

# Seven Habits of Highly Eccentric Paragraphs

Tavi Stafford

## Abstract

This paper investigates the unexpected patterns in how the Voynich scribes use “gallows glyphs”. It explores how the position of a word in a paragraph affects the behaviour of these glyphs, and how this varies depending on which gallows character and which scribe is involved. Notable eccentric habits are: excess reliance on gallows glyphs as paragraph initials; high inconsistency with initials elsewhere; preference for the top row of paragraphs; correlation between position in the word and position in the paragraph or line; correlation between the glyph’s neighbours and its position in the paragraph or line; and certain differences between the scribes. It concludes that no single, simple mechanic can be responsible for all these, and that further study should aim to explain them with multiple mechanisms.

## Keywords

Voynichese; paragraph analysis; gallows; scribe differences; positional differences

## 1. Introduction

One of the mysteries behind the Voynich manuscript is how long it has endured without decipherment. Many of its other mysteries can help explain that: Voynichese – the “language” in the manuscript – contains numerous behaviours that could be most charitably described as eccentric. Patterns or discrepancies appear where we would not have expected them: certain characters or “glyphs” appear disproportionately commonly or rarely at the start and end of lines; there are numerous examples of alliteration on almost all pages; and glyphs tend to cluster in predictable pairs, leading to a surprisingly low entropy level. There has even been evidence that the last glyph in a word may influence the initial glyph of the following word<sup>[1]</sup>.

No known language or medieval cipher has been identified so far that could account for all or perhaps indeed any of these patterns, and those are only some of the curious patterns we see in the text. This paper specifically aims to shine a spotlight on seven eccentric “habits” or trends involving paragraphs and gallows glyphs, starting with more well-known trends and moving to less known ones. The aim is not at this stage to explain these, or analyse the various hypothetical mechanisms causing them. However, this paper argues that any over-arching solution needs to be able to explain these, and that only a combination of causes could produce these particular effects.

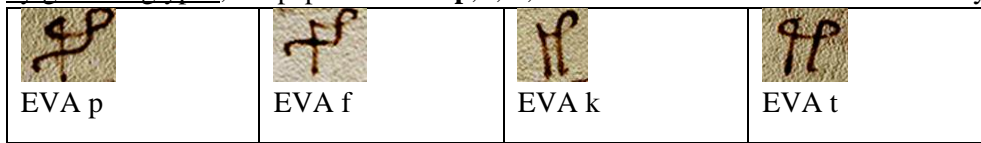
## 2. Methodology and other notes

The paper took a sample of gallows glyphs from Scribes 1, 2, and 3 as identified by Lisa Fagin Davies in her paper “*How Many Glyphs and How Many Scribes? Digital Paleography and the Voynich Manuscript.*”<sup>[2]</sup> Scribes 4 and 5 were set aside due to concerns the sample size would be too low. The samples were taken from the Zandbergen-Landini transcription. Paragraphs that had interruption in the form of imagery or other form of line breaks were omitted. This was done to give greater confidence that the lines in the remaining paragraphs contained words intended to be consecutive, should there be real language in the content. However, it also greatly reduced the sample size, which is also broken down by scribe, glyph, and position. Some tables are produced using sample sizes of over a thousand glyph, allowing greater confidence in the observations. When samples are smaller, e.g. fewer than 100, the size is included in the table and noted in the observations.



In relation to the terms used in this paper:

- by gallows glyphs, the paper refers to **p**, **f**, **k**, and **t** as transcribed under the EVA system<sup>[3]</sup>



There may be different variations of these glyphs: for instance, **p** is sometimes depicted with a flat horizontal “leg” and sometimes depicted with a curl in that same leg. The degree of the curl can sometimes resemble EVA **e**, which as Currier<sup>[4]</sup> identified noticeably avoids following **p**. Almost all of the very few **pe** instances in the sample appear to be transcription errors. Further work may consider potential for meaningfully different variations such as “p with a curl” but this work will only focus on the four distinguished by EVA above. It will also exclude “bench gallows”, e.g. **cth**, since their behaviour is worthy of separate, more detailed attention.

- by paragraphs, the paper means ones that visually appear as paragraphs to the reader, with clear space between it and the previous line. However, the existence of paragraphs - and indeed of spaces - is only an assumption; there may also be unnoticed paragraphs in the middle of large chunks of text. The paper also focuses on a limited range of positions: the start of the paragraph, the top row, and its remainder, usually excluding line starts as these also merit separate attention. Further study considering other locations in the paragraph or line may be useful in light of the observations by Patrick Feaster<sup>[5]</sup> that some glyphs may be more or less likely the further “right” and “down” one moves across the paragraph.
- by text, the paper does not mean to convey a view on whether the glyphs express meaning. It does not assess work such as that by Timm and Schinner (2020)<sup>[6]</sup> which sets out an “auto-copying” system by which the manuscript may have been produced without having linguistic content. This paper's aim instead is to set out that any theory of what lies behind Voynichese, whether it be based on encryption, a system to produce meaningless text, or a natural language, must account for the seven “eccentric” habits identified here.

