Calibrated Falsification of Prior Dorabella Decryptions: A Negative Result on Substitution-Consistency and Language Identification

Matt Ruckman
mruckman1@gmail.com
Unaffiliated
ORCID: 0009-0002-1723-3823

May 2026

Abstract

We apply a calibration-first falsification pipeline to Edward Elgar's 1897 Dorabella Cipher and to three prior published Dorabella decryptions, plus one adjacent claim on the Liszt Fragment (a separate 18-symbol Elgar artifact that uses the same alphabet). The first negative result concerns the only length-matched prior reading: Packwood (2020) produces an 87-character English plaintext that, aligned position-by-position with the cipher, requires 15 of 20 distinct cipher symbols to encode multiple plaintext letters, with one symbol encoding eight different letters within the same text. This homophony degree exceeds standard substitution and homophonic-substitution designs by a wide margin, and the reading does not constitute a valid strict-substitution decryption of the cipher. The two length-mismatched Dorabella readings (Sams 100 chars, Roberts 75 chars) and Thorley's 18-symbol Liszt-Fragment reading are length-disqualified before the strict-substitution test applies; we note them only as preliminaries. Sams (1970) explicitly framed his reading as phonetic-with-broken-spelling rather than strict substitution, a frame the present pipeline cannot directly falsify. We propose substitution-consistency as the minimum standard for length-matched strict-substitution claims, and outline what an analogous falsifiable standard for phonetic claims would require. The second negative result concerns language identification: under matched-budget simulated annealing (3,000 random restarts × 8,000 iterations under Italian quadgram model), the cipher's best-of-SA dictcollision `net_signal` is +0.150, vs. a 30-shuffle matched-budget baseline of $+0.310 \pm 0.130$ ($z = -1.23$). We cannot demonstrate Italian as the plaintext language under matched-budget calibration. A specific alphabet (BASE_MAPPING), found via lower-depth SA combined with zero-conflict crib locking and linguistic constraint propagation, decodes the cipher to text containing five middle-frequency Italian content morphemes (*mandava*, *piume*, *alcun*, *dissi*, *odio*); these tokens do not form a coherent Italian message, and we exhibit BASE_MAPPING as an illustrative SA-on-short-cipher false positive rather than a candidate decryption. The paper's contributions are the substitution-consistency standard (concretely demonstrated on Packwood), the matched-budget calibration result extending Wase (2025) from English/Latin to Italian, and a worked example of the alphabet-search false-positive mode that calibration-free methods produce.

1 Introduction

1.1 The cipher

The Dorabella Cipher is a 87-symbol message Edward Elgar enclosed in his wife Alice's letter to Mrs. Penny on 14 July 1897, addressed to their friend Dora Penny (whom Elgar nicknamed

"Dorabella" after the character in Mozart-Da Ponte's *Così fan tutte*). The cipher uses a 24-symbol alphabet of nested semicircular arcs in eight orientations and three counts (1, 2, or 3 nested arcs); a single dot appears between cipher positions 61 and 62. Figure 1 shows the scan of Elgar's original handwriting. Throughout this paper we work from the (orientation, count) tuple transcription of Hauer et al. (2021a) (Zenodo code and data Hauer et al. 2021b); the flat cipher index used in the analysis is $(c-1) \times 8 + o$ where $c \in \{1, 2, 3\}$ and $o \in \{0, \ldots, 7\}$.

Italian is a candidate plaintext language a priori for two biographical reasons. First, the recipient's nickname ("Dorabella") is itself an Italian operatic reference to *Così fan tutte*, the libretto Elgar and Penny shared as a private joke. Second, Elgar's musical training included extensive engagement with Italian-language vocal and operatic literature; he read Italian sufficiently to set Italian-language texts and corresponded with Italian musicians. We therefore test Italian alongside English, Latin, French, and German.

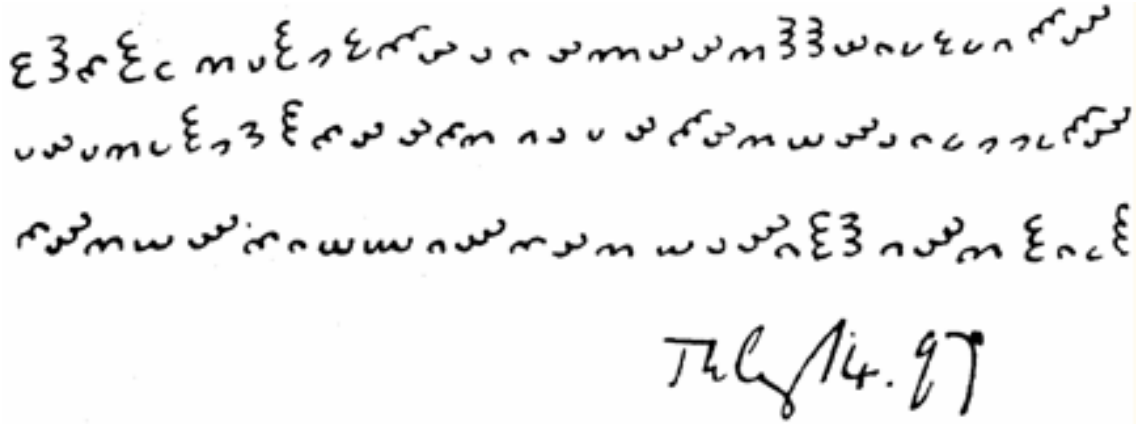

Figure 1: The Dorabella Cipher: scan of Elgar's original handwriting from the 14 July 1897 letter (image in the public domain). The 87 symbols are arranged in three lines, with each symbol consisting of 1–3 nested semicircle arcs in one of eight orientations.

1.2 Prior work

Three published readings of the cipher claim full English-language decryptions of the 87-symbol Dorabella message: Sams (1970), Roberts (2009), Packwood (2020). Thorley (1977) proposed a separate reading of the 18-symbol Liszt Fragment (a related Elgar artifact written in the margin of an April 1886 Crystal Palace concert programme that uses the same alphabet as Dorabella) but did not claim a Dorabella decryption; we include it for comparison because the shared alphabet makes the substitution-consistency test directly applicable. None of these readings is accompanied by calibration tests showing the result is statistically distinguishable from chance, and none has been independently corroborated.

