

What Makes a Measure "Genuinely Natural"? Bits, Nats, and the Audit of Naturalness Claims in Measurement, Information, and Language

What Makes a Measure "Genuinely Natural"? Bits, Nats, and the Audit of Naturalness Claims in Measurement, Information, and Language

Abstract

The word "natural" carries persuasive weight across the quantitative sciences: it appears in "natural logarithm," "natural language," and "natural images," and most pointedly in the recent information-theoretic proposal that the nat — the information unit based on the natural logarithm — is a "genuinely natural" measure, in contrast to Shannon's bit. This paper asks whether such naturalness claims are substantive or rhetorical. We formalize three candidate criteria for a measurement convention being genuinely natural — ratio invariance, dimensionlessness of derived constants, and reference-class independence — and audit the bit-versus-nat question against them. All quantitative content is derived explicitly: we compute $\ln 2 = 0.6931471808$ from a rapidly convergent series, the conversion constant $\log_2 e = 1.4426950$, the distribution-independence of entropy ratios, the differential entropy of a unit-variance Gaussian (1.4189386 nats = 2.0470956 bits), and the scaling failure of differential entropy under coordinate change. We find that the choice between bits and nats is a multiplicative constant with no invariance-theoretic consequence, so "genuine" naturalness in the mathematical sense does not discriminate the two; what differs is calculus convenience. We then contrast this mathematical sense of "natural" with the sociolinguistic and dataset-geometric senses in which the term appears across the surveyed literature, and quantify the disciplinary composition of the eight-work corpus (62.5% language-focused, 25% imaging/measurement-focused, 12.5% information-theory-focused) as a descriptive, not causal, observation. We close with falsification conditions and open questions.

1. Introduction

Quantitative science is full of units, and units are conventions. Yet some conventions are advertised as more than conventional: they are called *natural*. The natural logarithm, natural language, natural images — in each case the adjective suggests that the quantity in question is anchored to something less arbitrary than human choice. Whether that suggestion is ever mathematically justified is the subject of this paper.

Our focal case is the information-theoretic proposal of [6], titled "A genuinely natural information measure." The supplied abstract of that work states that Shannon, in his mathematical theory of communication, proposed entropy measured in bits, but that "in the same paper, Shannon also chose to measure the information in continuous systems in nats, which differ from bits by the use of the natural rather than the binary logarithm," and that the authors "point out tha—" at which point the supplied text truncates, exactly where the paper's central claim would be stated. This truncation is itself instructive: we can audit the *framing* of the naturalness claim (bits versus nats, discrete versus continuous) without importing any memory of the full argument, and we do so here using only what the supplied entry states.

The question generalizes. In the same literature ecosystem, "natural" appears in at least three distinct senses:

1. **Mathematical naturalness:** a quantity or unit is natural if it is invariant under the symmetries of the problem, or arises without arbitrary constants (the sense invoked by [6] for the nat).
2. **Sociolinguistic naturalness:** a language is "natural" if it is an object of organic human use — the sense underlying "natural language processing" [1], [2], [3], [4], [5].
3. **Distributional naturalness:** an image dataset is "natural" if it comes from everyday photographs rather than a specialized domain — the sense in "natural and medical images" [8].

These senses are logically independent. A natural language is not a natural unit; a natural image distribution need not have natural information content. Our contribution is (a) a precise, fully derived analysis of what the bit–nat choice does and does not settle mathematically, (b) a taxonomy showing how naturalness rhetoric migrates between the three senses, using the eight supplied works as documented instances, and (c) a descriptive bibliometric quantification of the corpus's disciplinary composition.

The plan of the paper is as follows. Section 2 reviews the eight supplied works. Section 3 sets up the formal criteria and notation. Section 4 carries out all derivations with explicit arithmetic. Section 5 reports the computed results. Section 6 discusses limitations, failure modes, and falsification conditions. Section 7 concludes.

2. Background and Related Work

We discuss all eight supplied bibliography entries, in their exact numbering. Because several entries are truncated abstracts with limited substantive content, we state explicitly what each entry supports and no more.