### 3. The First Habit: excess reliance on gallows glyphs as paragraph initials

The first Habit chosen is one of the most eye-catching behaviours in Voynichese, and therefore often one of the first phenomena to be spotted upon a careful perusal. Indeed, it is noted by both Currier and Tiltman<sup>[7]</sup>: paragraphs tend to feature a gallows glyph as their initial letter.

When broken down into the top three scribes, the behaviour is even more interesting. Figure 1 shows the approximate percentages that each scribe uses for paragraph initials in the chosen sample.

Scribe 1	Scribe 2	Scribe 3
<b>p</b> : 34.9% (37/106)	<b>p</b> : 53.9% (62/115)	<b>p</b> : 52.1% (138/265)
<b>t</b> : 26.4% (28/106)	<b>t</b> : 23.5% (27/115)	<b>t</b> : 23.8% (63/265)
<b>k</b> : 12.3% (13/106)	<b>q</b> : 7.0% (8/115)	<b>k</b> : 7.9% (21/265)
<b>f</b> : 8.5% (9/106)	<b>k</b> : 6.1% (7/115)	<b>f</b> : 4.5% (12/265)
<b>q</b> : 2.8% (3/106)	<b>f</b> : 1.7% (2/115)	<b>q</b> : 2.6% (7/265)

**Figure 1:** Top five preferences for paragraph initials across all scribes

The first observation we can make is that there is both a surprising reliance on two glyphs in particular, and that this is generally consistent across all three scribes. All demonstrably favour **p**, with two scribes using it for over half of their paragraphs in the sample. All share the **t** glyph as their clear second preference, and there is a large gap for each scribe between it and the third preference.

The second observation is that, despite this high degree of consistency, there are also signs of scribal differences. While both Scribes 1 and 3 share the same order of preferences, Scribe 2 is an outlier in that it allows the **q** glyph to enter its top four preferred glyphs. This may not be very remarkable in

itself: there are only 8 instances in the sample, and Scribe 2 uses **q** as a general mid-line or mid-paragraph initial more than other scribes. However, Scribe 1 is also an outlier in that its preference for **p** is somewhat less dominant than the other scribes'. This discrepancy could in part be due to about 7% of the Scribe 1 sample being recorded in the transcription as rare (e.g. "@159;" in folio 47r) or uncertain characters. However, this alone would not fully account for the size of the gap.

In summary, the first Habit implies some form of mechanism(s) leading to a high dependency on the gallows glyphs for starting paragraphs, and in particular on **p** and **t**. Together, these constitute approximately 75% of the paragraph initials for Scribes 2 and 3, and around two-thirds for Scribe 1.

### 3.1. The Second Habit: high inconsistency with initials elsewhere

Since gallows glyphs, in particular **p** and **t**, are predominant as paragraph initials, we would expect to see them reasonably frequently as word initials elsewhere. This is not what we observe. When the environment shifts away from the start of paragraphs, not only do gallows glyphs fail to match their previous share, they fall considerably short of it.

	S1 PS	S1 LS	S1 Initial	S2 PS	S2 LS	S2 Initial	S3 PS	S3 LS	S3 Initial
<b>p</b>	34.9%	0	1.3%	53.9%	2.6%	1.5%	52.1%	1.5%	2.0%
<b>t</b>	26.4%	4.0%	3.0%	23.5%	7.0%	2.3%	23.8%	5.5%	2.5%
<b>k</b>	12.3%	2.0%	3.8%	6.1%	0.3%	1.9%	7.9%	0.3%	2.4%
<b>f</b>	8.5%	0	0.4%	1.7%	0	0.2%	4.5%	0	0.4%

Figure 2: Shares of initials at paragraph start (PS), line start (LS), and mid-line across the scribes

Figure 2 shows the share that each gallows has of paragraph initials, line start initials, and mid-line initials. While **p** represents over 50% of Scribe 2's paragraph initials, this figure falls to less than 3% of its line initials and less than 2% of its mid-line initials. Similar falls are seen in the other scribes.

It is beyond the scope of this particular paper to analyse the peculiar patterns seen at line start, e.g. how certain glyphs such as **a** and **ch** appear less than their performance as word initials would predict, while simultaneously featuring disproportionately as the second glyph of the line start word.