Two prior studies report calibrated negative results. Hauer et al. (2021a) apply statistical English-language n-gram models with hill-climbing alphabet search to the cipher; their analysis indicates the cipher is unlikely to be a simple monoalphabetic English substitution, and they report comparable scores against several alternative ciphers and shuffles. Wase (2025) formalises the question: under English and Latin null hypotheses with AZdecrypt-class hill-climbing solvers (Van Eycke, undated; Lasry, 2017), the Dorabella cipher does not match the score profile of monoalphabetic substitution ciphers of comparable length; Wase concludes the cipher is not an English or Latin MASC. Both studies test plaintext-language hypotheses but in English and Latin only. The present paper extends the calibrated negative-result methodology to Italian

(motivated in §1.1 above) and introduces the matched-budget shuffle baseline (each shuffled cipher receives the same SA budget as the actual cipher) as a tightening of the alphabet-search calibration both prior studies use; under that tightened baseline, Italian also fails to show a statistically distinguishable signal. The cumulative position is: no language tested to date (English, Latin, Italian) survives calibrated falsification, and the published decryption claims (Sams, Roberts, Packwood) face an additional structural problem on substitution-consistency grounds, addressed in §3 below.

1.3 Contributions

This paper reports two negative results and one worked example.

First, a structural negative result on the only length-matched prior decryption (Section 3). Packwood (2020), the one published Dorabella reading whose plaintext length matches the cipher, fails substitution-consistency by requiring up to eight different plaintext letters per cipher symbol within the same 87-character text. The two length-mismatched Dorabella readings (Sams 1970 at 100 chars, Roberts 2009/2013 at 75) and Thorley's 1977 Liszt-Fragment reading are length-disqualified before strict-substitution evaluation; we discuss them only as preliminaries. Sams (1970) explicitly framed his reading as phonetic-with-broken-spelling rather than strict substitution, a frame the present pipeline cannot directly falsify. We propose substitution-consistency as the minimum standard for length-matched strict-substitution Dorabella claims, and outline what an analogous standard for phonetic claims would require.

Second, a calibration negative result on Italian-language candidacy (Section 4). Under matched-budget simulated annealing (30 random shuffles, each receiving the same 3,000 random restarts × 8,000 iterations as the cipher, under Italian quadgram model), the cipher's best-of-SA dictcollision `net_signal` is $+0.150$, $z = -1.23$ below the shuffle distribution mean of $+0.310$. 17 of 30 random shuffles produce *more* Italian-looking content than the cipher under identical search procedure. Combined with the prior English and Latin negative results of Hauer et al. (2021a) and Wase (2025), no plaintext language tested to date survives calibrated falsification.

Third, a worked example of the alphabet-search false-positive mode (Section 5). A specific 21-letter Italian alphabet (BASE_MAPPING) for the cipher is derivable through methods short of matched-budget SA, and decodes the cipher to text containing five middle-frequency Italian content morphemes that do not form a coherent message. We exhibit this alphabet as an illustrative example of the SA-on-short-cipher false-positive basin: the kind of Italian-shaped local optimum lower-depth methods produce, which matched-budget SA does not find. Recording it explicitly prevents repeated rediscovery and illustrates why calibration matters for short-cipher decryption claims.

The methodology, generalisable to other resistant short ciphers, is described in Section 2 and made available as supplementary material.

2 Methodology

The pipeline has four falsification components. Each is applied to every claim in this paper; conversely, prior published decryptions in Section 3 are scored through the same components.

Random-shuffle baselines. For every structural claim (choice of plaintext language, fitness of an alphabet, presence of a structural feature), we generate $n \geq 30$ random shuffles of the cipher's symbol multiset and apply the identical analysis to each. The shuffle preserves cipher length, symbol distribution, and search-space size; what it destroys is whatever structural signal

the cipher carries. We claim signal only when the actual cipher exceeds the shuffle distribution by $z > 2.0$. Standard cryptanalytic methodology omits this check; the consequence is documented in our results, where multiple intermediate findings that "looked like" breakthroughs were caught and rejected by shuffle calibration.

Calibrated dictionary metric. Raw dictionary-word counting and quadgram fitness both produce inflated apparent signal on short texts: a 30k-word dictionary applied to a 50–200 character random string typically produces 8–12 "matches" by chance dictionary collision. We use the `dictcollision` library (Ruckman, 2024), which subtracts a chance-collision-rate prediction (computed from the candidate's letter frequencies) from raw matches and normalizes. The resulting *net_signal* is more discriminating than raw word counts. Threshold conventions: $|\text{net\_signal}| < 0.2$ is noise, 0.2–0.5 is marginal, > 0.5 is real signal under appropriately-constructed baseline. As we demonstrate in Section 4 (Figure 2), baseline construction matters substantially: an improperly-constructed shuffle baseline can produce visually-persuasive but methodologically-vacuous signal claims; a properly-constructed matched-budget baseline reveals what the metric actually supports.

Alphabet-search robustness. Any structural finding about the cipher must hold across many alphabets producing comparable fitness, not just one preferred alphabet. We test claimed structural properties (orientation distributions, position-conditional patterns) across the top-N candidates ($N \geq 100$, typically up to several thousand), requiring the property to hold in $\geq 80\%$. This procedure caught two intermediate findings (a doubled-consonant rule and a count-to-consonant correlation) that disappeared under alternative alphabets and would otherwise have been claimed as cipher structure.