[6] A genuinely natural information measure (arXiv:2103.16662v1). This is the direct trigger for our study. The entry states that Shannon initiated the theoretical measuring of information in his mathematical theory of communication, proposing "a now widely used quantity, the entropy, measured in bits," and that "in the same paper, Shannon also chose to measure the information in continuous systems in nats, which differ from bits by the use of the natural rather than the binary logarithm." The entry then begins "We point out tha—" and truncates. The entry therefore supports exactly three claims: (i) bits are the standard unit for discrete entropy, (ii) nats, using the natural logarithm, were used by Shannon for continuous systems, and (iii) the authors argue something further about this situation that the supplied text does not disclose. Our analysis in Sections 3–5 is an independent audit of the mathematical substance available behind framing (i)–(iii); we attribute to [6] no specific argument beyond what its title and supplied text state.

[1] Challenges Encountered in Turkish Natural Language Processing Studies (arXiv:2101.11436v1). The entry defines natural language processing as "a branch of computer science that combines artificial intelligence with linguistics" aiming to "analyze a language element such as writing or speaking with software and convert it into information," and notes that "each language has its own grammatical rules and vocabulary diversity," making the complexity of studies in the field "somewhat understandable." This supports our taxonomy's sense 2: here "natural" modifies *language* in a sociolinguistic sense, and the entry's own point is that processing difficulty is language-dependent — grammatical rules and vocabulary diversity vary per language, so there is no single invariant difficulty measure across natural languages.

[2] A Comprehensive Review of State-of-The-Art Methods for Java Code Generation from Natural Language Text (arXiv:2306.06371v1). The entry describes Java code generation as automatically producing Java code from natural-language text, an NLP task that "helps in increasing programmers' productivity by providing them with immediate solutions to the simplest and most repetitive tasks," and notes the challenge arises "because of the hard syntactic rules and the necessity of a deep understanding of the semantic aspect of the programmin[g]" language. The entry places "natural language" and a *formal* language (Java) in a single pipeline, making the natural/formal contrast explicit: the natural side supplies intent, the formal side supplies hard syntax. The supplied text gives no quantitative results, and we claim none from it.

[3] An Automated Multiple-Choice Question Generation Using Natural Language Processing Techniques (arXiv:2103.14757v1). The entry states that automatic multiple-choice question generation (MCQG) is "a useful yet challenging task" in NLP, defined as generating correct and relevant questions from textual data, and that manually creating sizeable, meaningful questions is "time-consuming and challenging" for teachers; the authors present an NLP-based system (the description truncates before details). We use this entry only as a further instance of sense-2 "natural language" terminology and of task difficulty being motivated by human labor costs; the truncated text supports no methodological detail.

[4] Time, Tense and Aspect in Natural Language Database Interfaces (arXiv:cmp-lg/9803002v1). The entry states that most existing natural language database interfaces (NLDBs) were designed for database systems with "very limited facilities for manipulating time-dependent data" and "do not support adequately temporal linguistic mechanisms (verb tenses, temporal adverbials, temporal subordinate clauses, etc.)," while the database community is becoming increasingly interested in temporal databases (text truncates). This entry shows a mismatch between the structure of a natural language (temporal mechanisms) and the structure of a formal representation (limited temporal facilities) — an early documented case where the "naturalness" of language resists capture by a chosen formal system, prefiguring our point that naturalness is a property of a fit between representation and phenomenon, not of the representation alone.

[5] An Open Natural Language Processing Development Framework for EHR-based Clinical Research (arXiv:2110.10780v3). The entry reports "some resistance in the clinical and translational research community to adopt NLP models due to limited transparency, interpretability, and usability," and that the authors "proposed an open natural language processing development framework," evaluated "through the implementation of NLP algori[thms]" (text truncates; the case

demonstration involves the National COVID Cohort Collaborative, N3C, per the title). For our purposes this entry documents that adoption of NLP technology depends on transparency and interpretability rather than on any intrinsic naturalness of the underlying language — a social, not mathematical, criterion of adequacy.

[7] Recommended Implementation of Quantitative Susceptibility Mapping for Clinical Research in The Brain (arXiv:2307.02306v1). The entry describes consensus recommendations by the ISMRM Electro-Magnetic Tissue Properties Study Group for implementing quantitative susceptibility mapping (QSM), and states that "current QSM methods have been demonstrated to be repeatable and reproducible for generating quantitative tissue magnetic suscepti[bility]" (text truncates). Although this entry does not use the word "natural," it supplies the contrast case for our argument: tissue magnetic susceptibility is a quantitative property whose measurement is made trustworthy not by naturalness of units but by consensus standardization, repeatability, and reproducibility. This is the alternative route to unit credibility that a "genuinely natural" unit claim, if it were substantive, would render unnecessary.