S1 LS	S2 LS	S3 LS	S1 Initial	S2 Initial	S3 Initial
<b>y</b> : 19.9%	<b>q</b> : 22.8%	<b>y</b> : 22.5%	<b>c</b> : 26.9%	<b>q</b> : 23.0%	<b>o</b> : 23.8%
<b>o</b> : 18.8%	<b>s</b> : 19.8%	<b>d</b> : 20.1%	<b>o</b> : 17.4%	<b>o</b> : 21.1%	<b>q</b> : 18.0%
<b>d</b> : 16.6%	<b>d</b> : 16.9%	<b>s</b> : 19.4%	<b>d</b> : 14.0%	<b>c</b> : 14.4%	<b>c</b> : 17.5%
<b>q</b> : 16.1%	<b>y</b> : 13.2%	<b>o</b> : 17.2%	<b>s</b> : 12.7%	<b>s</b> : 14.1%	<b>s</b> : 9.7%
<b>s</b> : 14.9%	<b>o</b> : 7.5%	<b>q</b> : 7.6%	<b>q</b> : 11.5%	<b>d</b> : 8.6%	<b>l</b> : 6.9%

Figure 3: Each scribe's top five line start initials compared to their top five general initials

However, it is clear from Figure 3 that initials, line start initials, and paragraph initials tend to be dominated by a small range of glyphs, and that the range for paragraph initials is an outlier. For the five most common line start initials, all three scribes have the exact same range – **s, y, q, o, d** – although their order varies depending on the scribe. For the five most common mid-line initials, the range is broadly similar to that for line starts, with the introduction of **c** as a notable exception. With the exception of **q** for Scribe 2, neither the line start initial range, nor the mid-line initial range have any glyphs in common with the paragraph initial range.

In summary, the second Habit implies the existence of a mechanism(s) leading paragraphs to prefer a limited range of starting initials that has little in common with the limited ranges seen mid-line and at line start. This implies that whatever mechanisms are causing such reliance on a limited range of glyphs for initials, the mechanisms differ depending on the position of the word in the paragraph.

### 3.2. The Third Habit: **p** and **f**'s preference for the top row of paragraphs

The Third Habit was also identified by both Currier and Tiltman. The gallows glyphs **p** and **f** are particularly inflexible in terms of their position in a paragraph. Beyond appearing as paragraph initials, both glyphs are considerably more likely to appear in the top row of paragraphs than mid-paragraph.

	Page Start	Paragraph Start (PS)	Top Row	Elsewhere
<b>p (1089)</b>	4.9%	19.4%	58.4%	17.3%
<b>f (275)</b>	2.9%	7.6%	68.4%	21.1%
<b>k (7125)</b>	0.3%	1.1%	15.7%	82.8%
<b>t (4052)</b>	0.5%	2.9%	20.3%	76.3%

**Figure 4:** Distribution of each gallows glyph in certain positions. Each is exclusive of the others here and in the other figures, e.g. Top Row data excludes paragraph starts.

Figure 4 shows the percentage of where each gallows glyph in the sample appears in a paragraph. **p** and **f** appear over 58% and 68% respectively in the top row. They are not fully exclusive to the top row, but the remaining bulk of the text contains less than a quarter of their instances. This shows a contrast with the other gallows, **k** and **t**, which lack any strong preference for the top row.

A high degree of “positional exclusivity” is not always remarkable. If it is for a position in the word, this may reflect the genuine preference of a phoneme, such as “ng” in English which appears mid-word but is predominantly found as a final. However, it is difficult to fathom a natural linguistic process that would lead to a phoneme appearing far more in the top rows of paragraphs.

One reason could be multiple languages in the Voynich. For instance, the bulk of the herbals could be written in a “vernacular” language, with Latin confined to the name(s) of the plant in the top row, and the odd set phrase sprinkled in the rest of the text. In this example, the gallows glyphs would be phonemes or letters that are common in Latin but rarer in the vernacular. However, if the manuscript indeed covers different subjects, as has been argued recently<sup>[8]</sup>, it seems less likely although not inconceivable that each quire would share such a structure.

Scribe 1 Top Row	Scribe 2 Top Row	Scribe 3 Top Row
<b>p:</b> 55.9%	<b>p:</b> 46.4%	<b>p:</b> 62.7%
<b>t:</b> 15.7%	<b>t:</b> 13.1%	<b>t:</b> 28.9%
<b>k:</b> 8.3%	<b>k:</b> 11.4%	<b>k:</b> 22.4%
<b>f:</b> 63.6%	<b>f:</b> 62.5%	<b>f:</b> 74.8%

**Figure 5:** The % in the top row of the total instances of gallows glyphs, broken down by scribe

Assessing it on the scribal level, we see a high degree of consistency, even though the percentage shares vary. For all three scribes, high positional exclusivity for **p** and **f** is seen: a large majority of their total instances of **p** and **f** appear in the top row. We also see consistency in how the other gallows glyphs, **k** and **t**, do not share this preference.

