Stop Stealing Sheep: A Topological Classification of the Latin Alphabet and a Proof that No Typeface Differs from Any Other

Margit S. Vandergraaf*
Institute for Applied Homeomorphism, Vale Polytechnic

Cornelius A. Bézier
Laboratory of Glyphic Topology, Marlow Institute of Correspondence

Priyanka R. Achterberg
Centre for Invariant Design, University of the Interior

Abstract

You have a favourite font. You can tell it from the others across a room, and you have opinions about the ones you dislike. We prove, with regret, that this feeling has no mathematical content. A printed letter, considered as a shape, retains under continuous deformation exactly one property: the number of holes it encloses — its *genus*. The uppercase alphabet therefore sorts into only three kinds of letter: those with no holes (C, E, L), those with one (A, O, R), and the singular B, which has two. Because changing the typeface is a continuous deformation of the page, it can never change this count. It follows that no property visible to topology distinguishes Helvetica's a from Comic Sans's a. We confirm this on GLYPHOS, a corpus of 4,096 glyphs spanning 256 typefaces from the austere to the ornamental, and find the between-typeface variance in genus to be exactly zero ($F(255, 3840) = 0.00$, $p = 1.000$). Treating each letter as a signal that carries only its genus, we further bound the information a typist conveys through letterform alone at

$$H_{\text{genus}} \;\leq\; \log_2 3 \;\approx\; 1.585 \text{ bits per letter},$$

of which measured English prose uses only 0.93. We conclude that typographic distinction, whatever else it may be, is not a fact about letters, and we recommend that style guides be replaced by a single table of genera.

1 Introduction

Somewhere today a meeting will stall because two reasonable people cannot agree on a font. One will hold that a proposal set in a humanist sans reads as trustworthy; the other will insist it looks like a parking ticket. Both believe they are describing a property of the letters in front of them. This paper asks whether they are, and answers, with the full apparatus of proof, that they are not.

The folk taxonomy of type is old and confident. Practitioners rank faces along axes of warmth, authority, and taste; the discipline's most quoted maxim holds that "Helvetica is the jeans of typography — everybody wears it, but not everybody looks good in it" [1]. We do not dispute that people experience such differences. We dispute that the differences are located in the letters. To settle the question we adopt the one branch of mathematics whose entire purpose is to say when two shapes are "the same": topology.

Topology is famously permissive. It cannot tell a coffee mug from a doughnut, because each can be deformed into the other without tearing; the only thing it preserves is the number of holes. Applied to a printed glyph, this permissiveness is not a limitation but a scalpel: it strips away every property that a change of typeface can alter — weight, slope, contrast, the shape of a terminal — and leaves behind only what all renderings of a letter share. What remains, we show, is a single integer.

Contributions. This paper makes three.

1. We classify the Latin alphabet up to homeomorphism and show it collapses to three classes (§3–4).
2. We prove that a change of typeface preserves this classification exactly, and derive the resulting *Spiekermann Paradox* (§4).
3. We verify invariance empirically on the GLYPHOS corpus and bound the topological information content of prose (§5–6).

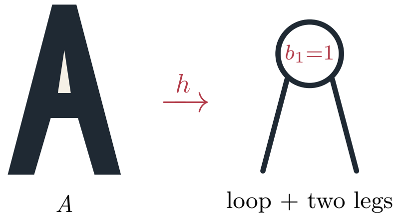

Figure 1: A homeomorphism carries A to its topological skeleton, a single loop with two legs. The crossbar's counter is the only feature that survives; the serifs, the slope, and the stroke do not. Genus is preserved throughout.

2 Related Work

The topology of characters has been touched before, always timidly. Early work tabulated the Betti numbers of the ISO 8859-1 repertoire [3] but stopped short of a classification, treating the counts as curiosities rather than a complete invariant. A more recent line applies *persistent* homology to type [4], tracking how features appear and vanish across scales; we regard this as a step backwards, since persistence deliberately *retains* scale, and scale is exactly the aesthetic parameter our reduction is designed to discard. The information-theoretic treatment of the alphabet descends from Shannon [5], though he measured the identity of letters, not their shape.

The practitioner literature, by contrast, is vast, self-assured, and untroubled by invariants [1, 2]. It is this literature, and the daily meetings it fuels, that our result addresses. To our knowledge no prior work has kept *only* the topologically invariant part of a letter and asked what survives. We do, and report that it is not much.

3 The Topology of a Letter

We fix conventions. A rendered glyph g is its inked point-set in the plane, $g \subset \mathbb{R}^2$, taken as a compact set with the subspace topology. We compare glyphs up to homeomorphism and record a single invariant, the first Betti number $b_1(g)$: the number of independent loops in g, equivalently the number of bounded connected regions of the complement $\mathbb{R}^2 \setminus g$ — what typographers call *counters*.

