explains why compression that destroys reconstruction is not thereby shown to destroy any particular answer.

7 Approximate Recovery

Let (Z, d) be a metric space. For a description p and question q define the optimal worst-case recovery error

$$E(p, q) = \inf_{r:\, p(X) \to Z} \sup_{x \in X} d\big(q(x),\, r(p(x))\big).$$

Exact losslessness gives $E(p, q) = 0$, with the infimum attained by the recovery map of Theorem 2.2. The error can be computed from the geometry of the fibers. For $y \in p(X)$ let $A_y = q(p^{-1}(y)) \subseteq Z$ be the set of answers compatible with the description y, and write

$$\operatorname{diam} A = \sup_{a,b \in A} d(a, b), \qquad \operatorname{rad}_Z A = \inf_{c \in Z} \sup_{a \in A} d(a, c)$$

for the diameter of a set and its Chebyshev radius relative to Z.

The proof below selects, for every fiber simultaneously, an approximately optimal center, and for an arbitrary family of fibers this uses the axiom of choice, in contrast with Theorem 2.2, whose proof requires none. The theorem is therefore stated in ordinary set theory with choice. Moreover $E(p, q)$ is an infimum and need not be attained by any recovery map: if $Z = \mathbb{R} \setminus \{\frac{1}{2}\}$ and some fiber image is $\{0, 1\}$, its radius is $\frac{1}{2}$ and no center in Z attains it, because centers approaching $\frac{1}{2}$ make the maximum distance approach $\frac{1}{2}$ while the excluded midpoint is the only center in $\mathbb{R}$ that achieves it. No compactness or completeness of Z is needed for the formula.

Theorem 7.1 (Exact error formula)**.** *For every description p and question q,*

$$E(p, q) = \sup_{y \in p(X)} \operatorname{rad}_Z A_y, \qquad \tfrac{1}{2} \sup_{y \in p(X)} \operatorname{diam} A_y \;\le\; E(p, q) \;\le\; \sup_{y \in p(X)} \operatorname{diam} A_y.$$

In particular $E(p, q) = 0$ if and only if p is lossless relative to q.

Proof. Write $f(y, c) = \sup_{a \in A_y} d(a, c)$, so that $E(p, q) = \inf_r \sup_y f(y, r(y))$. The inequality $\inf_r \sup_y f(y, r(y)) \ge \sup_y \inf_c f(y, c)$ holds because $f(y, r(y)) \ge \inf_c f(y, c)$ for every r and

every y. The reverse inequality is the substantive one: given $\varepsilon > 0$ choose for each y a point c_y with $f(y, c_y) \le \inf_c f(y, c) + \varepsilon$ and set $r(y) = c_y$; then $\sup_y f(y, r(y)) \le \sup_y \operatorname{rad}_Z A_y + \varepsilon$. This proves $E(p, q) = \sup_y \operatorname{rad}_Z A_y$.

For the bounds on the radius, fix $y \in p(X)$. The fiber $p^{-1}(y)$ is nonempty, so A_y is nonempty. For any $c \in Z$ and $a, b \in A_y$ the triangle inequality gives $d(a, b) \le d(a, c) + d(c, b) \le 2f(y, c)$, so $\operatorname{diam} A_y \le 2 \operatorname{rad}_Z A_y$. Taking $c \in A_y$ gives $f(y, c) \le \operatorname{diam} A_y$, so $\operatorname{rad}_Z A_y \le \operatorname{diam} A_y$.

If $E(p, q) = 0$ then every A_y has diameter zero, so q is constant on each fiber of p and Theorem 2.2 applies. The converse was noted above. □

The theorem removes any need for additional hypotheses to conclude exactness from zero error: because d is a metric, and in particular $d(a, b) = 0$ implies $a = b$, vanishing of the optimal error forces every fiber image to be a single point. It also makes the lower bound of Proposition 5.1 quantitative, since the error is at least half the largest spread of answers hidden inside a single fiber.

Example 7.2. Let $X = Z = \mathbb{R}$ with the usual metric, let $p(x) = \lfloor x \rfloor$, and let $q = \mathrm{id}$. The fibers are the intervals $[n, n + 1)$, whose Chebyshev radius in $\mathbb{R}$ is $\frac{1}{2}$, so $E(p, q) = \frac{1}{2}$, with the supremum over each fiber approached but not attained. Rounding to the nearest integer would change the fibers to $[n - \frac{1}{2}, n + \frac{1}{2})$, but their lengths, and hence their Chebyshev radii, would not change, so the error would remain $\frac{1}{2}$.

