

Information: Quantifying Alternatives and Dependence

IDE-002 · Fundamental Ideas

Concept essays for The Universe and Its Observers

Author: Juan Segurolo

Version: v0.6 EN · 19 September 2026

Status: Publication version

Article form: Concept essay

Scale / context: Statistical and physical information across scales

Boundary: Quantitative information and coding; physical signals and signalling are treated in IDE-024

Related works: IDE-003, IDE-008, IDE-024

Epistemic profile: Established knowledge

Licence at publication: All Rights Reserved

Article Abstract

Information is a quantitative language for distinguishable alternatives, statistical uncertainty and dependence. In Shannon's theory, a source is represented by possible messages with probabilities; entropy summarizes average uncertainty, while mutual information measures how much observing one variable reduces uncertainty about another. These quantities set rigorous limits on compression and reliable communication, but they do not by themselves supply semantic meaning, causal direction or a physical substance called information. A bit is a unit defined by logarithms to base two, and the same logical distinction can be embodied in many physical systems. Physical implementation nevertheless matters: memory must be represented, manipulated and sometimes erased, and logically irreversible operations have thermodynamic consequences. This article separates the mathematical measure from its physical carriers and from causal signalling, establishing the boundary with IDE-024, which asks what a signal is and how influence is physically conveyed.

In 60 Seconds

Start with a set of possible outcomes and probabilities. Shannon entropy, measured in bits when logarithms use base two, gives the average uncertainty of that distribution; mutual information measures statistical dependence between two variables. These are not vague metaphors. They support coding theorems: frequent outcomes can receive shorter descriptions, noisy channels have finite capacities, and processing data cannot create information about an upstream variable that was absent from the input.[1,2] Yet the mathematics intentionally abstracts away from meaning. Two bit strings can have the same Shannon statistics while differing completely in what a human interprets them to mean. [1] Information also is not a weightless fluid that exists independently of representation. A logical bit can be stored in voltage, magnetization or another physical degree of freedom, while erasing logically distinguishable states under thermal conditions carries a minimum thermodynamic cost in the ideal

limit.[3,4] Finally, shared information is not proof of communication or causation: common causes and prior interactions can correlate variables without any present signal between them.[5]

Why This Question Matters

The word information moves easily between telecommunications, physics, biology, computing and everyday speech, and that portability is both useful and dangerous. Without stating the random variables, probabilities and coding assumptions, a numerical information claim may be undefined; without a causal model, mutual information can be mistaken for influence; without a physical model, Landauer's principle can be misquoted as a universal energy cost for every computational step. A disciplined account lets us use the same mathematical tools across fields while keeping distinct questions distinct: how uncertain a source is, how strongly variables depend on one another, how a message is encoded, and how a physical signal actually travels.

1. What Shannon Information Measures

Shannon's starting point is not a substance but a set of possible messages and their probabilities. For a discrete random variable X with outcomes x and probabilities $p(x)$, the entropy $H(X) = -\sum p(x) \log_2 p(x)$ is the expected logarithmic surprise. The base of the logarithm fixes the unit: base two gives bits. Entropy is zero when one outcome is certain and grows when probability is spread among alternatives.[1,2]

That definition is operational because it connects to coding. For long sequences generated by a source, entropy sets the fundamental scale for average lossless description length: highly predictable sources can be compressed more than highly uncertain ones. Shannon's noisy-channel theory likewise links reliable communication to channel capacity rather than to any particular alphabet or hardware.[1,2] The abstraction is powerful precisely because it ignores the material form of the symbols.

Mutual information asks a different question. $I(X;Y)$ quantifies how much the joint distribution of X and Y departs from independence, equivalently how much knowing one reduces uncertainty about the other. It is non-negative and vanishes when the variables are statistically independent under the specified distribution.[2] A high value therefore means strong statistical dependence, not that X caused Y or that a message physically travelled from X to Y .

Where That Leaves Us

Information theory begins with explicit alternatives and probability distributions, then turns them into quantities that constrain compression and communication. Entropy and mutual information are properties of a probabilistic description, not hidden fluids inside objects. That makes the next question unavoidable: if the mathematics is abstract about hardware, what changes when the same logical distinctions are implemented in real physical devices?