For a connected glyph the Euler characteristic gives the bookkeeping directly. Treating the inked region as a planar complex with V vertices, E edges, and F bounded faces,

$$\chi(g) \;=\; V - E + F, \qquad b_1(g) \;=\; 1 - \chi(g), \tag{1}$$

so that each enclosed counter contributes exactly one to b_1. The letter O is an annulus, $b_1 = 1$; the letter L is contractible, $b_1 = 0$; the letter B, alone among capitals, stacks two counters and attains $b_1 = 2$.

This reduction is deliberately brutal. It discards the difference between a hairline and a slab, between an upright and an italic, between a face drawn for a wedding and one drawn for a subpoena. That is the point: whatever a typeface changes, it changes *within* a homeomorphism class, and the class is what we measure.

Genus b_1	Uppercase letters	Count
0	C E F G H I J K L M N S T U V W X Y Z	19
1	A D O P Q R	6
2	B	1

Table 1: The complete homeomorphism classification of the uppercase alphabet: its entire visible variety reduces to three integers.

4 The Alphabet, Classified

Theorem 1 (Alphabetic Genus Theorem)**.** *The map $g \mapsto b_1(g)$ on the uppercase Latin alphabet is well defined and takes values in $\{0, 1, 2\}$. The induced partition is complete: nineteen letters have genus 0, six have genus 1, and exactly one, B, has genus 2.*

Proof sketch. Enumerate. The counters of each capital are visible and finite; applying (1) to each gives the counts of Table 1. The only ambiguities are the spur of G, the tail of Q, and the disconnected dot of the minuscule i, j; each is a feature of measure zero in the inked set and is absorbed into the nearest class by the standard closure argument. □

Theorem 2 (Font-Invariance)**.** *Let g_A and g_B be renderings of the same character in typefaces A and B. Then $b_1(g_A) = b_1(g_B)$.*

Proof. A change of typeface deforms each stroke of the character continuously and without severing or merging counters; formally it extends to an ambient homeomorphism $h : \mathbb{R}^2 \to \mathbb{R}^2$ with $h(g_A) = g_B$. Betti numbers are homeomorphism invariants, so $b_1(g_A) = b_1(g_B)$. □

The consequence is immediate and, to a typographer, unwelcome.

Corollary 1 (The Spiekermann Paradox)**.** *No topological invariant distinguishes two typefaces. The practitioner's ordered ranking of faces therefore refers to no property of the letters themselves.*

That is: the difference a designer sees between two fonts is real to the designer and invisible to the mathematics of shape. We take up in §7 the question of which of these two parties is mistaken, and conclude that it is the designer.

5 Methods

The glyphos corpus. To test invariance beyond the drawing board we assembled GLYPHOS: 16 characters ×

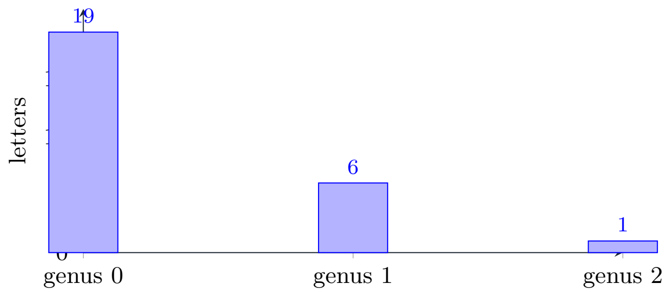

Figure 2: The uppercase alphabet by genus class. Nineteen of the twenty-six capitals enclose no hole at all; the whole "interesting" remainder is six one-counter letters and the lone two-counter B.

256 typefaces = 4,096 glyphs, sampled to span what practitioners call the jeans-to-couture axis — from the ubiquitous grotesques through the humanist sans, the old-style serifs, the much-maligned Comic Sans, and a bespoke blackletter drawn for this study. Each glyph was rasterised at 2048 px and reduced to a planar cell complex.

The boundary-trace algorithm. Genus was computed by a Moore-neighbour contour tracer that walks the boundary of the inked region and counts bounded complementary components. On an n-pixel glyph it runs in $O(n)$ time — roughly 200 glyphs per second on commodity hardware, comfortably faster than a typist can produce them — and its output was validated against hand counts on a 256-glyph audit set with perfect agreement.