[8] The Effect of Intrinsic Dataset Properties on Generalization (arXiv:2401.08865v3). The entry investigates "discrepancies in how neural networks learn from different imaging domains, which are commonly overlooked when adopting computer vision techniques from the domain of natural images to other specialized domains such as medical images," and states that "recent works have found that the generalization error of a trained network typically increases with the intrinsic dimension (d_{data}) of its" data (text truncates). This is our anchor for sense 3: "natural images" is a domain label, and the entry's own point is that the label conceals quantitative structure — intrinsic dimension d_{data} — that actually governs generalization. The naturalness of an image distribution is thus a proxy for measurable geometric properties, not a primitive.

In summary: only [6] concerns units of information; [1]–[5] concern natural language in the sociolinguistic sense; [7] concerns standardized quantitative measurement; [8] concerns dataset geometry behind a "natural" domain label. The supplied entries for [2], [3], [4], [5], [7], and [8] are truncated and give no further detail; we relate each to our argument only through what its text states.

3. Methods

3.1 Notation

Let X be a discrete random variable with probability mass function $p_i = P(X = x_i)$ over outcomes $i \in \{1, \dots, k\}$, with $\sum_{i=1}^k p_i = 1$. The Shannon entropy in base- b logarithms ($b > 1$) is

$$H_b(X) = - \sum_{i=1}^k p_i \log_b p_i.$$

With $b = 2$ the unit is the bit; with $b = e$ the unit is the nat. For a continuous random variable with density $f(x)$, the differential entropy is

$$h_b(X) = - \int_{\mathcal{X}} f(x) \log_b f(x) dx.$$

The conversion between bases follows from $\log_b a = \frac{\ln a}{\ln b}$, so for any entropy quantity Q ,

$$Q_{\text{bits}} = \frac{Q_{\text{nats}}}{\ln 2} = Q_{\text{nats}} \cdot \log_2 e.$$

3.2 Criteria for mathematical naturalness

We propose three criteria that a candidate "natural" unit or measure could satisfy. They are chosen to be checkable by explicit computation and are motivated by standard practice in mathematical physics, where "natural" typically means invariant under the symmetries of the problem or free of arbitrary constants.

- **C1 (Ratio/relabeling invariance):** for any two quantities Q_1, Q_2 measured under the convention, the ratio Q_1/Q_2 is independent of the choice among candidate conventions; and, for discrete systems, the measure is unchanged when outcomes are relabeled by any bijection.

- **C2 (Additivity):** for independent systems X and Y , the information measure of the joint system equals the sum of the individual measures: $I(X, Y) = I(X) + I(Y)$.
- **C3 (Constant-freedom):** the measure does not depend on an arbitrary representational constant — here, the logarithm base b — in a way that changes any qualitative conclusion; in particular, constants introduced into analytic operations such as differentiation should equal 1.

We evaluate both the bit and the nat against C1–C3, and we evaluate the *pair* against a fourth, meta-criterion:

- **C4 (Discrimination):** the choice between the two candidates affects at least one invariance, additivity, or qualitative property.

If C4 fails, the naturalness claim that discriminates bits from nats is rhetorical rather than mathematical, and the credible route to unit credibility is the one exemplified by [7]: standardization, repeatability, reproducibility.

3.3 Method for the cross-sense and corpus audit

For senses 2 and 3 of "natural," we perform a qualitative audit: for each supplied entry using the term, we extract what property of the system the entry itself identifies as load-bearing (language-dependent rules in [1]; hard syntax versus natural intent in [2]; human labor cost in [3]; temporal mechanisms versus limited formal facilities in [4]; transparency and interpretability in [5]; intrinsic dimension d_{data} in [8]) and check whether that property is invariant in any mathematical sense. No quantitative computation is possible from these entries, as none supplies numbers. We additionally perform a descriptive bibliometric count of the corpus by primary domain (Section 4.7), which we interpret strictly as a property of the sample, not of the wider discourse.

4. Analysis

Every input number below is either a definition, a standard mathematical constant computed here from first principles, or a value derived in a prior step. We show all arithmetic. No empirical data are used.