8 Probabilistic and Entropic Versions

Let $(X, \mathcal{X}, \mu)$ be a probability space, let $(Y, \mathcal{Y})$ and $(Z, \mathcal{Z})$ be measurable spaces, and let $p : X \to Y$ and $q : X \to Z$ be measurable. Say that p is *almost surely lossless relative to* q if there is a measurable $r : Y \to Z$ with $q = r \circ p$ μ-almost surely. Unlike Definition 2.1, the recovery map is here defined on all of Y so that it can be measurable; its values off $p(X)$ are immaterial. Write $\sigma(p) = p^{-1}(\mathcal{Y})$ for the σ-algebra generated by p. It plays the role in the measurable setting that $\ker p$ plays in the deterministic one: it is the information p retains.

Theorem 8.1 (Measurability criterion)**.** *Suppose $(Z, \mathcal{Z})$ is a standard Borel space. Then p is almost surely lossless relative to q if and only if there is a $\sigma(p)$-measurable $q' : X \to Z$ with $q = q'$ μ-almost surely.*

Proof. If $q = r \circ p$ almost surely, then $q' = r \circ p$ is $\sigma(p)$-measurable and agrees with q almost surely. Conversely, suppose q' is $\sigma(p)$-measurable; it takes values in Z, which is standard Borel by hypothesis. By the Doob–Dynkin lemma [7, Lemma 1.13] there is a measurable $r : Y \to Z$ with $q' = r \circ p$, and then $q = r \circ p$ almost surely. □

Thus almost sure recovery of the answer from the description is the statement that the answer is measurable with respect to the information the description retains, up to null sets. When X, Y, and Z are countable and every point has positive mass, the condition reduces to the exact condition of Theorem 2.2, and in general the identity $q = r \circ p$ is required to hold only outside a single measurable null set. That exceptional set may contain many states and many conflicting fibers, uncountably many in general, and nothing is asserted about its size or structure beyond its measure. The entropic form is as follows.

Proposition 8.2 (Entropy criterion)**.** *Let Y and Z be countable and let $H(q \mid p)$ denote the conditional Shannon entropy of $q(X)$ given $p(X)$ under μ. Then p is almost surely lossless relative to q if and only if $H(q \mid p) = 0$, where $H(q \mid p) \in [0, \infty]$.*

Proof. Write $H(q \mid p) = \sum_y \mu(p = y)\, H(q \mid p = y)$ with the sum over values of positive mass. This vanishes exactly when each conditional distribution of q given $p = y$ has zero entropy, which for a discrete distribution means that it is a point mass at some $r(y)$. Since Y and Z are countable no measurability issue arises, and setting $r(y)$ to any fixed element of Z on values of null mass gives $q = r \circ p$ almost surely. The converse is immediate because a deterministic function of the conditioning variable has zero conditional entropy. □

Remark 8.3. Under the countability hypotheses of Proposition 8.2, the condition $H(q \mid p) = 0$ is equivalent to the existence of a $\sigma(p)$-measurable version of q, which is the criterion of Theorem 8.1. The entropic form is the quantitative face of the same condition.

Remark 8.4. Let $P_e^* = \inf_r \Pr[r(p(X)) \neq q(X)]$ be the optimal error probability, a distribution-dependent criterion distinct from the worst-case error $E(p, q)$ of Section 7. When Z is finite, Fano's inequality [3] gives $H(q \mid p) \leq h(P_e^*) + P_e^* \log(|Z| - 1)$, with logarithms to a common base and h the binary entropy. Positive conditional entropy therefore forces a positive lower bound on the error probability of every recovery, which is the probabilistic counterpart of Proposition 5.1.

Remark 8.5. The relation to statistical sufficiency should be stated carefully. A statistic is sufficient in the sense of Fisher [4] and of Halmos and Savage [6] when the conditional distribution of the data given the statistic does not depend on the parameter of a family of distributions, and Blackwell [2] compares experiments by the decisions they support. Those notions concern the preservation of the probabilistic structure of a family. The condition studied here concerns recovery of one deterministic answer. The two answer different questions, and the quotient constructions of Section 3 are analogous to minimal sufficient statistics [9] without being identical to them.