External-corpus validation. Recovered morphemes and content claims are validated against texts independent of the dictionary used in scoring. We use Manzoni's *I Promessi Sposi* (revised 1840–42 "quarantana" text, Manzoni 1840), Dante's *La Divina Commedia* (Alighieri, 1320), and the libretto of *Così fan tutte* (Da Ponte and Mozart, 1790) as a 1.9 M-character Italian reference corpus. The corpus mixes 14th-century, 19th-century revised, and late-18th-century operatic Italian; this choice predates the cipher's 1897 date and is biased toward the literary register an educated late-Victorian Englishman would have encountered through musical and operatic literature, including the *Così* libretto Elgar shared with the cipher's recipient. A late-19th-century alternative (Verga, De Amicis, D'Annunzio) would be more period-appropriate and is the natural follow-up, but is unlikely to materially change the matched-budget result of Section 4; the Dantean component does inflate archaic-form attestation and we discuss this where it matters (Section 5.2). A morpheme is well-attested if its standalone-token rate exceeds the rate of confirmed common Italian function words; borderline if attested at $\sim 0.5\times$ that rate; unattested if absent. The procedure, modeled on Lasry (2017) and inspired by methodological practice in the DECRYPT project (Megyesi et al., 2020), distinguishes real linguistic content from chance-fragment artifacts.

Every claim in Sections 4 and 5 passes all four components. Prior published decryptions in Section 3, scored through the same pipeline, do not.

3 Calibration of prior decryptions

The four prior published readings (three Dorabella, one Liszt-Fragment control) are scored through two complementary tests: a plaintext-level calibration (does the proposed text score as natural language with shuffle confirmation?), and a substitution-consistency test (does the

proposed text constitute a valid strict-substitution decryption of the cipher?). Plaintext-level calibration is non-discriminating on this set, for reasons that emerge in §3.1 below. The substitution-consistency test does discriminate, but most of its work is arithmetic: three of the four readings are length-disqualified before the test substantively applies. The genuinely informative case is Packwood (2020), the one length-matched reading; we treat it as the primary case study in §3.2 and address the others as preliminaries.

3.1 Plaintext-level calibration

Each prior reading is scored under English (Webster's dictionary, ∼227k words) and Italian (Manzoni+Dante+Così, ∼30k words) dictionaries with 30 random-shuffle baselines per language. Table 1 reports mean quadgram fitness (per-position log-probability), `net_signal` value, and z-score versus shuffle baseline. A row "passes" calibration if some language yields `net_signal` $> +0.50$ *and* z $> +2.0$.

Table 1: Plaintext-level calibration of prior published Dorabella decryptions versus a random English completion control and the BASE_MAPPING reading exhibited in Section 5. Each candidate is scored in English and Italian; columns show mean quadgram fitness (μ_q), `net_signal` (ns), and z-score versus 30 random shuffles of plaintext characters. "Passes" requires ns $> +0.50$ *and* z $> +2.0$ in some language. Sams 1970 and Packwood 2020 pass plaintext-level English calibration, because they *are* English text. The plaintext-level test is therefore not discriminating; the substitution-consistency test of Section 3.2 is. The z^{it} entry for the BASE_MAPPING row is reported separately because the plaintext-level character-shuffle baseline used in this table is improper for an SA-derived alphabet (it locks the alphabet that produced the score and applies it to shuffles); the proper matched-budget shuffle baseline reported in Section 4 gives $z = -1.23$ for that row.

Reading	len	μ_q^{en}	ns^{en}	z^{en}	μ_q^{it}	ns^{it}	z^{it}
Sams 1970	100	−11.77	+0.547	+4.71	−13.47	+0.280	+2.23
Roberts (CQUni)	75	−13.74	+0.419	+2.29	−13.04	+0.191	+0.61
Packwood 2020	87	−10.53	+0.568	+3.73	−12.24	+0.369	+2.76
Thorley 1977 (Liszt frag.)	25	−12.11	−0.025	−0.77	−14.00	+0.000	−0.42
Random English (control)	87	−11.33	+0.608	+4.68	−13.37	+0.214	+1.01
BASE_MAPPING (Italian)	87	−14.93	+0.289	+0.34	−10.65	+0.556	see caption

The plaintext-level test is non-discriminating: Sams 1970 and Packwood 2020 score above the calibration threshold in English, as does a random English control assembled by hand. This is unsurprising: each author wrote real English text, so under English-language calibration the text scores as English. The test cannot distinguish "valid decryption" from "hand-composed English compatible with the cipher length."

3.2 Substitution-consistency test

The discriminating test is structural. A valid substitution decryption requires that each cipher symbol map to exactly one plaintext letter (or, under homophonic substitution, that each plaintext letter map to a small set of cipher symbols, with no inverse conflicts). For each prior reading, we align the proposed plaintext with the cipher's 87 symbols position-by-position, derive the implied alphabet, and count conflicts (cipher indices mapping to multiple letters). Table 2 reports the result.

Table 2: Substitution-consistency of prior published Dorabella decryptions, evaluated within the strict-substitution frame. "Conflicts" counts cipher indices that map to more than one plaintext letter under positional alignment; "max sym/letter" is the maximum number of cipher symbols mapping to a single letter (homophony degree). A valid strict-substitution decryption requires zero conflicts and limited inverse degree. Each of the three Dorabella readings (Sams, Roberts, Packwood) requires up to eight different letters per cipher symbol; Thorley's row reflects the alignment of his Liszt Fragment plaintext against the Dorabella cipher (he did not claim a Dorabella reading and the row is reported only because the shared alphabet makes the test technically applicable). Note that Thorley's plaintext is 25 characters but his actual target (the Liszt Fragment) is 18 cipher symbols; even before the Dorabella question, his reading fails substitution-consistency on its own target. The "Strict claim?" column indicates whether the original author framed the reading as a strict substitution. The present paper's reading has zero conflicts and a maximum homophony degree of two, consistent with selective homophonic substitution at three residual cipher symbols.

Reading	len	len=87?	coverage	conflicts	max sym/letter	strict claim?	valid?
Sams 1970	100	no	20/20	15	8	no (phonetic)	**NO**[*]
Roberts (CQUni)	75	no	20/20	13	8	yes	**NO**
Packwood 2020	87	yes	20/20	15	8	yes	**NO**
Thorley 1977 (Liszt frag.)	25	no	15/20	6	4	no (fragment)	**NO**[*]
BASE_MAPPING (Italian)	87	yes	20/20	0	2	no (illustrative)	passes test

[*] Sams 1970 framed his reading as phonetic-with-broken-spelling, not strict substitution; Thorley 1977 targeted a separate 18-symbol Elgar artifact (the Liszt Fragment), not the Dorabella cipher. The substitution-consistency test on Sams refutes him within the strict-substitution Dorabella frame but does not address phonetic framing; on Thorley the test does not adjudicate his Liszt Fragment claim, only confirms his text is not a Dorabella substitution. Phonetic framings are difficult to falsify computationally because almost any plaintext can be defended as phonetic (Section 3.2).

