

Why The Right Event Space Compresses Best

An Explainer for MCP Grafts 30, 31, 32, 33, and 34

Claude and MJC

March 12, 2026

Abstract

The recent MCP grafts add up to one coherent picture. The Universal Model's forward pass is not arbitrary: it is the max-min operation forced by a support lattice. Prediction and compression are the same calculation seen from two directions: the predictor assigns conditional probabilities, and the compressor pays their negative log-probabilities as code length. The remaining question is then architectural: which event space should the model use? The tick-tock fixed-point view says that the right event space is a self-consistent one, discovered by alternating counting with event-space revision. The renormalization view says the same thing in coarse-graining language: good event spaces are the ones that preserve mutual information and make further coarsening expensive. This note explains MCP grafts 30–34 as one thread.

1 The Five Grafts Verbatim

The MCP equations are:

$$L = (\{0, \dots, 255\}, \min, \max, 0, 255) \quad (1)$$

$$a \rightarrow b = \max\{c \in L : \min(a, c) \leq b\} \quad (2)$$

$$\min(a, c) \leq b \iff c \leq (a \rightarrow b) \quad (3)$$

$$F_j(t) = \max_i \min(t_i, p_{ij}) \quad (4)$$

$$\hat{P}(o | i) = \frac{c(i, o)}{c(i, \cdot)} \quad (5)$$

$$L(\hat{P}, D) = -\frac{1}{N} \sum_{t=1}^N \log_2 \hat{P}(d_t | \text{context}_t) \quad (6)$$

$$|C| \approx N \cdot L(\hat{P}, D) = N \cdot \text{bpc} \quad (7)$$

$$R(E) = \log_2 |I|, \quad D(E) = H(O) - I(I; O) \quad (8)$$

$$\Phi(E) = \Theta_2(\Theta_1(E, D), K) \quad (9)$$

$$\Phi(E^*) = E^* \quad (10)$$

$$\text{bpc}(E^*) = H(O | I_{E^*}) \quad (11)$$

$$s(e) = 0 \wedge s(\bar{e}) = 0 \quad (12)$$

$$s(e) = 0 \wedge s(\bar{e}) > 0 \quad (13)$$

$$\min(0, p_{ij}) = 0, \quad \min(t_i, 0) = 0 \quad (14)$$

$$c'(e'_1, e'_2) = \sum_{e_1 \in \phi^{-1}(e'_1)} \sum_{e_2 \in \phi^{-1}(e'_2)} c(e_1, e_2) \quad (15)$$

$$I(E') \leq I(E) \quad (16)$$

$$\beta(k) = I(E_k) - I(E_{k+1}) \quad (17)$$

$$\text{bpc}(k) = H(O) - \sum_{j=0}^{k-1} \beta(j) \quad (18)$$

2 The Forward Pass Is Forced

Graf 30 starts with a support lattice

$$L = (\{0, \dots, 255\}, \min, \max, 0, 255).$$

Think of each support value as a discretized amount of evidence. The meet min means the weaker premise limits the conclusion, and the join max means keep the strongest available support.

The residual

$$a \rightarrow b = \max\{c : \min(a, c) \leq b\}$$

is the algebraic version of implication. The adjunction law

$$\min(a, c) \leq b \iff c \leq (a \rightarrow b)$$

is what makes support propagation and implication fit together.

Proposition 1. *Once support lives in this lattice, the forward pass*

$$F_j(t) = \max_i \min(t_i, p_{ij})$$

is the natural existential inference rule: each input pattern contributes up to the weaker of its support and its pattern strength, and the output keeps the strongest such contribution.

The first point of the thread is that the UM update rule is not arbitrary. It is algebraically forced by the support semantics.

3 Prediction and Compression Are The Same Quantity

Graf 31 changes viewpoint. Given counts,

$$\hat{P}(o | i) = \frac{c(i, o)}{c(i, \cdot)},$$

the model becomes a predictor. Its loss is

$$L(\hat{P}, D) = -\frac{1}{N} \sum_t \log_2 \hat{P}(d_t | \text{context}_t),$$

which is exactly the bpc.

But the same number is also the compressed size per symbol:

$$|C| \approx N \cdot \text{bpc}.$$

Remark 1. *Prediction and compression are not merely related. In the UM they are the same computation read in two directions.*

If the model assigns high probability, the code is short. If it assigns low probability, the code is long. So whenever we ask which event space predicts best, we are also asking which event space compresses best.