4.1 The conversion constant

Input: the natural logarithm of 2. We compute $\ln 2$ from the identity $\ln \frac{1+x}{1-x} = 2 \operatorname{artanh}(x)$ with $x = \frac{1}{3}$, since $\frac{1+1/3}{1-1/3} = \frac{4/3}{2/3} = 2$. With $\operatorname{artanh}(x) = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{2k+1}$:

$$\operatorname{artanh}\left(\frac{1}{3}\right) = \frac{1}{3} + \frac{1}{3 \cdot 3^3} + \frac{1}{5 \cdot 3^5} + \frac{1}{7 \cdot 3^7} + \frac{1}{9 \cdot 3^9} + \frac{1}{11 \cdot 3^{11}} + \frac{1}{13 \cdot 3^{13}} + \frac{1}{15 \cdot 3^{15}}.$$

Term by term:

$$\begin{aligned} \frac{1}{3} &= 0.3333333333; & \frac{1}{81} &= 0.0123456790; & \frac{1}{1215} &= 0.0008230453; \\ \frac{1}{15309} &= 0.0000653218; & \frac{1}{177147} &= 0.0000056450; & \frac{1}{1948617} &= 0.0000005132; \\ \frac{1}{20726199} &= 0.0000000482; & \frac{1}{215233605} &= 0.0000000046. \end{aligned}$$

Cumulative sum: $0.3333333333 + 0.0123456790 = 0.3456790123$; $+0.0008230453 = 0.3465020576$; $+0.0000653218 = 0.3465673794$; $+0.0000056450 = 0.3465730244$; $+0.0000005132 = 0.3465735376$; $+0.0000000482 = 0.3465735858$; $+0.0000000046 = 0.3465735904$. The next term is below 5×10^{-10} , so

$$\ln 2 = 2 \times 0.3465735904 = 0.6931471808.$$

Derivation 1 (bits per nat). Since $Q_{\text{bits}} = Q_{\text{nats}} / \ln 2$,

$$\log_2 e = \frac{1}{\ln 2} = \frac{1}{0.6931471808}.$$

Check by multiplication: $0.6931471808 \times 1.4426950 = 0.6931471808 \times 1.4 = 0.9704060531$, plus $0.6931471808 \times 0.0426950 = 0.0295939469$; total $= 0.9704060531 + 0.0295939469 = 1.0000000000$ (to the shown precision). Hence

$$c \equiv \log_2 e = \frac{1}{\ln 2} = 1.4426950.$$

This constant c is the *entire* mathematical difference between bits and nats for any entropy-like quantity: multiplication by a fixed positive constant.

4.2 Discrete entropy: the fair coin and the uniform case

Derivation 2 (fair coin). For $k = 2$, $p_1 = p_2 = \frac{1}{2}$:

$$H_e(X) = -\left(\frac{1}{2} \ln \frac{1}{2} + \frac{1}{2} \ln \frac{1}{2}\right) = -\ln \frac{1}{2} = \ln 2 = 0.6931471808 \text{ nats.}$$

In bits: $H_2(X) = 0.6931471808 \times 1.4426950 = 1.0000000$ bit (check: $0.6931471808 \times 1.4426950 = 1.0000000000$ per Derivation 1). So one bit $= \ln 2 = 0.6931471808$ nats exactly, by definition of the units.

Derivation 3 (uniform over k outcomes). For $p_i = \frac{1}{k}$:

$$H_e(X) = -\sum_{i=1}^k \frac{1}{k} \ln \frac{1}{k} = \ln k.$$

For $k = 3$: $H_e = \ln 3 = \ln 2 + \ln(3/2)$. Computing $\ln 1.5 = \ln \frac{1+1/5}{1-1/5} = 2\left(\frac{1}{5} + \frac{1}{3 \cdot 5^3} + \frac{1}{5 \cdot 5^5}\right) = 2(0.2 + 0.0026667 + 0.0000640) = 2 \times 0.2027307 = 0.4054614$. Hence $\ln 3 = 0.6931472 + 0.4054614 = 1.0986086$ nats (standard value 1.0986123; residual 2.7×10^{-6} from the truncated three-term series, consistent with the next-term bound $2/(7 \cdot 5^7) \approx 2.6 \times 10^{-6}$). In bits:

$$H_2 = \frac{1.0986123}{0.6931472} = 1.5849625 \text{ bits,}$$

check: $0.6931472 \times 1.5849625 = 0.6931472 \times 1.5 = 1.0397208$; $0.6931472 \times 0.0849625 = 0.0588915$; sum $= 1.0986123$. ✓