9 Stability Under Refinement of the Question

A description adequate for the question originally asked may be inadequate for a later and more discriminating one. The factorization theorem allows this to be stated without treating the original description as mistaken. Say that a question q' *refines* a question q if $\ker q' \sqsubseteq \ker q$, that is, if q' distinguishes at least as many states as q and q can be computed from q'.

Theorem 9.1 (Stability under refinement)**.** *Let p be lossless relative to q and let q' refine q.*

(a) *p is lossless relative to q' if and only if q' separates no pair of states identified by p, that is,* $\ker p \sqsubseteq \ker q'$.

(b) *The refinements of q that p can still answer are exactly those q' with* $\ker p \sqsubseteq \ker q' \sqsubseteq \ker q$, *the order interval* $[\ker p, \ker q] = \{\kappa : \ker p \sqsubseteq \kappa \sqsubseteq \ker q\}$ *of* $\mathrm{Part}(X)$. *The interval contains both endpoints, and it is the single point* $\{\ker q\}$ *if and only if* $\ker p = \ker q$.

(c) *If $q_1, q_2, \ldots$ is a sequence of refinements of q with* $\ker q_{n+1} \sqsubseteq \ker q_n$, *then p is lossless relative to every q_n if and only if it is lossless relative to the joint question $(q_n)_{n \geq 1}$, that is, if and only if* $\ker p \sqsubseteq \bigwedge_n \ker q_n$.

Proof. (a) is Theorem 2.2 applied to q'. (b) The conditions $\ker p \sqsubseteq \ker q'$ and $\ker q' \sqsubseteq \ker q$ together describe the stated interval; the first is (a) and the second is the hypothesis that q' refines q. Both endpoints belong to the interval, so it is a single point exactly when its two endpoints coincide, that is, when $\ker p = \ker q$. (c) is Proposition 3.3(a). □

The theorem identifies the obstruction to adequacy for future questions with a specific object: a pair of states that the description identifies and that some later question separates. A description for which $\ker p = \ker q$ is minimal in the sense of Proposition 3.2(c) and has no room to grow, since every strict refinement of q separates a pair that p identifies. A description with a larger gap between $\ker p$ and $\ker q$ is adequate not only for q but for every question whose kernel falls in the interval, and that interval measures the headroom for future inquiry. The interval describes the admissible refinements under unrestricted deterministic recovery; it does not say which future questions are likely, computationally tractable, or scientifically useful. Injective descriptions have the full interval $[\Delta_X, \ker q]$, which contains every refinement of q.

Part (c) shows that adequacy for a research program, in the sense of an unbounded sequence of increasingly discriminating questions, is a single condition on the meet of their kernels. Example 3.5 is the special case in which the questions q_z range over all continuations and the relevant meet is the Nerode equivalence.

Example 9.2. Let $X = \mathbb{R}^n$ with $n \geq 2$, let p be the sorted rearrangement, and let q be the sum. The refinement of q that returns the sum together with the maximum lies in $[\ker p, \ker q]$ and is answered by p. The refinement that returns the sum together with the first entry does not lie in the interval, because $(1, 2)$ and $(2, 1)$ have the same sorted rearrangement. Since $\ker p \neq \ker q$, the interval has more than one point.

Corollary 9.3 (Capacity of two descriptions)**.** *Let p_1 and p_2 both be lossless relative to q. The questions answerable from each of p_1 and p_2 separately are exactly those whose kernels lie above $\ker p_1 \vee \ker p_2$. The questions answerable from the joint description (p_1, p_2) are exactly those whose kernels lie above $\ker p_1 \wedge \ker p_2$. Both families contain q, and the second contains the first. (A question lies above a partition when its kernel is coarser, so the family above the meet is the larger one.)*

Proof. A question f is answerable from p_i exactly when $\ker p_i \sqsubseteq \ker f$, and from both exactly when $\ker p_1 \vee \ker p_2 \sqsubseteq \ker f$. By Proposition 3.3(a) the kernel of the joint description is $\ker p_1 \wedge \ker p_2$, so f is answerable from it exactly when $\ker p_1 \wedge \ker p_2 \sqsubseteq \ker f$. Both $\ker p_i$ lie below $\ker q$, so their join does too, and so does their meet. Finally $\ker p_1 \wedge \ker p_2 \sqsubseteq \ker p_1 \vee \ker p_2$ gives the inclusion of the families. □

10 Scope and Computability

The theory above is deliberately minimal. It treats single-valued questions, so a question with several acceptable answers, or an answer relation instead of an answer function, would require replacing equivalence relations with covers or with relations and is not covered by the factorization theorem. It does not say how a description should be chosen, only how adequacy is decided once the description and question are given. And it does not concern the computational difficulty of recovery, to which the following remarks apply.