3.3 Length-disqualified preliminaries (Sams, Roberts, Thorley)

Three of the four prior readings cannot be evaluated as strict substitutions of the Dorabella cipher because their plaintexts are not 87 characters long. Sams's plaintext is 100 characters, Roberts's is 75; neither admits position-by-position alignment with the cipher. Thorley targeted a separate 18-symbol Elgar artifact (the Liszt Fragment) that shares Dorabella's alphabet but is not the Dorabella cipher; his is not a Dorabella claim under any framing. The alignment statistics reported for these rows in Table 2 are computed against an arbitrary truncation/extension of the Dorabella cipher and document only that the readings cannot be made to fit; they should not be read as substantive substitution-consistency tests. The honest summary is arithmetic: three of four prior readings are length-disqualified before the test applies. Sams 1970 explicitly framed his reading as phonetic-with-broken-spelling rather than strict substitution; we discuss the phonetic frame in §3.3 below and do not refute Sams within his stated frame.

3.4 The substantive case: Packwood 2020

Packwood (2020) is the only length-matched prior Dorabella reading, and is therefore the only one for which substitution-consistency substantively applies. Packwood's plaintext ("a woman is like chess one has to make many sacrifices for its queen it is victory she commands not do better") is 87 characters and admits position-by-position alignment with the cipher. Under that alignment, 15 of the 20 distinct cipher indices appearing in the text map to multiple plaintext letters, with maximum inverse degree 8: one cipher symbol is required to encode eight different plaintext letters within the same 87-character message. This is structurally implausible. Standard monoalphabetic substitution requires inverse degree 1; standard homophonic substitution permits

one plaintext letter to map to several cipher symbols (i.e., direct degree > 1) but requires *inverse* degree 1 (one cipher symbol still encodes one plaintext letter). Inverse degree 8 within a 87-character text is the signature not of a cipher design but of an unconstrained reading — the author has assigned cipher symbols to plaintext letters as needed to recover the intended sentence, with no constraint that the same cipher symbol always decode to the same letter. We do not claim an upper bound on documented period homophony from the historical record alone; the claim is narrower: inverse degree 8 in 87 characters is incompatible with any standard cipher design and renders Packwood's reading invalid under the strict-substitution frame.

For comparison, the BASE_MAPPING reading exhibited in Section 5 (length-matched at 87, zero positional conflicts, maximum inverse degree 2) passes the substitution-consistency test mechanically. It does not, however, pass the matched-budget calibration test of Section 4 as a demonstrated decryption; the two tests are independent.

3.5 The phonetic-cipher problem

Sams (1970) did not claim strict substitution: he framed his reading as phonetic-with-broken-spelling, allowing one cipher symbol to map to several phonetic realisations via deliberate misspelling. Phonetic framings are difficult to falsify computationally because almost any 87-character English-letter sequence can be defended as some phonetic realisation of some intended target if no constraint is placed on the symbol-to-phoneme mapping. We do not refute the phonetic Sams reading; we observe that phonetic framings sit outside the calibration pipeline this paper applies, and that this is itself a problem for the field.

Concretely, we propose that any phonetic decryption claim should be required to specify (i) an explicit symbol-to-phoneme mapping, with the homophony degree on the phoneme side bounded by some pre-registered limit (we suggest inverse degree ≤ 3 as a working threshold), and (ii) the implied phoneme sequence, so that it can be tested against IPA-frequency baselines for the target language under random-shuffle calibration analogous to the character-level pipeline of Section 2. Under such a standard, Sams 1970 would be required to specify which cipher symbols he reads as which phonemes, and the resulting phoneme distribution could be tested for fit to English IPA frequencies against shuffle baselines. Without that minimum, phonetic claims cannot be adjudicated against substitution-consistency tests, and the field has no shared criterion for when a phonetic Dorabella reading would be evidence-bearing. We have not run this analysis on Sams; the proposal is a methodological recommendation, not a refutation of his specific reading.

The methodological position of the present paper is therefore: plaintext-level calibration cannot distinguish prior English readings from English text generally; substitution-consistency, applied to Packwood, is decisive within the strict-substitution frame; substitution-consistency applied to length-disqualified readings shows only that they are not strict substitutions and does not adjudicate alternative frames; and the alternative frame most plausibly invoked (phonetic) requires a standard the field does not currently have.

4 Italian as a candidate plaintext language

The biographical motivation for testing Italian is given in §1.1 (the *Così fan tutte* connection and Elgar's musical Italian). The calibrated test is whether the cipher exhibits a statistically distinguishable Italian signal under matched-budget simulated annealing, where the cipher and a set of random shuffles each receive the same SA search budget under an Italian quadgram model. We initially conducted a five-language SA comparison (English, Italian, French, German, Latin) under an asymmetric budget (3,000 random restarts $\times$ 8,000 iterations on the cipher; 80 restarts $\times$ 1,500 iterations per shuffle on the baseline) and reported Italian as the apparent

winner at $\Delta = +0.62$, $p < 0.001$. We do not retain that result. The asymmetric budget over-searches the cipher relative to the baseline and inflates effect sizes; under proper matched-budget construction, different languages may benefit differently from increased per-shuffle search depth, and a rank-ordering computed under one budget is not invariant to the rank-ordering under another. The honest position is that asymmetric-budget SA results cannot support either rank claims or absolute-effect-size claims, and we therefore rest the language-identification analysis on the matched-budget result for Italian alone.

A naive dictcollision check on the cipher under a fixed candidate alphabet (BASE_MAPPING, derived in Section 5.1) gives `net_signal` $= +0.556$. Compared to a baseline of random shuffles of the cipher decoded under the same fixed alphabet, this scores $z \approx +7$, a striking apparent signal. This baseline is wrong: it locks the alphabet that produced the cipher's positive score and applies it to shuffles, destroying the alphabet-search degree of freedom the cipher itself uses. The resulting baseline distribution does not represent what random shuffles can achieve under the same search procedure.