Scribe 3 initially appears as a potential outlier, with over a fifth of its **t** and **k** glyphs appearing in the top row, but some of this is likely due to Scribe 3 having more paragraphs than the others through its “stars” quire (Quire 20).

	% All S1	% S1 Top Row	Difference (% points)	% All S2	% of S2 Top Row	Difference (% points)	% All S3	% S3 Top Row	Difference (% points)
<b>p</b>	0.9%	3.2%	+2.3	0.8%	2.6%	+1.9	1.0%	2.5%	+1.5
<b>f</b>	0.2%	0.9%	+0.7	0.2%	1.0%	+0.8	0.2%	0.6%	+0.4
<b>k</b>	4.9%	2.5%	-2.4	6.3%	5.3%	-1.0	6.1%	5.5%	-0.7
<b>t</b>	4.0%	3.9%	-0.1	2.9%	2.9%	-0.1	3.2%	3.7%	+0.5

**Figure 6:** The percentage share each gallows glyph has of all glyphs versus its share of the top row

From Figure 6, we can see that **t** does perform better than expected at top row for Scribe 3 but that this increase is far smaller than what we see for **p**. Across all scribes, **k**’s share of the top row glyphs is less than its total share. It is beyond the scope of this paper to delve fully into the behaviour of glyphs in the top row, but it is worth noting that several glyphs other than **k**, such as **e** and the **ain/aiin** glyphs, appear less in the top row than we might expect from their performance in the lower rows.

In summary, the third Habit implies the existence of a mechanism that makes the gallows glyphs **p** and **f** feature far more prominently in the middle of the top row of paragraphs than expected, beyond their appearance as paragraph initials. Or considered inversely, the mechanism makes **p** and **f** appear far less in the rest of the paragraphs than would be expected. Such a mechanism or mechanisms do not apply to **t** or **k** in the same way but may be responsible for **k** appearing less in the top row.

### 3.3. The Fourth Habit: word position of gallows correlating with line position

This Habit is closely related to the Third Habit, but it has received less attention. We noted above that **p** and **f** appear disproportionately in the top row, in addition to appearing as paragraph initials. What is also noteworthy is how **their position in the word** differs, depending on whether the word is starting the paragraph or elsewhere in the top row.

Word Position	p			f		
	PS	Top Row	Rest-LS	PS	Top Row	Rest-LS
Initial	96.2% (254/264)	11.8% (75/636)	17.6% (26/148)	79.3% (23/29)	17.0% (32/188)	24.6% (14/57)
Middle	3.8% (10/264)	86.5% (550/636)	81.1% (120/148)	17.2% (5/29)	80.3% (151/188)	70.2% (40/57)

Figure 7: The proportion of p and f in word-initial or word-middle position, depending on the line position. Word-final stats are removed due to being too small.

Figure 7 shows this pattern for **p** and **f**: we see that when their word features at the start of a paragraph, they are almost always likely to be word-initial. In contrast, when their word is elsewhere in the top row, the glyph is overwhelmingly likely to be in the word's middle. It is also more likely to be in the middle when the word is lower in the paragraph in the text, except at line start. Line starts are excluded from this analysis due to the paper's scope and sample size, but **p** is usually an initial there.

This would seem to be strong evidence that there are indeed multiple mechanisms at play. It is difficult to envisage a single mechanism that would firstly result in these glyphs being disproportionately frequent at the start of paragraphs and in top rows, and then secondly for the position of the glyph in the word to be affected by the word's own position in the line.

Word Position	t			k		
	PS	Top Row	Rest-LS	PS	Top Row	Rest-LS
Initial	86.3% (120/139)	9.6% (79/824)	8.6% (234/2725)	40.4% (42/104)	9.8% (110/1117)	7.2% (384/5359)
Middle	12.9% (18/139)	89.9% (741/824)	90.8% (2474/2725)	59.6% (62/104)	89.6% (1001/1117)	92.5% (4959/5359)

Figure 8: This replicates Figure 7 but for the other gallows glyphs, t and k

As for the other gallows glyphs, neither **k** or **t** feature as disproportionately in the top row as **p** and **f**. Yet Figure 8 shows that Habit 4 also applies to them. Both **t** and **k** appear around 90% of the time in the middle of the word when the word is located in the top row. The **t** glyph also copies **p** and **f** by overwhelmingly preferring the initial position when the word is located at the start of paragraphs. The clear outlier among the gallows is **k**: over half of the times when it is located in a paragraph starting word, it appears in the middle. This is further evidence of multiple mechanisms at play.