For finite X given explicitly, losslessness is decided in time linear in $|X|$ by tabulating q on each fiber of p, in a model where evaluating p and q, testing equality, and looking up table

entries cost constant time; for other representations the cost depends on how p, q, and their values are encoded. For $X = \mathbb{N}$ and computable p and q, the situation splits. The recovery map is the restriction to $p(\mathbb{N})$ of a partial computable function: on input y one searches $n = 0, 1, 2, \ldots$ for the first n with $p(n) = y$ and outputs $q(n)$. This halts exactly when $y \in p(\mathbb{N})$ and is correct whenever p is lossless relative to q, but it need not halt outside the image, so it is not in general a total computable function. The property itself is undecidable.

Proposition 10.1 (Undecidability for programs)**.** *There is no algorithm that, given indices of total computable functions $p, q : \mathbb{N} \to \mathbb{N}$, decides whether p is lossless relative to q.*

Proof. Let p be the constant function 0. By Theorem 2.2, p is lossless relative to q exactly when q is constant. For a Turing machine e let $q_e(0) = 0$ and, for $n \geq 1$, let $q_e(n) = 1$ if e halts within n steps on the empty input and $q_e(n) = 0$ otherwise. Each q_e is total and computable, with an index computable from e by the s-m-n theorem, and q_e is constant if and only if e never halts. A decision procedure for losslessness, applied to the constant function 0 and q_e, would therefore decide the complement of the halting problem. The reduction is explicit and self-contained. □

The proposition concerns the extensional condition on all of X. It does not obstruct the use of the theorem in settings where the state space is finite, where the fibers are known by construction, or where a proof of the factorization is supplied together with the description.

Two extensions appear natural and are left aside. The first replaces the deterministic question by a family of conditional distributions and quantifies loss by a divergence, which leads toward the treatments of Blackwell [2] and of Tishby, Pereira, and Bialek [13]. The second asks which descriptions are lossless relative to a given question while also satisfying structural constraints, such as being computed by a bounded circuit or by a local update rule, in which case the lattice picture persists but the available kernels form a proper subset of $\mathrm{Part}(X)$.

One limitation deserves emphasis. The formalism takes the question q as already specified. It can decide whether a description is sufficient for that question and which refinements of it remain answerable, but it cannot say which questions ought to be asked, which future distinctions will matter, or which losses are acceptable for a given purpose. Those problems require additional structure, and the theory makes no claim about them.

References

[1] G. Birkhoff, *Lattice Theory*, 3rd ed., American Mathematical Society Colloquium Publications, vol. 25, American Mathematical Society, Providence, 1967.

[2] D. Blackwell, Equivalent Comparisons of Experiments, *Annals of Mathematical Statistics* **24**(2) (1953), 265–272.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, Hoboken, 2006.

[4] R. A. Fisher, On the Mathematical Foundations of Theoretical Statistics, *Philosophical Transactions of the Royal Society of London, Series A* **222** (1922), 309–368.

[5] P. Gács and J. Körner, Common Information Is Far Less Than Mutual Information, *Problems of Control and Information Theory* **2**(2) (1973), 149–162.

[6] P. R. Halmos and L. J. Savage, Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics, *Annals of Mathematical Statistics* **20**(2) (1949), 225–241.

[7] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed., Springer, New York, 2002.

[8] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed., Springer, New York, 1998.

[9] E. L. Lehmann and H. Scheffé, Completeness, Similar Regions, and Unbiased Estimation: Part I, *Sankhyā* **10** (1950), 305–340.

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

[11] A. Nerode, Linear Automaton Transformations, *Proceedings of the American Mathematical Society* **9**(4) (1958), 541–544.

[12] O. Ore, Theory of Equivalence Relations, *Duke Mathematical Journal* **9**(3) (1942), 573–627.

[13] N. Tishby, F. C. Pereira, and W. Bialek, The Information Bottleneck Method, in *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing* (1999), 368–377.