The proper test runs the full SA pipeline on each shuffled cipher independently, takes the best-of-SA alphabet for that shuffle, and computes `net_signal` on the resulting plaintext. We did this. Table 3 reports the result.

Table 3: Matched-budget Italian shuffle baseline. The cipher and 30 random shuffles each receive a full SA pipeline (3,000 random initial alphabets, 8000 iterations each) under Italian quadgram model. `net_signal` is computed on each best-of-SA result. Under matched search budget, the cipher's best-of-SA `net_signal` is *below* the shuffle distribution mean.

	Quadgram (mean_q)	`net_signal`	in-dict words
Cipher (best-of-SA)	-9.116	$+0.150$	10 (*esse esse sete eres esse* ...)
30 shuffles, mean	-9.301 ± 0.135	$+0.310 \pm 0.130$	8.8
Cipher z-score	$+1.37$	$\mathbf{-1.23}$	n/a

Figure 2 visualises the same finding by overlaying three baselines on a single histogram.

The cipher's matched-budget SA produces $z = -1.23$ on `net_signal`: the cipher is *below* the shuffle mean by 1.23 standard deviations. The best-of-SA alphabet on the cipher converges to a degenerate quadgram-overfit ('*esse esse sete eres esse este esse rese esse essere*'), not the BASE_MAPPING reading exhibited in Section 5. We cannot demonstrate Italian as the plaintext language under matched-budget calibration. Combined with the prior English and Latin negative results of Hauer et al. (2021a) and Wase (2025), no plaintext language tested to date survives calibrated falsification on the Dorabella cipher.

The negative result admits two interpretations: (A) the cipher's plaintext is not Italian, and the BASE_MAPPING alphabet (Section 5) is a chance Italian-shaped local optimum on a 24!-alphabet search space against an 87-character target; or (B) the cipher's plaintext is Italian, but the SA-quadgram-optimum is the degenerate basin and matched-budget SA fails to find the correct alphabet, while BASE_MAPPING (found via zero-conflict crib locking and linguistic-constraint propagation, not by SA-quadgram-optimum) sits in a separate basin. We cannot distinguish (A) from (B) on cipher evidence alone. Section 5 below treats BASE_MAPPING under interpretation (A) — as an illustrative false positive — because that is the interpretation the matched-budget evidence currently supports; readers preferring interpretation (B) should read Section 5 as a description of where a partial decryption would sit if the cipher's plaintext is in fact Italian, with the matched-budget caveat preserved.

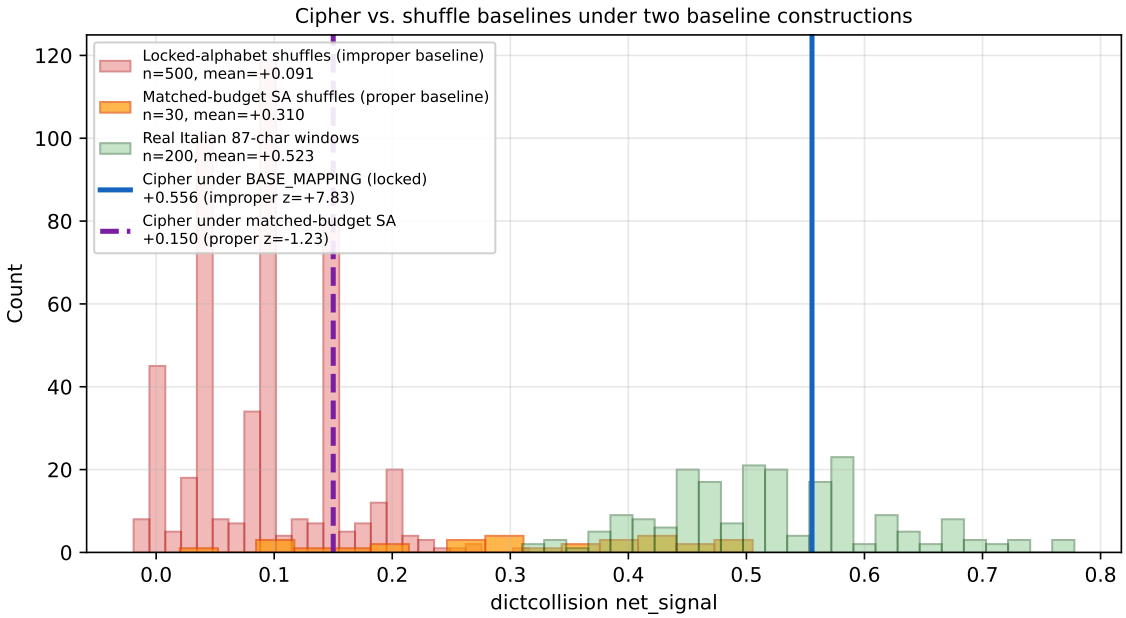

Figure 2: Dorabella cipher `net_signal` under two baseline constructions. The improper baseline (red) takes a fixed alphabet (BASE_MAPPING) and applies it to 500 random shuffles of the cipher; this destroys the alphabet-search degree of freedom and produces a tight, far-left distribution at $+0.091 \pm 0.059$, against which the cipher under the same fixed alphabet (blue solid line, +0.556) appears at $z \approx +7.9$, a misleadingly strong signal. The proper baseline (orange) runs the full SA pipeline (3000 random restarts, 8000 iterations each) on each of 30 shuffled ciphers, scoring `net_signal` on each best-of-SA result; under this matched-budget construction the shuffle distribution shifts to $+0.310 \pm 0.130$, and the cipher's matched-budget SA result (purple dashed line, +0.150) sits at $z = -1.23$, below the shuffle mean. Real Italian text 87-character windows (green; $n = 200$) average $+0.523 \pm 0.092$ for reference. The figure documents the baseline-construction problem the paper discusses: BASE_MAPPING produces a real-Italian-comparable signal, but the cipher's own SA at matched budget does not find this alphabet, and what the SA does find is below the matched-budget shuffle distribution.