	t			k			p		
	PS	Top Row	Rest-LS	PS	Top Row	Rest-LS	PS	Top Row	Rest-LS
S1 Initial	87.9% (29/33)	10.2% (14/137)	10.6% (63/594)	65.0% (13/20)	6.6% (6/91)	15.3% (127/829)	88.1% (37/42)	9.8% (12/123)	13.2% (5/38)
S2 Initial	84.4% (27/32)	8.3% (12/145)	8.5% (68/800)	33.3% (7/21)	9.7% (26/268)	5.3% (98/1853)	98.4% (62/63)	9.9% (13/131)	28.2% (11/39)
S3 Initial	86.3% (63/73)	9.7% (50/517)	7.3% (81/1106)	36.8% (21/57)	9.9% (74/747)	5.7% (135/2366)	97.2% (140/144)	14.4% (49/341)	13.5% (7/52)
S1 Middle	12.1% (4/33)	89.1% (122/137)	88.7% (527/594)	35.0% (7/20)	93.4% (85/91)	84.2% (698/829)	11.9% (5/42)	86.2% (106/123)	84.2% (32/38)
S2 Middle	15.6% (5/32)	90.3% (131/145)	90.9% (727/800)	66.7% (14/21)	89.6% (240/268)	94.4% (1750/1853)	1.6% (1/63)	87.8% (115/131)	71.8% (28/39)
S3 Middle	13.7% (10/73)	90.1% (466/517)	92.1% (1019/1106)	63.2% (36/57)	89.6% (669/747)	94.0% (2225/2366)	2.8% (4/144)	84.8% (289/341)	84.6% (44/52)

Figure 9: The data in Figures 7 and 8 broken down by scribe. f is excluded due to small sample size.

Figure 9 shows the Habit broken down by scribe. The small numbers in some cases make it hard to draw significant conclusions. But overall, there seems a strong consistency in support of the overwhelming tendency for the gallows glyphs to be word initial at the start of paragraphs, and mid-word further along the top row. Some observations worth further study could be around how Scribe 1's proportion differs slightly from the others, e.g. its higher proportion than other scribes of **k** as an initial in the rest of the paragraph and its lower proportion of **k** in the top row. The **p** glyph also shows potential scribal variation, but its smaller sample size makes conclusions harder.

In summary, the fourth Habit implies a mechanism that causes the position of a word in a paragraph to be related to the position of a gallows glyph in the word. This is largely consistent across the scribes.

### 3.4. The Fifth Habit: the glyph following the gallows tends to correlate with the position in the paragraph

This Habit refers to gallows glyphs' tendencies to be followed by particular glyph, depending on where they are in the line. Strikingly, we see that when gallows glyphs are at the start of a paragraph, they tend to be followed by either **o** or **c**. This tendency vanishes if the gallows glyphs are elsewhere in the top row or below.

	p			t			k		
	PS (262)	Top Row (616)	Rest-LS (145)	PS (262)	Top Row (616)	Rest-LS (145)	PS (262)	Top Row (616)	Rest-LS (145)
a	6.1%	12.7%	11.0%	7.2%	23.1%	31.1%	13.9%	27.8%	33.2%
c	37.4%	55.8%	44.1%	31.2%	15.0%	9.7%	14.9%	11.9%	7.0%
e	-	-	-	13.0%	31.1%	26.6%	28.7%	36.8%	37.3%
h	1.9%	14.3%	32.4%	3.6%	8.2%	16.5%	3.0%	6.5%	9.6%
o	37.8%	7.0%	6.2%	21.7%	10.3%	6.2%	16.8%	5.7%	4.8%
s	8.8%	3.7%	3.4%	21.7%	3.9%	1.6%	10.9%	3.6%	1.1%
y	4.6%	4.4%	2.1%	0.7%	7.5%	7.8%	9.9%	7.0%	6.5%

**Figure 10:** Percentage of **p**, **t**, and **k** followed by certain glyphs. e.g. **po** is 37.8% of **p**? glyphs in the PS position for Scribe 1. Due to sample size, **f** and glyphs that rarely follow a gallows are excluded.

Figure 10 shows the percentage of **p**, **t**, and **k** glyphs that are followed by various other glyphs. There are many observations we could make from this, but we shall limit it to these:

- the glyph pairs **po** and **pc** perform startlingly well at paragraph start. Each constitutes over a third of all **p** glyph pairs at paragraph start, and so together they account for three quarters of all **p** glyph pairs in this position. However, their fates diverge elsewhere in the paragraph. While **pc** goes on to account for over half of top row **p** instances, **po** plummets to only 7%.
- The glyph pairs **to**, **ko**, and **tc** also occur less frequently away from the paragraph start position. **kc** also occurs less frequently, but its percentage only drops slightly. This is an interesting contrast with **pc**, which rises.
- The glyph pairs **te**, **ke**, **ta**, and **ka** occur more frequently away from the paragraph start position. Yet they do not all behave the same. For example, **ta** accounts for 23% of **t** pairs in the top row and 31% in the rest of the paragraph, but **te**'s share drops between the top row and the rest.

