Tock-2 Predictions: Second-Order Event Spaces

Claude (Opus 4.5) and MJC

2026-01-31

Abstract

We make concrete predictions about the second-order Event Spaces that should emerge from the augmented model trained in tick-1. These predictions are testable once training completes. We identify high-value ES-pairs, positional ESs, structural ESs, and within-ES refinements.

1 Setup

After tick-1, the model has:

- Input: 261 dimensions (256 bytes + 5 ES features)
- First-order ESs: Digits, Vowels, Whitespace, Punct, Other
- Training: enwik9 (1B chars), 3 epochs

Tock-2 will extract second-order structure: patterns that depend on ES history, not just current ES.

2 ES-Pair Predictions

The joint space $\text{ES}_{\text{prev}} \times \text{ES}_{\text{next}}$ has $5 \times 5 = 25$ cells. Not all are equally informative.

2.1 High-Value ES-Pairs

Prediction 1 (Dominant ES-Pairs)**.** *The following ES-pairs will have the highest mutual information with next-byte:*

Pair	*Frequency*	*Structure*
Other → Vowel	*High*	*Consonant clusters end; vowel distribution highly non-uniform (context-dependent)*
Vowel → Other	*High*	*Post-vowel consonants are constrained ("th", "ng", "st" patterns)*
Whitespace → Other	*Medium*	*Word-initial consonants (t, w, a, s dominate)*
Other → Whitespace	*Medium*	*Word-final consonants (e, d, s, n, t dominate)*
Punct → Whitespace	*Medium*	*Sentence/clause boundaries*
Digit → Digit	*Low*	*Year patterns (19xx, 20xx)*

Prediction 2 (Low-Value ES-Pairs)**.** *These pairs will show near-uniform within-ES distributions:*

- *Whitespace → Whitespace (rare, mostly uniform)*
- *Vowel → Vowel (rare in English, "aa", "oo" uncommon)*
- *Digit → Vowel (very rare)*

2.2 Quantitative Predictions

Prediction 3 (ES-Pair Entropy Reduction)**.** *For high-value pairs, conditional entropy will be significantly below unconditional:*

Pair	$H(byte\|ES)$	$H(byte\|pair)$	*Reduction*
Other → Vowel	*2.06*	*∼1.2*	*∼0.8 bits*
Whitespace → Other	*4.60*	*∼2.5*	*∼2.1 bits*
Other → Whitespace	*0.32*	*∼0.1*	*∼0.2 bits*

3 Positional ES Predictions

Position within word is a natural second-order ES.

Hypothesis 1 (Word Position ES)**.** *The model will learn implicit representations for:*

- ***Word-initial****: First character after Whitespace*
- ***Word-medial****: Characters with Other/Vowel on both sides*
- ***Word-final****: Last character before Whitespace*

Prediction 4 (Position-Dependent Distributions)**.** *Character distributions vary dramatically by position:*

Position	*Dominant characters*
Word-initial	*t ("the"), a ("and"), w ("was"), s ("she"), capitals*
Word-medial	*e, a, i, o, n, r, s, t (high-frequency letters)*
Word-final	*e (silent e), s (plural), d (past tense), n, t*

Prediction 5 (Positional Neurons)**.** *We will find hidden neurons that activate specifically for:*

- $h_i \approx 1$ *iff position = word-initial*
- $h_j \approx 1$ *iff position = word-final*

These form a 3-way positional ES.

4 Structural ES Predictions (XML)

enwik9 is Wikipedia XML. The model should learn XML-specific structure.

Hypothesis 2 (XML Context ES)**.** *The model will distinguish:*

- ***Tag-name****: Characters after “<” or “< /”*
- ***Attribute-name****: Characters after whitespace inside tag*
- ***Attribute-value****: Characters after “=” inside tag*
- ***Text content****: Characters outside tags*
- ***Entity****: Characters after “&”*

Prediction 6 (Tag-Name Distribution)**.** *After “<”, the character distribution is highly constrained:*

Character	*Probability*
/	∼0.45 (closing tag)
t	∼0.15 (text, title, table)
r	∼0.08 (ref, revision)
p	∼0.06 (page, p)
!	∼0.05 (comment)

This gives $H(char|\text{“}<\text{”}) \approx 2.5$ *bits vs* $H(char) \approx 4.5$ *bits.*

Prediction 7 (Entity Distribution)**.** *After “&”, only a few patterns occur:*

- *“amp;” (ampersand)*
- *“lt;” (less than)*
- *“gt;” (greater than)*
- *“quot;” (quote)*
- *“#” (numeric entity)*

$H(char|\text{“\&”}) \approx 1.5$ *bits.*

5 Within-ES Refinement Predictions

The “Other” ES (205 bytes) is too coarse. It should split.

Hypothesis 3 (Other ES Splits)**.** *The Other ES will naturally factor into:*

1. ***High-frequency consonants****: {t, n, s, r, h, l, d, c, m, f, p, g, b, w, y, v, k}*
2. ***Low-frequency consonants****: {j, x, q, z}*
3. ***Uppercase letters****: {A-Z} \ {A, E, I, O, U}*
4. ***XML special****: {<, >, /, =, &, ;}*

5. ***Other punctuation**: remaining*

Prediction 8 (Sub-ES Sizes).

Sub-ES	*Size*	$H_{\max}$
High-freq consonants	*17*	*4.09 bits*
Low-freq consonants	*4*	*2.00 bits*
Uppercase consonants	*21*	*4.39 bits*
XML special	*6*	*2.58 bits*
Other	*∼157*	*7.29 bits*

6 Bigram Pattern Predictions

Certain bigrams within "Other" are highly predictive.

Prediction 9 (Special Bigrams). *These bigram contexts will show entropy reduction* > 2 *bits:*

Context	*Next char*	*Entropy*
"th"	*vowel (e dominates)*	*∼1.0 bit*
"qu"	*always u*	*0 bits*
"wh"	*vowel (a, e, i, o)*	*∼1.5 bits*
"ch"	*vowel or consonant*	*∼2.0 bits*
"sh"	*vowel (e, o, a)*	*∼1.5 bits*
"ng"	*vowel or space*	*∼1.5 bits*

7 Testing Protocol

7.1 After Tick-1 Completes

1. **ES-Pair Analysis**: Compute $H(\text{byte}|\text{ES}_{\text{prev}}, \text{ES}_{\text{curr}})$ for all 25 pairs
2. **Neuron Probing**: Find neurons that correlate with positional/structural features
3. **Context Analysis**: Run extended context analysis with ES-pair conditioning
4. **Clustering**: Cluster hidden activations to discover implicit sub-ESs

7.2 Success Criteria

Prediction	Test
ES-pairs reduce entropy	Measure $H(\text{byte}|\text{pair}) < H(\text{byte}|\text{ES})$
Positional neurons exist	Find h_i with > 0.8 correlation to position
XML context learned	$H(\text{char}|\text{``}<\text{''}) < 3$ bits
Other ES splits	Clustering gives > 3 coherent sub-groups

8 Expected Outcome

If predictions hold, tock-2 will yield:

- 6–10 high-value ES-pairs (from 25 possible)

- 3 positional ESs (initial, medial, final)
- 4–5 structural ESs (tag, attribute, text, entity, comment)
- 3–5 sub-ESs within Other

Total second-order ESs: ∼15–20, explaining an additional ∼1–2 bpc beyond first-order.

References

[1] Claude & MJC (2026). Tick: Training with Factored Event Spaces. This archive.

[2] Claude & MJC (2026). Quotient Spaces as Bias. This archive.