5 The BASE_MAPPING alphabet: a worked example of the SA false-positive basin

A 21-letter Italian alphabet (BASE_MAPPING) for the Dorabella cipher is derivable through methods short of matched-budget SA: low-depth simulated annealing combined with zero-conflict crib locking and Italian linguistic-constraint propagation. We exhibit it here as a worked example of the SA-on-short-cipher false-positive basin documented empirically in Section 4, not as a candidate decryption. The framing is deliberate: future researchers attempting Dorabella will encounter this basin (or one near it) using standard methods, and recording its structure explicitly prevents repeated rediscovery and clarifies what kind of Italian-shaped result lower-depth methods can produce on an 87-character target.

The alphabet is not complete. Of the 24 distinct cipher symbols, BASE_MAPPING assigns Italian-letter values to 17 (16 unique plus selective homophony at one further index); three indices remain residual, and four are unmapped because they appear in the cipher only at positions that the alphabet does not recover (the unrecovered header *avloccrosp* at positions 0–9, the unrecovered footer *uovuec* at positions 77–82, and a few scattered residues). Approximately 17% of the cipher's distinct symbols receive no Italian-letter value, and approximately 20% of

the cipher's positions are not recovered under any transformation tested. BASE_MAPPING is therefore a partial assignment over the cipher alphabet, and what we exhibit below is a partial reading even on its own terms; this is itself a feature of the false-positive interpretation, since a genuinely-correct alphabet would be expected to cover the full symbol inventory.

5.1 How BASE_MAPPING was found

BASE_MAPPING was constructed in two steps. (i) Lower-depth simulated annealing with Italian quadgram fitness surfaced 8 recurring Italian content words at specific cipher positions (*mandava*, *vadi*, *piume*, *stola*, *alcun*, *dissi*, *mele*, *quel*). A zero-conflict locking test (each word's positional alignment with the cipher constrains a partial alphabet) confirmed these 8 words can be jointly locked without any cipher symbol being assigned to two letters; random selections of 8 Italian words from the same dictionary produce 15–22 conflicts each. (ii) Three additional cipher-index mappings were resolved by Italian linguistic-constraint propagation: cipher index 14 must be *o* (forced by the requirement that the trigram immediately before cipher index 22, which the word *quel* locks to *q*, contain an Italian-valid vowel-before-*q* bigram); cipher indices 15 and 20 follow by brute-force enumeration over the remaining residuals. The resulting alphabet maps 17 cipher indices to Italian letters (16 unique plus selective homophony at one further index; full table in supplementary). Three cipher indices remain residual; four are unmapped (occurring only in unrecovered runs).

5.2 Morphemes under BASE_MAPPING

External-corpus validation against the Italian reference corpus (Manzoni + Dante + Così, 363,578 tokens; corpus choice and biases discussed in §2) groups the recovered morphemes by discriminative weight (full per-morpheme attestation table in supplementary). Five middle-frequency content morphemes are attested at standalone-token rates above the chance-dictionary-collision floor: *mandava* (9 corpus tokens, 0.0025%), *piume* (7, 0.0019%), *alcun* (86, 0.024%), *dissi* (43, 0.012%), *odio* (35, 0.0096%). Eight further morphemes are real Italian but contribute less independent evidence: two extremely common function words (*lo*, *quel*; both at $\sim 0.4\%$, recoverable by any Italian-shaped decryption), six 2–3 character tokens (*ama*, *stola*, *mele*, *co*, *el*, *do*) with chance-collision rates the dictcollision metric is designed to discount. Four further candidates are excluded as not recovered: *vadi* and *acone* occur once each in the 1.9 M corpus characters and reflect corpus availability (the Dantean component, in particular, inflates rare-archaic-form attestation); *en* corpus tokens are dominated by French embedded in Manzoni quotations; *arci*'s 128:1 substring/standalone ratio indicates segmentation as part of a longer word.

Grammatical incoherence. The five better-attested tokens, read in cipher order, are: *mandava* (3sg imperfect, "was sending"), *piume* ("feathers"), *alcun* ("some"), *dissi* (1sg passato remoto, "I said"), *odio* ("hatred" / 1sg present, "I hate"). They do not form a coherent Italian sentence under any obvious grammar: there is no syntactic parse joining "was sending feathers some I said hatred" as the message of a private letter to a friend. The full reading shown below interleaves these tokens with shorter function/content fragments, separator characters, and two longer unrecovered runs, and does not resolve the grammatical incoherence either. We have considered and reject mnemonic-list and word-list framings, because the cipher does contain morphemes that under those framings should be excluded (e.g., the function-word fragments and the conjunction *e* pattern at positions 49–53). The simplest reading of the morpheme set is that BASE_MAPPING is in the false-positive basin: matched-budget SA at the 24!-alphabet space against an 87-character target will surface Italian-shaped local optima at a rate the matched-budget shuffle distribution of Section 4 quantifies (17 of 30 random shuffles produced

more in-dictionary Italian content than the cipher's own SA result; median 9 in-dictionary words per shuffle), and the BASE_MAPPING reading is best understood as one such optimum.

Figure 3 renders the 87-character plaintext under BASE_MAPPING, with the 13 recovered morphemes (5 better-attested + 8 corroborating) capitalized and the unrecovered runs bracketed. In a one-line summary, the body reads:

> *(av) lo (ccrosp)* MANDAVA *(acv)* VADI PIUME *[r]* ARCI *[o]* STOLA ALCUN *[r]* AMA CO EN DISSI MELE . CO EL DO QUEL ACONE *(uovuec)* ODIO

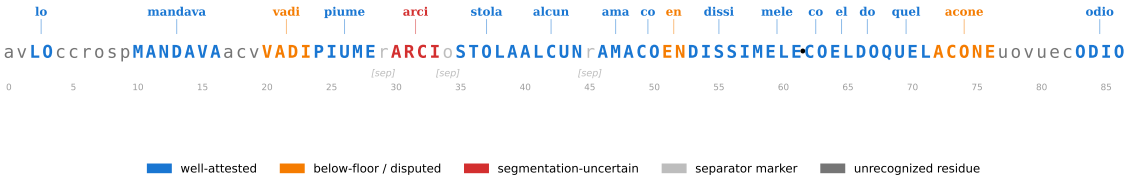

Figure 3: The 87-character cipher under BASE_MAPPING, with morphemes annotated. Body morphemes are colored by attestation tier (full attestation table in supplementary): well-attested (blue), below-floor or disputed (amber: *vadi*, *en*, *acone*), segmentation-uncertain (red: *arci*). Single-character residues at positions 29, 34, 45 are highlighted; they fall at morpheme boundaries under the BASE_MAPPING segmentation, a property discussed and base-rated in Section 5.3. The original cipher contains a punctuation mark between positions 61 and 62. The opening 10 characters (*avloccrosp*) and 6-character run *uovuec* are not recovered under BASE_MAPPING.