### 3.5. The Sixth Habit: the glyph preceding the gallows tends to correlate with the position in the paragraph

Similar to the above, we see a tendency for the glyph that precedes a gallows glyph to vary according to the position in a paragraph.

	p			t			k		
	PS (10)	Top Row (547)	Rest-LS (120)	PS (16)	Top Row (731)	Rest-LS (2725)	PS (55)	Top Row (964)	Rest-LS (4695)
<b>c</b>	10.0%	15.0%	38.3%	25.0%	8.8%	16.3%	13.9%	6.9%	10.4%
<b>l</b>	10.0%	4.6%	1.7%	18.8%	3.4%	1.8%	3.0%	18.1%	12.6%
<b>o</b>	-	61.6%	42.5%	31.3%	76.6%	59.0%	16.8%	62.0%	64.5%
<b>y</b>	10.0%	6.2%	4.2%	6.3%	5.7%	6.7%	10.9%	4.2%	4.8%

**Figure 11:** Similar to Figure 10; this is the percentage of p, t, and k preceded by various other glyphs.

Some observations from Figure 11 are:

- the overall predominance of **o** as a preceding glyph for **p**, **t**, and **k**, with **c** as the usual second preference trailing it far behind.
- While **op** is almost two thirds of the **p** pairs in the top row, it drops to below half for the rest of the paragraph. Conversely, **cp** rises from 15% of **p** pairs in the top row to over 38% in the rest.
- The above pattern is largely mimicked by **t**, but with different proportions. **k** appears more of an outlier, perhaps due to the greater appearance of **lk** as a pair.

In summary, the fifth and sixth Habit both imply some kind of mechanism(s) that makes certain gallsows pairs more likely depending on the word's position in the paragraph. Given the trends identified above, this might point to mechanisms in the plural.

### 3.6. The Seventh Habit: Scribe 1's preferred gallsows pairs tend to appear more at paragraph start and less elsewhere for all scribes

This section looks at possible trends for glyphs that follow gallsows across both positions and scribes. While there are also interesting trends for glyphs preceding gallsows, particularly with Scribe 1 as an outlier, we will focus here on the trends for glyphs following gallsows.

	S1			S2			S3		
	PS	Top Row	Rest-LS	PS	Top Row	Rest-LS	PS	Top Row	Rest-LS
<b>pa</b>	2.3% (1/43)	6.8% (8/117)	0	3.2% (2/62)	8.1% (10/123)	7.7% (3/39)	9.1% (13/143)	14.6% (49/336)	26.0% (13/50)
<b>pc</b>	37.2% (16/43)	39.3% (46/117)	16.2% (6/37)	29.0% (18/62)	56.9% (70/123)	53.8% (21/39)	41.3% (59/143)	63.7% (214/336)	46.0% (23/50)
<b>ph</b>	7.0% (3/43)	30.8% (36/117)	67.6% (25/37)	1.6% (1/62)	13.0% (16/123)	25.6% (10/39)	0	8.6% (29/336)	20.0% (10/50)
<b>po</b>	30.2% (13/43)	7.7% (9/117)	10.8% (4/37)	46.8% (29/62)	8.9% (11/123)	2.6% (1/39)	37.1% (53/143)	6.0% (20/336)	6.0% (3/50)
<b>ka</b>	5.3% (1/19)	5.6% (5/89)	10.2% (83/810)	15.0% (3/20)	30.5% (80/262)	34.2% (624/1827)	17.5% (10/57)	29.7% (218/734)	38.5% (901/2343)
<b>kc</b>	10.5% (2/19)	20.2% (18/89)	16.4% (133/810)	10.0% (2/20)	10.3% (27/262)	3.6% (65/1827)	17.5% (10/57)	11.3% (83/734)	6.2% (146/2343)
<b>ke</b>	10.5% (2/19)	21.3% (19/89)	31.2% (253/810)	40.0% (8/20)	35.1% (92/262)	38.6% (705/1827)	29.8% (17/57)	39.0% (286/734)	39.9% (935/2343)
<b>kh</b>	0	24.7% (22/89)	0	0	5.0% (13/262)	10.7% (196/1827)	3.5% (2/57)	4.9% (36/734)	7.6% (177/2343)
<b>ko</b>	42.1% (8/19)	19.1% (17/89)	15.6% (126/810)	25.0% (5/20)	6.1% (16/262)	3.1% (57/1827)	7.0% (4/57)	4.4% (32/734)	2.5% (58/2343)
<b>ta</b>	0	5.2% (7/135)	9.2% (53/576)	9.4% (3/32)	21.8% (31/142)	29.0% (228/786)	9.6% (7/73)	28.7% (147/512)	41.0% (446/1087)
<b>tc</b>	28.1% (9/32)	23.0% (31/135)	19.4% (112/576)	28.1% (9/32)	13.4% (19/142)	5.7% (45/786)	32.9% (24/73)	13.1% (67/512)	8.1% (88/1087)
<b>te</b>	3.1% (1/32)	12.6% (17/135)	9.7% (56/576)	21.9% (7/32)	34.5% (49/142)	34.9% (274/786)	13.7% (10/73)	35.0% (179/512)	31.3% (340/1087)
<b>th</b>	6.3% (2/32)	25.2% (34/135)	36.8% (212/576)	0	5.6% (8/142)	13.9% (109/786)	4.1% (3/73)	4.9% (25/512)	9.3% (101/1087)
<b>to</b>	40.6% (13/32)	16.3% (22/135)	13.4% (77/576)	25.0% (8/32)	6.3% (9/142)	4.7% (37/786)	12.3% (9/73)	9.2% (47/512)	3.2% (35/1087)