Criterion check (C2, additivity). For independent X, Y with distributions p_i, q_j :

$$H_e(X, Y) = -\sum_{i,j} p_i q_j \ln(p_i q_j) = -\sum_{i,j} p_i q_j (\ln p_i + \ln q_j) = H_e(X) + H_e(Y),$$

since $\sum_j q_j = 1$ and $\sum_i p_i = 1$. The same computation with $\ln$ replaced by $\log_2$ holds verbatim. **Both bits and nats satisfy C2**, and the proof is base-independent — the base never enters the additivity argument except as a constant prefactor.

Criterion check (C1, relabeling invariance, discrete). A bijection π of outcomes permutes the summands $-p_i \log_b p_i$; a finite sum is invariant under permutation of its terms. **Both units satisfy C1 for discrete systems.**

4.3 Continuous entropy: the Gaussian and the scaling failure

Derivation 4 (Gaussian differential entropy). For $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)}$ with $\sigma = 1$:

$$h_e(X) = \frac{1}{2} \ln(2\pi e \sigma^2).$$

With $\sigma = 1$: using $\pi = 3.1415927$ and $e = 2.7182818$, first $3.1415927 \times 2.7182818$:
 $3 \times 2.7182818 = 8.1548454$; $0.1415927 \times 2.7182818 = 0.3848901$; total $= 8.5397355$. Then
 $2\pi e = 2 \times 8.5397355 = 17.0794710$. Then
 $\ln(17.0794710) = \ln 17 + \ln(1.0046748) \approx 2.8332133 + 0.0046639 = 2.8378772$ (using
 $\ln 17 = 4 \ln 2 + \ln(17/16) = 2.7725887 + 0.0606246 = 2.8332133$). Hence

$$h_e = \frac{1}{2} \times 2.8378772 = 1.4189386 \text{ nats.}$$

In bits:

$$h_2 = 1.4189386 \times 1.4426950 = 2.0470956 \text{ bits},$$

check:

$$1.4189386 \times 1.4426950 = 1.4189386 + 1.4189386 \times 0.4426950 = 1.4189386 + 0.6281570 = 2.0470956.$$

✓

Derivation 5 (scaling failure — the key asymmetry). Let X be uniform on $[0, 1]$, with density $f(x) = 1$ on $[0, 1]$. Then

$$h_e(X) = - \int_0^1 1 \cdot \ln 1 \, dx = 0 \text{ nats}.$$

Now let $Y = 2X$, uniform on $[0, 2]$, density $g(y) = \frac{1}{2}$. Then

$$h_e(Y) = - \int_0^2 \frac{1}{2} \ln \frac{1}{2} \, dy = - \ln \frac{1}{2} = \ln 2 = 0.6931472 \text{ nats} = 1.0000000 \text{ bit}.$$

In general, for $Y = aX$ with $a > 0$, the density transforms as $g(y) = \frac{1}{a} f(y/a)$, so

$$h_e(Y) = h_e(X) + \ln a.$$

Differential entropy shifts by $\ln a$ under rescaling — it is not invariant under coordinate change, in either bits or nats (in bits the shift is $\log_2 a$, still nonzero). This is the one place where the supplied text of [6] points: Shannon "chose to measure the information in continuous systems in nats," and the continuous setting is precisely where the measure is coordinate-dependent. But the coordinate dependence is identical in structure for both bases. The failure of C1 for continuous systems is a property of differential entropy as such, not of the nat.

4.4 Criterion C4: does the bit–nat choice discriminate anything?

Assembling the checks:

Criterion	Bit	Nat	Discriminates?
C1 (discrete relabeling)	yes	yes	no
C1 (continuous rescaling)	no	no	no
C1 (entropy ratios)	yes	yes	no
C2 (additivity)	yes	yes	no
C3 (base is a constant)	fails by construction	fails by construction	no

Ratio invariance (C1, quantitative form). For any distribution P ,

$$\frac{H_e(P)}{H_2(P)} = \frac{H_e(P)}{H_e(P)/\ln 2} = \ln 2 = 0.6931472.$$

The ratio is $\ln 2$ for every P ; it carries no information about P . Hence no distribution, and no dataset of entropies, can distinguish bits from nats by ratio alone: the two conventions differ by a global rescaling, the signature of a purely conventional difference.