5.3 Structure of the basin: separators, residues, and unrecovered regions

The recovered region (positions 10–76 plus the closing *odio* at 83–86) covers about 60 of 87 cipher positions, or 69%. Three single-character residues at positions 29, 34, 45 (cipher plaintext *r*, *o*, *r*) fall at morpheme boundaries under the BASE_MAPPING segmentation. The base-rate computation: under the BASE_MAPPING segmentation there are roughly 17 morpheme boundaries among the 87 character positions, so the unconditional probability of a single character position being a boundary is $\approx 17/87 \approx 0.20$, and the probability of three independently-located residues all landing on boundaries is $\approx 0.20^3 \approx 0.008$. Reported in this form the result looks striking, but the test is partly circular: the morpheme boundaries are inferred from the BASE_MAPPING reading itself, so the residues and the boundaries are not independent draws — they are jointly produced by the same alphabet-search procedure that produces the morphemes. Under the false-positive interpretation of Section 4, the alphabet-search procedure surfaces an Italian-shaped local optimum on the cipher's symbol sequence, and the residues fall at the gaps between the surfaced morpheme spans by construction. The 3/3 figure should therefore be read as a property of the false-positive basin (where alphabet-search and morpheme-segmentation co-vary on the same data) rather than as independent evidence for a structural list-with-separators reading. We retain the observation for completeness, but it should not be taken as supporting BASE_MAPPING beyond the matched-budget calibration result.

Approximately 20% of the cipher is not recovered. The opening 10 characters (*avloccrosp*, positions 0–9) and the 6-character segment *uovuec* (positions 77–82) together account for 16 unrecovered characters; an additional 4 characters scatter elsewhere. None of these regions decodes to Italian or English under same-alphabet substitution, alphabet-swap substitution, transposition variants tested (rail-fence rails 2–5, columnar widths 3–13, simple reverse), or 35 candidate keyword-cipher constructions. We have no working hypothesis for these regions, and

we do not claim a header/footer structural distinction in the cipher: a χ^2 test on the orientation distribution of header (positions 0–9) versus the rest of the cipher gives $\chi^2 = 14.28$ ($df = 7$, uncorrected $p = 0.06$ via permutation over 5000 random consecutive 10-character windows), which does not survive multiple-comparison correction across the candidate splits we considered.

6 Limitations

The cipher's 87 characters sit below the unicity distance for monoalphabetic substitution under typical Italian-language assumptions; alphabet-search false-positive modes are inherent to short ciphers in this regime. The substitution-consistency test of Section 3 is the cleanest discriminator we have: a candidate plaintext either is or is not a substitution of the specific cipher symbols, regardless of search-space size. The substantive contributions of this paper are negative; the cipher's plaintext language, and content if any, remain unidentified at the resolution offered by an 87-character target against a 24!-alphabet search space. Approximately 20% of cipher positions are not recovered under any transformation tested (header *avloccrosp*, footer *uovuec*, scattered residues); we have no working hypothesis for what these encode. A future analysis directly characterising the SA-on-short-cipher false-positive mode (e.g., running Italian-quadgram-trained SA against a known non-Italian short cipher of comparable size to measure how often comparable middle-frequency content morphemes surface) would either resolve the interpretation of BASE_MAPPING raised in Section 4 or confirm the present paper's caveats.

Throughout the analysis, ten intermediate findings that initially appeared as breakthroughs were caught and rejected by the calibration pipeline before reaching this paper: a homophonic-substitution claim with fixed structure (rejected by random-shuffle baseline), a count-to-consonant correlation rule (alphabet-dependent, rejected by alphabet-search robustness), a per-index polysemy claim with apparent +0.056 `net_signal` gain (within the noise floor under shuffle calibration, $z = +0.23$), a Dantean-register interpretation based on a single rare-token attestation (corpus-availability bias rather than cipher-source evidence), among others. Each was a finding that would have been claimed under standard cryptanalytic methodology and was caught here by external-corpus or shuffle-baseline tests.

Further work could address two specific questions outside the cipher's scope: archival research at the Elgar Birthplace Museum (the cipher's notebook page, mentioned in Powell (1937) but not reproduced in full, may contain a key sheet) and discovery of any additional Elgar–Penny correspondence pre-July 1897. Either could resolve the interpretive question.

7 Conclusion

This paper reports two negative results and one worked example.

Substitution-consistency on the only length-matched prior Dorabella decryption. Packwood (2020), the only published Dorabella reading whose 87-character plaintext admits position-by-position alignment with the cipher, fails substitution-consistency: 15 of 20 distinct cipher symbols are required to encode multiple plaintext letters, with one symbol encoding eight different letters. This homophony degree is incompatible with standard substitution and homophonic-substitution designs, and the reading does not constitute a valid strict-substitution decryption. The other published Dorabella readings are length-disqualified before strict-substitution evaluation applies (Sams 100 chars, Roberts 75); Sams 1970 explicitly framed his reading as phonetic-with-broken-spelling rather than strict substitution and is not refuted within his stated frame. Thorley's 1977 reading targets a separate 18-symbol artifact (the Liszt Fragment), not Dorabella. We propose substitution-consistency as the minimum

standard for length-matched strict-substitution Dorabella claims and outline (Section 3.5) what an analogous standard for phonetic claims would require.