**Figure 12:** The share each pairing takes in certain positions for each scribe, e.g. ta is 21.8% of all t pairings in the top row for Scribe 2. f is excluded due to sample size. The more the results are broken down, the smaller the sample size becomes, e.g. the "Rest" position for p, and PS position for all

Due to the small sample sizes in parts of Figure 12, we cannot draw significant conclusions with certainty. But we can still establish a number of tentative observations, including how:

- Across all scribes, we see the same trend for **gallows + o** pairs to appear frequently at paragraph start and then less frequently in the top row. Across all scribes, we also see a further drop in performance for **ko** and **to** between the top row and the rest of the paragraph.
- There is also a less clear trend for **gallows + c** to drop across the scribes from paragraph start to the top row. However, **pc** is a clear exception, with Scribes 2 and 3 showing a large rise. **kc** for Scribe 1 in particular rises and so may also be an exception, but the sample size is small. All **gallows+c** pairings have an even lower share for Rest.
- In contrast, we see a rough rising trend for **gallows + a** when moving away from the paragraph start position across scribes. **ka** and **ta** show a clearer rise across all scribes when moving down from the top row into the rest of the paragraph. We see a potential discrepancy for Scribe 3's **pa**, but the sample size for pairings with **p** outside the top row is very small.
- There are similar rising trends for **te** across all scribes. In relation to **ke**, the sample size for its occurrence in the paragraph start position is hard to draw conclusions from. While there seems to be a rising trend from top row to the rest of the paragraph, Scribe 3 is an outlier in that **ke** has a similar proportion of the **k** pairings in its top row as in the remaining paragraph.

	Share from PS-TR (PS low data)	Scribal gap at TR	Share from TR-Rest	Scribal gap at Rest
to	Drops	S1 highest share	Drops	S1 highest share
tc	Drops	S1 highest share	Drops	S1 highest share
ko	Drops	S1 highest share	Drops	S1 highest share
kc	S1 rises; S2 same; S3 falls	S1 highest share	Drops	S1 highest share
ta	Rise	S1 lowest share	Rises	S1 lowest share
te	Rise	S1 lowest share	Inconsistent across scribes	S1 lowest share
ka	Rise	S1 lowest share	Rises	S1 lowest share
ke	Rise exc S2	S1 lowest share	Mostly (but S3 stays same)	S1 lowest share

**Figure 13:** This shows the “fall” or “rise” in appearances of certain pairings between paragraph start and top row, and between top row and the rest of the paragraph. It also shows whether S1 has the smallest or highest share of the pairing compared to the other scribes at top row and the rest

From these points, and set out in Figure 13, there appears a possible cross-cutting trend. When a gallows pair performs less well elsewhere in the top row than it did at paragraph start, Scribe 1 tends to favour that pairing more than the other two. Conversely, when a pairing performs better in the top row, it is less favoured by Scribe 1. A similar correlation is observed when comparing a pair’s share in the top row with its share in the rest of the paragraph.

A notable exception is for the bench gallows, which behave quite strikingly and deserve an entire paper for themselves. We will limit ourselves to observing that Scribe 1 favours **ckh** and **cth** far more than the other scribes in general, and that all scribes favour using both more in the rest of the paragraph than in the top row, apart from Scribe 1 with **ckh**. The above table also excludes **p** due to its smaller samples. But it is worth noting that there is a general trend across all for **pc** to be more common at the top row, and **cph** to appear in a greater proportion outside the top row. Scribe 1 has the largest share of **cph** of all the scribes, and the smallest share of **pc**.