2. A Bit Is Logical, but Its Implementation Is Physical

A bit is a binary distinction, not a microscopic bead. The logical states 0 and 1 can be represented by two voltage ranges, two magnetization directions, two optical states or many other physically distinguishable configurations. Information theory can treat these realizations as equivalent when

only the logical alternatives matter, while engineering must also track noise, stability, energy barriers, error probabilities and the dynamics used to read or change the state.[1,2]

Landauer identified a specific bridge between logical and thermodynamic irreversibility. When a physical operation maps two distinguishable logical states onto one, as idealized erasure does, information about the prior state is discarded. In a thermal environment, carrying out that logically irreversible operation reversibly as far as possible still requires entropy to be exported to the surroundings; for erasing one unbiased bit the familiar ideal lower bound is $kT \ln 2$ of dissipated heat.[3]

The principle does not say that every logical operation must dissipate $kT \ln 2$. Bennett showed that general computation can, in principle, be arranged through logically reversible steps, retaining enough information to retrace intermediate states and avoid erasure until it is actually required.[4] Real computers dissipate far more energy for many engineering reasons, but the conceptual lesson is precise: information is abstract enough to admit many embodiments, while operations on embodied information remain subject to physical law.

Where That Leaves Us

The same logical bit can inhabit very different hardware, so information cannot be identified with one material carrier. Yet implementation is not irrelevant: storing, reading and erasing distinctions are physical processes with noise and thermodynamic constraints. This separates information as a quantitative relation among alternatives from the signal that physically carries a change between locations, the subject reserved for IDE-024.

3. Information Is Not the Same as Meaning

Shannon explicitly designed his communication theory so that semantic aspects of messages were not required for the engineering problem.[1] A code can preserve a sequence with arbitrarily low error without understanding whether that sequence is a poem, a genome, a software file or random-looking output. The mathematics cares about symbol probabilities, code lengths, channel statistics and recoverability under a specified task.

This is not a claim that meaning is unreal. It is a boundary statement about what Shannon quantities measure. Meaning depends on users, conventions, reference, goals and larger models that are not determined by $H(X)$ alone. Two messages can have similar length and entropy while differing radically in semantic content; conversely, the same proposition can be encoded in many strings with different statistical properties. Quantitative information measures therefore do not become a universal theory of semantics merely because the word information is shared.

Context nevertheless enters the mathematics in a disciplined way. One must choose variables, an alphabet or state space, a probability model and often a coding or observational protocol before entropy or mutual information can be computed. Different choices can answer different legitimate questions. Once those choices are fixed, however, the resulting quantity is not arbitrary: the probability model and definitions determine the value and the theorems constrain what transformations can do.[2]

Where That Leaves Us

Shannon information deliberately answers a narrower question than ordinary-language meaning. It quantifies uncertainty and dependence once variables and probabilities are specified; it does not infer what symbols refer to or why they matter to an organism or observer. That boundary prevents a common category error and prepares the next one: statistical dependence can be strong even when no causal influence or present communication exists.

4. Dependence Is Not Causation or Signalling

Mutual information detects statistical dependence. If two variables share a common cause, were correlated by an earlier interaction or are both effects of a third process, observing one can reduce uncertainty about the other even when neither currently sends anything to the other. Information theory alone therefore cannot decide whether the dependence arose by direct influence, common cause, selection or some other causal structure.[2,5]

Causal inference requires additional assumptions or interventions. Pearl's graphical framework makes this explicit by combining probability distributions with structural information about which variables can directly affect which others; only under such assumptions can observational data identify causal effects, and sometimes they cannot.[5] This is a different task from measuring dependence. The distinction matters whenever phrases such as 'information flows from A to B' are used as though a non-zero correlation had already established a physical route.

The data-processing inequality provides a useful but limited constraint. If X influences Y and Y is then processed into Z through a Markov chain $X \rightarrow Y \rightarrow Z$, ordinary processing cannot increase the mutual information that Z contains about X beyond what Y already contained.[2] The theorem constrains statistical structure given a model; it does not establish that the assumed causal chain is the true physical history. Model and measurement must therefore be kept separate.

Where That Leaves Us

Shared information can arise without a contemporaneous signal, so dependence, causation and communication must not be collapsed into one word. Information theory constrains probability distributions; causal inference adds assumptions about intervention and direction; signal physics asks what interaction can actually carry a change. With those roles separated, the final section can state what information does unify across disciplines without turning it into a universal substance.

