Why Do Humans Love High-Density Information Flow?

NeoThinksPhilosophy • 21 Sep 2025

Why Do Humans Love High-Density Information Flow?

Late last week, I found myself scrolling through Hacker News well past midnight. Each short piece felt like a condensed information capsule—sometimes the core conclusion of a research paper, other times a sharp take on tech trends from a programmer. I simply couldn't stop, and before I knew it, it was 2 AM.

That "can't put it down" feeling is actually everywhere in life. When watching *Rick and Morty*, you never know what's coming next: one second Rick's tinkering with a wild invention, the next the timeline warps due to some accident. Those absurd twists keep you on the edge of your seat, hooked. In just a few minutes, the dense plot turns and mind-bending ideas glue viewers to the screen.

On the flip side, scrolling through certain short videos leaves you feeling empty, even though your fingers were moving a mile a minute. It's like you watched a lot, but remembered nothing. We say someone is "hard to talk to" usually because conversations with them circle around trivial gossip or repetitive complaints—never a fresh thought. And a show feels "draggy" mostly because the dialogue is filled with fluff, and you can guess the ending after just two episodes—no surprises at all.

We all know the difference between "interesting" and "boring," but we rarely stop to ask why. Or say, **what's the quantitive definition for "interesting" v.s. "boring"**. It wasn't until I revisited the meaning of "information" that I realized: what humans really crave is *high-density information flow*. The things we find "impressive" or "nutritious" are essentially carriers of more valuable information.

In information theory, **Shannon entropy** is a key measure of uncertainty in a random variable. Its formal definition goes like this: For a discrete random variable X with possible values $\{x_1, x_2, ..., x_n\}$ and probability distribution $P(X = x_i) = p_i$ $(i = 1, 2, ..., n)$, the Shannon entropy $H(X)$ is calculated as:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

This formula shows that the more uniform the probability distribution of the random variable's values, the higher the entropy. Conversely, when the probability of one value approaches 1, the entropy approaches 0. For example, flipping a fair coin (with a 0.5 chance of heads and 0.5 for tails) gives an entropy of $H(X) = -(0.5 \log_2 0.5 + 0.5 \log_2 0.5) = 1$ bit. But a rigged coin that always lands heads ($p_{heads} = 1, p_{tails} = 0$) has an entropy of 0.

Within the framework of **Bayesian theory**, entropy expands into *posterior entropy*, which measures how the uncertainty of a random variable X changes after new evidence E is obtained. The posterior entropy $H(X|E)$ is defined as:

$$H(X|E) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(X = x_i, E = e_j) \log_2 P(X = x_i | E = e_j)$$

Here, $P(X = x_i, E = e_j)$ is the joint probability that both event $X = x_i$ and evidence $E = e_j$ occur, while $P(X = x_i|E = e_j)$ is the conditional probability of $X = x_i$ given $E = e_j$. By comparing the prior entropy $H(X)$ (uncertainty before evidence) and posterior entropy $H(X|E)$ (uncertainty after evidence), we can quantify how much new evidence reduces uncertainty.

From Shannon entropy to Bayesian extensions, it's clear that valuable information isn't just about the probability distribution of symbols. What matters is whether it helps us *improve our ability to predict the world*. Take the sentence, "Tesla will launch a new product tomorrow, so its stock might rise." In Bayesian terms, this new evidence drastically reduces uncertainty about stock market trends—we can update our predictions for supply chains, stock prices, and related events based on it.

In contrast, a jumble of characters like "a7$pG2*…" can't connect to real-world scenarios, so it provides no useful evidence and does nothing to reduce uncertainty. Casual small talk like "Did you eat?" "Yeah, I did" is meaningful, but it brings no new evidence and doesn't change our understanding of the world —so it barely boosts predictive ability. The takeaway? **The more a thing improves our ability to predict the world, the more valuable information it carries**—that's the core of information's worth.

Humans' love for high-density information is actually a genetic habit. And I think **this is the first principle of "interesting" v.s. "boring" judgement.**

Maybe we could all try to be “carriers of high-density information”—it’s a small shift that makes a big difference.