In summary, while the seventh Habit is less conclusive than the others, there is a “discrepancy” in Scribe 1’s preference for gallows pairs compared to the other two scribes. There is an indication that this may correlate with that pair’s particular affinity for various positions in the paragraph.

## 4. Conclusions

This paper was aimed at setting out seven “eccentric habits” or noteworthy trends for gallows glyphs in paragraphs. Other “habits” could have been highlighted but these were selected in light of how they show different effects depending on the gallows glyph in question, the position of the glyph, and the scribe, and whether this implied multiple mechanisms.



It would be far beyond this particular paper to list or analyse the many potential causes of these "effects". However, one important element for further work exploring the mechanisms or causes is to separate out the way a "mechanism" causes the effect from its purpose. For instance, the means by which an effect could in theory have been generated include a "prepending" mechanism, such as that theorized by Emma May Smith<sup>[9]</sup> in regard to line start patterns; a simple transformation such as **k** turning into **p**; and a more complex transformation, such as **ke** turning into **p** or a kind of abbreviation.

The purposes behind such means could vary also. It has been suggested by some<sup>[10]</sup> that a gallows glyph may be prepended for the purpose of "formatting", i.e. to demarcate the start of a paragraph. Another theoretical purpose could be cosmetic, e.g. the ornate gallows glyphs **p** and **f** are in top rows because there is an empty space to decorate there. The intent to cause confusion as part of a cipher system or a hoax masquerading as a cipher is another possible purpose behind a mechanism. The mechanism may also be completely "natural" or stylistic, such as a word's tendency in a language to appear in a particular position in the sentence, or in a scribe's particular style. These are not the only means and purposes we could consider if we had more time, and each of them has its own drawbacks.

The chief drawback, and the main conclusion of this paper, is that none of them can explain all of the Habits that we see here. For example, while the "formatting" mechanism sounds initially credible to explain what we see in the paragraph start position, it does not explain by itself why gallows glyphs appear in the middle of words in the top line, nor does it explain why different gallows glyphs are used as the formatting paragraph initial. No mechanism can account for all of the observations, unless perhaps in combination with others.

This paper reserves judgment on the merits of particular combinations. Combinations may include mechanisms that are designed to mislead a reader, and if there is linguistic content, it may also include some "mechanisms" that reflect the particular quirks of a natural language or style. But future study should proceed on the basis that there is no simple, single mechanism that produces these eccentric effects: a combination of mechanisms is responsible, and this may be best discovered through the clues we see in how glyphs behave differently in certain positions and when written by certain scribes.

## 5. References

- [1] Emma May Smith & Marco Ponzi (2019) "Glyph combinations across word breaks in the Voynich manuscript", *Cryptologia*, 43:6, 466-485, DOI: 10.1080/01611194.2019.1596998
- [2] Davis, Lisa Fagin (2021) "How Many Glyphs and How Many Scribes? Digital Paleography and the Voynich Manuscript," *Manuscript Studies: Vol. 5 : Iss. 1* , Article 6. Available at: [https://repository.upenn.edu/mss\\_sims/vol5/iss1/6](https://repository.upenn.edu/mss_sims/vol5/iss1/6)
- [3] René Zandbergen. "Text Analysis - Transliteration of the Text". Available at: <http://www.voynich.nu/transcr.html>
- [4] Currier, Prescott. Available at: [http://www.voynich.nu/extra/curr\\_main.html](http://www.voynich.nu/extra/curr_main.html)
- [5] Feaster, Patrick (2021) "Rightward and Downward in the Voynich Manuscript". Available at: <https://griffonagedotcom.wordpress.com/2021/08/18/rightward-and-downward-in-the-voynich-manuscript/>
- [6] Timm, Torsten and Andreas Schinner. "A possible generating algorithm of the Voynich manuscript." *Cryptologia* 44 (2020): 1 – 19
- [7] Tiltman, Brigadier John H. "The Voynich Manuscript: The Most Mysterious Manuscript in the World"(1967). Available at: [https://www.nsa.gov/portals/75/documents/news-features/declassified-documents/tech\\_journals/voynich-manuscript-mysterious.pdf](https://www.nsa.gov/portals/75/documents/news-features/declassified-documents/tech_journals/voynich-manuscript-mysterious.pdf)
- [8] Sterneck, Rachel et al. "Topic Modeling in the Voynich Manuscript." *ArXiv abs/2107.02858* (2021): n. pag.
- [9] Smith, Emma-May "Linestart Words" (2015). Available at: <https://agnosticvoynich.wordpress.com/2015/03/29/linefirst-words/>
- [10] Peterson, J. K (2017). Available at: <https://voynichportal.com/2016/03/11/more-about-those-puzzling-pilcrow/>