5. What the Concept Unifies - and Where It Stops

Information theory unifies problems that can be expressed as alternatives, probabilities, coding constraints and statistical dependence. The same formal tools can describe source compression, noisy communication, inference and parts of statistical physics because each problem contains a structure of distinguishable states and uncertainty.[1,2] That common mathematics is substantive, but it does not imply that every use of the word information names the same physical mechanism.

The concept also has boundaries inside mathematics. Differential entropy can be negative and changes under coordinate transformations and scaling, whereas mutual information has stronger invariance properties.[2] Statements such as 'the system contains N bits' therefore require the representation, resolution and probabilistic model to be specified.

The division of labor across this series is deliberate. IDE-003 treats entropy as a broader concept, including its thermodynamic role; IDE-008 treats probability itself; IDE-024 asks what a signal is, focusing on physical carriers, propagation and the conditions for signalling. IDE-002 stops at quantitative information: how alternatives, uncertainty, dependence and coding are represented and constrained. The handoff is conceptual, not evidential, so it needs no citation.

Where That Leaves Us

Information is a rigorous family of quantitative relations, not a universal material ingredient. Its power comes from separating what depends on probability and coding from what requires semantics, causal structure or a physical transmission mechanism. That boundary makes the neighboring articles complementary rather than redundant: entropy, probability and signal each inherit part of the story, but none is simply another name for information.

What We Know, and How

- For discrete sources, Shannon entropy quantifies average uncertainty and sets the fundamental scale for lossless compression.[1,2]
- Mutual information measures statistical dependence and obeys the data-processing inequality under the specified probabilistic structure.[2]
- Logically irreversible erasure has a thermodynamic lower bound in the ideal thermal limit, while logically reversible computation can in principle avoid paying that cost at every step.[3,4]
- Statistical dependence alone does not identify a causal direction; causal claims require additional structural assumptions or interventions.[5]

What We Still Don't Know

- How best to connect information-theoretic quantities to semantic or functional notions in specific biological and cognitive systems without conflating the definitions.
- Which information measures are most informative for strongly non-stationary, high-dimensional systems when the relevant state variables themselves are uncertain.
- How close future physical computing architectures can approach reversible-computation limits while remaining fast, reliable and manufacturable.

What Could Change Our Minds

- A reproducible violation of the lossless source-coding or channel-coding limits under their stated assumptions would require revision of the corresponding information-theoretic framework.
- A controlled erasure experiment that, after accounting for all reservoirs and errors, repeatedly dissipated less than the Landauer bound for an unbiased bit in the quasistatic thermal regime would challenge the standard thermodynamic account.[3]
- A causal-inference procedure that identified direction uniquely from arbitrary observational joint distributions without additional assumptions, while succeeding on cases with common causes and observational equivalence, would overturn the boundary drawn here between dependence and causation.[5]

If You Remember Only One Thing

Information quantifies uncertainty and statistical dependence under explicit choices of variables and probabilities; it is neither semantic meaning, causal influence nor a physical substance.

Related Articles

- IDE-003 — Entropy
- IDE-008 — Probability
- IDE-024 — What Is a Signal?

References

- [1] Claude E. Shannon. A Mathematical Theory of Communication. Bell System Technical Journal 27, 379-423 and 623-656 (1948). DOI 10.1002/j.1538-7305.1948.tb01338.x; 10.1002/j.1538-7305.1948.tb00917.x.
- [2] Thomas M. Cover; Joy A. Thomas. Elements of Information Theory, 2nd ed. Wiley (2006). DOI 10.1002/047174882X.
- [3] Rolf Landauer. Irreversibility and Heat Generation in the Computing Process. IBM Journal of Research and Development 5(3), 183-191 (1961). DOI 10.1147/rd.53.0183.
- [4] Charles H. Bennett. Logical Reversibility of Computation. IBM Journal of Research and Development 17(6), 525-532 (1973). DOI 10.1147/rd.176.0525.
- [5] Judea Pearl. Causal diagrams for empirical research. Biometrika 82(4), 669-688 (1995). DOI 10.1093/biomet/82.4.669.