4 Event Spaces Live On A Rate-Distortion Curve

An event space is not free. A richer input partition uses more bits to specify context:

$$R(E) = \log_2 |I|.$$

Its benefit is lower residual uncertainty:

$$D(E) = H(O) - I(I; O).$$

This is the rate-distortion tradeoff of event-space design. A very small event space is cheap to describe but weak for prediction. A very fine event space is highly predictive but expensive and possibly wasteful.

5 Tick-Tock Finds A Self-Consistent Event Space

Graf 32 packages the event-space search as an operator:

$$\Phi(E) = \Theta_2(\Theta_1(E, D), K).$$

First count with the current event space. Then discover the best new event space from those statistics, subject to complexity budget K .

A fixed point satisfies

$$\Phi(E^*) = E^*.$$

That means the event space is self-consistent: when you count through it and then re-optimize, you get the same event space back.

The associated prediction loss is

$$\text{bpc}(E^*) = H(O | I_{E^*}).$$

Proposition 2. *The fixed point is the natural stopping condition for event-space search. Before that point, the data is still telling you that your partition is wrong or unnecessarily crude.*

6 Zero Support Does Not Mean False

Graf 33 inserts an epistemic correction that matters for compression. If

$$s(e) = 0 \quad \text{and} \quad s(\bar{e}) = 0,$$

the system is ignorant. It has no evidence either way.

Only when

$$s(e) = 0 \quad \text{and} \quad s(\bar{e}) > 0$$

do we have evidence against e .

That is why the update rule uses cautious propagation:

$$\min(0, p_{ij}) = 0, \quad \min(t_i, 0) = 0.$$

No support in means no support out. No pattern means no inference.

Remark 2. *This is not the same as assigning probability zero. It is better read as: the model has no grounded basis to commit here.*

7 Renormalization Explains The Factorization Tower

Graf 34 reframes event-space search as coarse-graining. A coarsening

$$\phi: E \rightarrow E'$$

pushes counts forward by summing over fibers:

$$c'(e'_1, e'_2) = \sum_{e_1 \in \phi^{-1}(e'_1)} \sum_{e_2 \in \phi^{-1}(e'_2)} c(e_1, e_2).$$

Mutual information can only go down:

$$I(E') \leq I(E).$$

The discrete beta-function

$$\beta(k) = I(E_k) - I(E_{k+1})$$

measures how much predictive information is lost at each coarse-graining step. Then

$$\text{bpc}(k) = H(O) - \sum_{j < k} \beta(j)$$

says that every lost bit of mutual information shows up directly as worse compression.

Proposition 3. *A good event space is one where further coarsening becomes expensive. That is exactly what a natural factorization means in RG language.*

8 Why These Five Grafts Belong Together

The thread is now tight:

1. Graf 30 says what support and inference are.
2. Graf 31 says why better inference is better compression.
3. Graf 32 says how to search for the right event space.
4. Graf 33 prevents the search from collapsing ignorance into falsehood.
5. Graf 34 explains the whole search as an RG flow over event spaces.

Together they answer one question:

What kind of event space gives the best compression without inventing unsupported structure?

9 Why This Prepares Compression Work

If we are about to compress, the practical questions are:

1. what event space should the compressor count over?
2. how much predictive information does each refinement buy?
3. when should we stop refining?
4. how should we treat unseen events?

Grafts 30–34 answer exactly those questions. They do not yet specify a particular compressor implementation, but they constrain what a good one must do.

10 Conclusion

The event-space optimization thread says that compression quality is ultimately a question of representation. The model's arithmetic, epistemics, and coarse-graining strategy all matter because they decide which distinctions are preserved and which are thrown away.

That is why this is the right explainer to write before switching to compression work. It isolates the design principles we will need when turning the MCP into an actual compressor.

References

- [1] Claude and MJC. *MCP*. Working manuscript, March 2026.
- [2] Claude and MJC. *Algebraic Semantics of the Universal Model*. Hutter archive, 12 February 2026.
- [3] Claude and MJC. *The Compression–Prediction Duality in Universal Model Terms*. Hutter archive, 12 February 2026.
- [4] Claude and MJC. *The Fixed-Point Theorem for the Universal Model*. Hutter archive, 12 February 2026.
- [5] Claude and MJC. *No Support Is Not Disbelief*. Hutter archive, 12 February 2026.
- [6] Claude and MJC. *Renormalization and the Factorization Tower*. Hutter archive, 12 February 2026.