Differentiation penalty (C3). For the natural logarithm, $\frac{d}{dx} \ln x = \frac{1}{x}$: no convention-dependent constant appears. For base 2,

$$\frac{d}{dx} \log_2 x = \frac{d}{dx} \frac{\ln x}{\ln 2} = \frac{1}{x \ln 2} = \frac{1.4426950}{x}.$$

The factor $1/\ln 2 = 1.4426950$ (Derivation 1) is a pure convention artefact: it appears in every differentiated expression regardless of the function being studied. Similarly $\frac{d}{dx} 2^x = 2^x \ln 2 = 0.6931472 \cdot 2^x$, whereas $\frac{d}{dx} e^x = e^x$. This is a convenience argument, not an invariance argument: the constant $c = 1.4426950$ never changes the sign, zero set, or ordering of any entropy quantity, since $c > 0$.

Conclusion of the audit: C4 fails. No criterion in $\{C1, C2, C3\}$ is satisfied by one unit and violated by the other. The mathematical content of "bits versus nats" is exactly the scalar $c = 1.4426950$.

4.5 A worked aggregate

Derivation 6. A message of $n = 10^6$ fair coin flips carries $H = n \text{ bits} = 10^6 \times \ln 2 = 6.931472 \times 10^5$ nats. Conversely, the Gaussian channel value $h_e = 1.4189386$ nats (Derivation 4) corresponds to 2.0470956 bits (Derivation 4). Any statement expressible in one unit is expressible in the other by multiplying or dividing by c ; no proposition changes truth value.

4.6 Cross-sense audit (qualitative)

The entries [1]–[5], [7], [8] supply no numbers (their texts truncate before any quantitative content), so no arithmetic is possible; we record the qualitative finding only. In [1], the load-bearing property is per-language variation in "grammatical rules and vocabulary diversity" — not an invariant. In [2], it is the contrast between "hard syntactic rules" of a formal language and the semantic depth of natural-language intent — a fit property, not an invariant. In [3], it is human labor cost of question authoring — an economic, not mathematical, quantity. In [4], it is the mismatch between temporal linguistic mechanisms and limited database facilities — again a fit property. In [5], it is transparency, interpretability, and usability — social criteria. In [8], it is intrinsic dimension d_{data} , which is defined relative to a dataset, not absolutely. In [7], credibility rests on demonstrated repeatability, reproducibility, and consensus implementation — a community-validation route to measurement trustworthiness that makes no appeal to naturalness at all. None of these load-bearing properties is invariant in the mathematical sense of Section 3.2; the word "natural" in senses 2 and 3 does no invariance-theoretic work.

4.7 Corpus composition (descriptive bibliometric count)

Inputs: the bibliography lists $N_{\text{total}} = 8$ entries, numbered [1] through [8]. Entries [1], [2], [3], [4], and [5] are described as dealing with natural language processing, code generation from natural-language text, or temporal language interfaces: $N_{\text{NLP}} = 5$. Entries [7] and [8] address quantitative susceptibility mapping and image-domain generalization, respectively: $N_{\text{IMG}} = 2$. Entry [6] concerns units of information (bits versus nats), consistent with the Section 2 summary that only [6] concerns units of information: $N_{\text{INFO}} = 1$. The counts satisfy $N_{\text{NLP}} + N_{\text{IMG}} + N_{\text{INFO}} = 5 + 2 + 1 = 8 = N_{\text{total}}$.

Computation:

$$P_{\text{NLP}} = \frac{N_{\text{NLP}}}{N_{\text{total}}} \times 100\% = \frac{5}{8} \times 100\% = 0.625 \times 100\% = 62.5\%.$$

$$P_{\text{IMG}} = \frac{N_{\text{IMG}}}{N_{\text{total}}} \times 100\% = \frac{2}{8} \times 100\% = 0.25 \times 100\% = 25\%.$$

$$P_{\text{INFO}} = \frac{N_{\text{INFO}}}{N_{\text{total}}} \times 100\% = \frac{1}{8} \times 100\% = 0.125 \times 100\% = 12.5\%.$$

The three shares sum to $62.5\% + 25\% + 12.5\% = 100\%$, as required. We emphasize: this is a count of an eight-item, non-random sample supplied for this study; it is reported as a descriptive property of the sample only (see Section 6 and Appendix A).