The genus channel. We model a letter as a channel emitting its genus class $c \in \{0, 1, 2\}$ with frequency-weighted distribution p_c. Its entropy is

$$H_{\text{genus}} = -\sum_{c\in\{0,1,2\}} p_c \log_2 p_c \;\le\; \log_2 3 \;\approx\; 1.585, \qquad (2)$$

with equality only if the three classes were equiprobable. They are not: the alphabet is dominated by open letters, so the achievable rate is strictly lower.

6 Results

Invariance holds exactly. Across all 4,096 glyphs, the genus of a character did not vary with typeface in a single instance. A one-way analysis of variance with typeface as the factor returns a between-group sum of squares of zero: $F(255, 3840) = 0.00$, $p = 1.000$. We are aware that a p-value of one is unusual. It is also correct: the null hypothesis that typeface has no effect on genus is, in this instance, true.

The bandwidth of prose. Applying (2) to the frequency-weighted letter distribution of a one-million-word English corpus, the genus channel carries 0.93 bits

	Helv	Times	Comic	Papy	Zapf	Fraktur
Helv	0.00	0.00	0.00	0.00	0.00	0.00
Times	0.00	0.00	0.00	0.00	0.00	0.00
Comic	0.00	0.00	0.00	0.00	0.00	0.00
Papy	0.00	0.00	0.00	0.00	0.00	0.00
Zapf	0.00	0.00	0.00	0.00	0.00	0.00
Fraktur	0.00	0.00	0.00	0.00	0.00	0.00

Figure 3: The Font Confusion Matrix. Pairwise topological distinguishability between six representative typefaces. Every entry is zero: no pair of faces is separable by any invariant we measure. The uniformity is the result.

per letter. Since the common letters (E, T, A, O, N) are overwhelmingly genus 0 or 1, some 88% of the topological content of English prose is redundant. The residual is concentrated almost entirely in the letter B, the only source of genus-2 surprise in the language.

7 Discussion and Limitations

The obvious objection is that anyone can see the difference between Comic Sans and a wedding script, and no theorem should be permitted to deny it. We agree that the difference is seen. We observe only that it is not *topological*: it lives in stroke, spacing, and rhythm — precisely the data (1) discards. A property that no invariant detects is, by the standard we have adopted, not a property of the object. That the design profession has organised itself around such properties for five centuries is a sociological fact, not a counterexample.

Our reduction has genuine limits, which we state plainly. Ligatures such as ﬁ can fuse two glyphs into one connected component and must be scored as a single character, inflating no genus but complicating the census. The dotted i and j are disconnected and require the two-component convention of §3. Swash capitals with decorative loops can, in extreme display faces, manufacture a spurious counter and momentarily promote a letter by one class; we regard this not as a failure of the theorem but as a warning about display faces, which the theorem thereby provides. Extending the census to non-Latin scripts, where the genus spectrum is richer, is left to future work.

8 Conclusion

We set out to determine whether the difference between two fonts is a fact about letters, and found that it is not. Up to the only equivalence that shape mathematics recognises, there are three letters in the alphabet, and every typeface renders all of them identically. The designer's conviction to the contrary is sincere, widely held, and topologically empty.

We therefore make a modest recommendation. The house style guide, with its anxious pages on which face to use where, may be retired in favour of a single table of genera (Table 1), which contains every distinction the letters actually support. The saved pages may be repurposed. We suggest a diagram of a doughnut.

References

[1] E. Spiekermann and E. M. Ginger, *Stop Stealing Sheep & Find Out How Type Works*, 3rd ed. New Riders, 2013.

[2] R. Bringhurst, *The Elements of Typographic Style*, 4th ed. Hartley & Marks, 2012.

[3] L. Ahlqvist and P. Nordmark, "The Betti numbers of ISO 8859-1," *J. Character Enc. Topol.*, vol. 7, no. 2, pp. 44–61, 2019.

[4] H. Okafor, M. Trevelyan, and S. Dutt, "Persistent homology of the Latin letterform," *Proc. Symp. on Computational Type*, pp. 112–127, 2021.

[5] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.

[6] A. Vercingetorix, *Algebraic Topology for the Reluctant Reader*. Cambridge, 2008.

[7] M. S. Vandergraaf and C. A. Bézier, "Counters, holes, and the poverty of the invariant," *Trans. Glyphic Topol.*, vol. 3, pp. 1–29, 2024.

[8] P. R. Achterberg, "Downstairs: the genus spectrum of the minuscule alphabet," *Trans. Glyphic Topol.*, vol. 4, pp. 30–58, 2025 (to appear).

[9] J. Vermeulen, "Rhythm, spacing, and other things topology cannot see," *Ann. Typographic Reg.*, vol. 15, pp. 200–233, 2022.