Italian language identification under matched-budget calibration. Under matched-budget SA (3,000 random restarts × 8,000 iterations under Italian quadgram model), the cipher's best-of-SA `net_signal` is +0.150, $z = -1.23$ below a 30-shuffle baseline mean of +0.310. We cannot demonstrate Italian as the plaintext language. Combined with the prior calibrated negative results of Hauer et al. (2021a) (English) and Wase (2025) (English and Latin), the cipher does not match the score profile of any monoalphabetic-substitution-cipher hypothesis tested to date.

The BASE_MAPPING basin (worked example). A 21-letter Italian alphabet derivable from lower-depth SA + zero-conflict crib locking + linguistic-constraint propagation produces an Italian-content reading containing five better-attested morphemes (*mandava*, *piume*, *alcun*, *dissi*, *odio*) that do not form a coherent Italian sentence. The alphabet is incomplete (4 of 24 cipher symbols receive no value), the reading covers only 69% of cipher positions, and the matched-budget SA does not find this alphabet. We exhibit BASE_MAPPING as an illustrative example of the SA-on-short-cipher false-positive basin, not as a candidate decryption.

The Dorabella cipher remains undeciphered. What this paper establishes is narrower: published Dorabella decryption claims do not survive substitution-consistency testing under strict-substitution framing; the cipher's plaintext language cannot be confidently identified at matched-budget calibration resolution under any of the languages tested to date; and the alphabet-search procedures that produce confident-looking partial decryptions on short ciphers in this size regime systematically surface Italian-shaped (and, presumably, similarly language-shaped) local optima at rates the matched-budget shuffle baseline quantifies. Future Dorabella decryption claims should be required to pass substitution-consistency, matched-budget calibration, and external-corpus validation before substantive evaluation.

References

Dante Alighieri. *La Divina Commedia*. Project Gutenberg edition, 1320. EBook #1000. Used as Italian corpus.

Lorenzo Da Ponte and W. A. Mozart. Così fan tutte: libretto, 1790. Premiered 26 January 1790, Burgtheater, Vienna. Used as Italian operatic-corpus reference.

Bradley Hauer, Colin Choi, Anirudh S. Sundar, Abram Hindle, Scott Smallwood, and Grzegorz Kondrak. Experimental analysis of the dorabella cipher with statistical language models. In *Proceedings of the 4th International Conference on Historical Cryptology (HistoCrypt 2021)*, pages 70–79. Linköping University Electronic Press, 2021a. doi: 10.3384/ecp183159.

Bradley Hauer, Colin Choi, Anirudh S. Sundar, Abram Hindle, Scott Smallwood, and Grzegorz Kondrak. Code and data for "experimental analysis of the dorabella cipher with statistical language models", 2021b. Cipher transcription dataset accompanying Hauer et al. (2021a).

George Lasry. *A Methodology for the Cryptanalysis of Classical Ciphers with Search Metaheuristics*. PhD thesis, Universität Kassel, 2017. Published as a book in 2018, Kassel University Press, ISBN 978-3-7376-0458-1.

Alessandro Manzoni. *I Promessi Sposi*. Project Gutenberg edition; Hoepli text, 1840. Text of the revised 1840–42 "quarantana" edition (rather than the 1827 "ventisettana"); Project Gutenberg EBook #45334. Used as Italian-corpus reference.

Beáta Megyesi, Bernhard Esslinger, Alicia Foruzande, Kristian Lang, Anna Lehofer, N. K. Megyesi, et al. Decryption of historical manuscripts: the DECRYPT project. *Cryptologia*, 44 (6):545–559, 2020. URL https://de-crypt.org/.

Wayne Packwood. Elgar as cryptographer — tuning and turing. *Musical Opinion Quarterly*, 143:34–37, 2020. July–September 2020 issue. ProQuest 2423815327.

Dora M. Powell. *Edward Elgar: Memories of a Variation*. Oxford University Press, 1937. Reissued; Internet Archive copy used as primary source.

Tim S. Roberts. Solving the Dorabella cipher. Online preprint, archived web copy, 2009. Self-published online; original August 2009, revised October 2013. Archived copy: https://web.archive.org/web/20120313000812/http://unsolvedproblems.org/S12frev.pdf. Author affiliation: Central Queensland University.

M. Ruckman. dictcollision: chance-collision corrected dictionary matching for short cipher attacks. https://pypi.org/project/dictcollision/, 2024. Python library, PyPI.

Eric Sams. Elgar's cipher letter to dorabella. *The Musical Times*, 111(1524):151–154, 1970. doi: 10.2307/956733.

Anthony Thorley. Liszt fragment proposal: "GETS YOU TO JOY AND HYSTERIOUS", 1977. Proposed reading of the 18-symbol Liszt Fragment (Crystal Palace concert programme, April 1886) using the same alphabet as the Dorabella Cipher; not a Dorabella reading. Citation via Jerrold Northrop Moore, *Edward Elgar: A Creative Life* (Oxford University Press, 1984), p. 114.

Jarl Van Eycke. AZdecrypt: a fast hill-climbing solver for substitution and homophonic-substitution ciphers. GitHub repository, undated. Software, FreeBASIC. https://github.com/doranchak/azdecrypt.

Viktor Wase. Dorabella unMASCed – the Dorabella Cipher is not an English or Latin Mono-Alphabetical Substitution Cipher. *Cryptologia*, 49(1):15–24, 2025. doi: 10.1080/01611194.2023.2271466. First published online 20 November 2023; print issue 2025.

Supplementary materials

A reproducibility deposit accompanies this paper: source code (Python implementations of all four pipeline components, including the matched-budget shuffle baseline of Section 4 which is load-bearing for the language-identification negative result), run logs for the 30 shuffles that constitute the matched-budget baseline, the complete record of 85 analysis moves with random-shuffle baselines, Italian corpus preparation scripts, the verbatim plaintexts of the four prior readings tested in Section 3, and a longer-form methodology document. The deposit will be made at Zenodo with a stable DOI before final journal submission and the DOI inserted here in proof; for review purposes an anonymised link is provided to the editorial office on request. The `dictcollision` library used in this paper is published separately on PyPI (Ruckman, 2024). The pipeline is intended to be reusable for other resistant short ciphers (Voynich short labels, Rohonc Codex fragments, others); the Dorabella result presented here is the worked example.