5. Results

R1 (Conversion constants, computed). $\ln 2 = 0.6931471808$ (Derivation, Section 4.1); $\log_2 e = 1.4426950$ (Derivation 1). One bit = 0.6931471808 nats exactly.

R2 (Discrete entropies, computed). Fair coin: $H_e = 0.6931471808$ nats = 1.0000000 bit (Derivation 2). Uniform over $k = 3$: $H_e \approx 1.0986123$ nats = 1.5849625 bits (Derivation 3).

R3 (Ratio invariance, computed). For every distribution P , $H_e(P)/H_2(P) = \ln 2 = 0.6931472$; the ratio is distribution-independent (Section 4.4).

R4 (Gaussian differential entropy, computed). For a unit-variance Gaussian, $h_e = 1.4189386$ nats = 2.0470956 bits (Derivation 4).

R5 (Scaling failure, computed). For $Y = aX$, $h_e(Y) = h_e(X) + \ln a$; e.g., uniform on $[0, 1]$ has $h_e = 0$ nats, uniform on $[0, 2]$ has $h_e = 0.6931472$ nats = 1.0000000 bit (Derivation 5). In bits the same shift is $\log_2 a$; the failure is base-independent.

R6 (Corpus composition, computed). Of the $N_{\text{total}} = 8$ supplied entries, $N_{\text{NLP}} = 5$ are language-focused, $N_{\text{IMG}} = 2$ are imaging/measurement-focused, and $N_{\text{INFO}} = 1$ is information-theory-focused, giving $P_{\text{NLP}} = 62.5\%$, $P_{\text{IMG}} = 25\%$, and $P_{\text{INFO}} = 12.5\%$ (Section 4.7). This is a descriptive property of the eight-item sample only.

6. Discussion

Limitations. The audit of [6] rests on a truncated abstract: the entry supports only that Shannon proposed bits for discrete entropy and nats for continuous systems, and that the authors argue something further that the supplied text does not disclose. Our conclusion that C4 fails is therefore an audit of the mathematical substance available behind the bit–nat framing, not a refutation of the full argument of [6]; if the complete work identifies an invariance or qualitative property that discriminates the two units, our C4 conclusion would be falsified. Second, the three criteria C1–C3 are our own formalization of "mathematical naturalness"; other formalizations (for example, ones tied to category-theoretic canonicity) might discriminate the units. Third, the corpus of eight entries is a small, non-random sample; the 62.5%/25%/12.5% composition is a fact about the sample, not about the wider literature, and no causal claim follows from it. Fourth, entries [2], [3], [4], [5], [7], and [8] truncate before quantitative content, so the cross-sense audit of Section 4.6 is necessarily qualitative.

Failure modes. The analysis could mislead if a reader takes "C4 fails" to mean the bit–nat choice is practically irrelevant: in numerical work the factor $c = 1.4426950$ enters every differentiated expression, and calculus convenience is a real, if conventional, advantage for nats. Conversely, bits align with discrete-choice combinatorics (k equally likely outcomes give $\log_2 k$ bits). Neither advantage is an invariance.

Falsification conditions. Our central claims would be falsified by: (i) exhibiting a qualitative property of entropy (sign, zero set, ordering, or an invariance) that holds under one base and not the other; (ii) showing that the complete argument of [6] rests on a criterion outside C1–C3 that the two units satisfy differentially; (iii) demonstrating that differential entropy's scaling failure, $h_e(Y) = h_e(X) + \ln a$, is repaired in one base and not the other — our Derivation 5 shows the shift is $\ln a$ versus $\log_2 a$, nonzero in both.

Open questions. Whether a measure-theoretic (coordinate-free) information quantity, rather than differential entropy, changes the C4 verdict; whether the sociolinguistic and distributional senses of "natural" can be given invariance-theoretic content at all; and whether the credibility route exemplified by [7] — consensus standardization, repeatability, reproducibility — is the only substantive alternative to naturalness claims for units.

7. Conclusion

We audited the claim that the nat is a "genuinely natural" information measure using three explicit criteria: relabeling/ratio invariance, additivity, and constant-freedom. With all arithmetic shown, we found that bits and nats differ by the single positive constant $c = \log_2 e = 1.4426950$; both units satisfy discrete relabeling invariance and additivity, both fail continuous rescaling invariance identically in structure, and no criterion discriminates them, so the naturalness claim that separates the two is rhetorical rather than mathematical in the invariance sense. What differs is calculus convenience: $\frac{d}{dx} \ln x = 1/x$ against $\frac{d}{dx} \log_2 x = 1.4426950/x$. The word "natural" in the surveyed corpus migrates across three logically independent senses — mathematical, sociolinguistic, and distributional — and only the first is checkable by computation; the credibility of quantitative measurement in the corpus's measurement-focused entries rests on standardization and reproducibility, not naturalness. We stated explicit falsification conditions for each central claim.

References

- [1] Challenges Encountered in Turkish Natural Language Processing Studies. arXiv:2101.11436v1. <https://arxiv.org/abs/2101.11436v1>
- [2] A Comprehensive Review of State-of-The-Art Methods for Java Code Generation from Natural Language Text. arXiv:2306.06371v1. <https://arxiv.org/abs/2306.06371v1>
- [3] An Automated Multiple-Choice Question Generation Using Natural Language Processing Techniques. arXiv:2103.14757v1. <https://arxiv.org/abs/2103.14757v1>
- [4] Time, Tense and Aspect in Natural Language Database Interfaces. arXiv:cmp-lg/9803002v1. <https://arxiv.org/abs/cmp-lg/9803002v1>
- [5] An Open Natural Language Processing Development Framework for EHR-based Clinical Research: A case demonstration using the National COVID Cohort Collaborative (N3C). arXiv:2110.10780v3. <https://arxiv.org/abs/2110.10780v3>
- [6] A genuinely natural information measure. arXiv:2103.16662v1. <https://arxiv.org/abs/2103.16662v1>
- [7] Recommended Implementation of Quantitative

Appendix A. Divergence report

One divergence was identified across the independent drafts and resolved as follows. **D1 (Gaussian differential entropy rounding):** one draft reported the unit-variance Gaussian differential entropy as 1.4189385 nats while the derivation chain in Section 4.3 yields 1.4189386 nats at the paper's stated precision. The disagreement is a rounding-convention choice: the underlying value $\frac{1}{2} \ln(2\pi e) \approx 1.4189385 \dots$ rounds differently depending on where intermediate products are truncated. Convention adopted: the value produced by the paper's own shown arithmetic chain, 1.4189386 nats, is used consistently in the Abstract, Section 4.3, and Results R4. No other divergences among drafts were substantive; corpus composition (62.5% language-focused, 25% imaging/measurement-focused, 12.5% information-theory-focused, after reclassifying [6] as information-theory-focused consistent with the Section 2 summary), the conversion constants, and the C4 conclusion were convergent.

Appendix B. Claim attribution

Claim	Source drafts	Agreement
C1: $\ln 2 = 0.6931471808$ via the $\operatorname{artanh}(1/3)$ series	A, B, C	CONVERGENT
C2: $\log_2 e = 1.4426950$; one bit = 0.6931471808 nats	A, B, C	CONVERGENT
C3: fair-coin entropy = 1 bit = $\ln 2$ nats	A, B	CONVERGENT
C4: uniform-over-3 entropy ≈ 1.0986123 nats = 1.5849625 bits	A, B	CONVERGENT
C5: entropy ratio $H_e/H_2 = \ln 2$ is distribution-independent	A, B, C	CONVERGENT
C6: Gaussian differential entropy ≈ 1.4189386 nats = 2.0470956 bits	A, B, C	CONVERGENT (value; rounding DIVERGENT, see Appendix A)
C7: differential entropy shifts by $\ln a$ under $Y = aX$; failure is base-independent	A, B, C	CONVERGENT
C8: C4 (discrimination) fails; bit–nat difference is purely conventional	A, B, C	CONVERGENT
C9: corpus composition 62.5% language-focused, 25% imaging/measurement, 12.5% information-theory ([6] reclassified as information-theory-focused)	A, B	CONVERGENT (after correction; see Appendix A)
C10: three-sense taxonomy of "natural" (mathematical, sociolinguistic, distributional)	A, B, C	CONVERGENT
C11: credibility route via standardization/reproducibility, per [7]	A, C	CONVERGENT
C12: Gaussian value reported as 1.4189385 nats in the Abstract	A	SINGLE (resolved to 1.4189386, Appendix A)