# Information-theoretic Neural Decoding Reproduces Several Laws of Human Behavior

**S. Thomas Christie**
NWEA
St. Paul, Minnesota
thomas@christie.one

**Paul R. Schrater**
University of Minnesota
Minneapolis, Minnesota
schrater@umn.edu

## Abstract

Features of tasks and environments are often represented in the brain by neural firing rates. Representations must be decoded to enable downstream actions, and decoding takes time. We describe a toy model with a Poisson process encoder and an ideal observer Bayesian decoder, and show the decoding of rate-coded signals reproduces classic patterns of response time and accuracy observed in humans, including the Hick-Hyman Law, the Power Law of Learning, speed-accuracy trade-offs, and response times matching lognormal distributions. The decoder is equipped with a codebook, a prior distribution over signals, and an entropy stopping threshold. We argue that historical concerns of the applicability of such information-theoretic tools to neural and behavioral data arises from a confusion about the application of discrete-time coding techniques to continuous-time signals.

## 1 Introduction

Whatever the task at hand, neurons performing task-related computations must infer features of the environment encoded in the firing rates of other neurons. Due to inherent biological constraints, this decoding process is noisy, imperfect, and takes time. In turn, decoding time enforces a lower bound on reaction time to stimuli. Despite the complex and chaotic nature of neural coding and decoding, simple changes in experimental conditions produce consistent and reliable effects on reaction times, described by 'laws' like the Hick-Hyman law [Hick, 1952, Hyman, 1953] and the Power Law of Practice [Newell and Rosenbloom, 1981]. We propose that such consistencies are a direct result of the mechanics of neural decoding.

In this paper, we use a toy model to consider information transmission from the environment, through the brain, to behavior. We focus on the encoding of discrete messages in the firing rates of simulated neurons and characterize the time it takes for an ideal observer to infer which messages are being transmitted. We show that decoding time and accuracy in this system replicates behavioral 'laws' and produces human-like response time distributions. The decoder maintains a codebook and a prior probability over possible encoded symbols, over which entropy is computed and compared to a pre-set threshold.

This information-theoretic approach affords a principled way to connect levels of analysis [Marr, 1982] by integrating energetic resource availability (implemented as limited firing rates), message encoding and decoding schemes, and task performance characteristics into a single framework. It also provides a unified, normative, and (to our knowledge) novel explanation for several behavioral phenomena.

Information-Theoretic Principles in Cognitive Systems Workshop at the 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

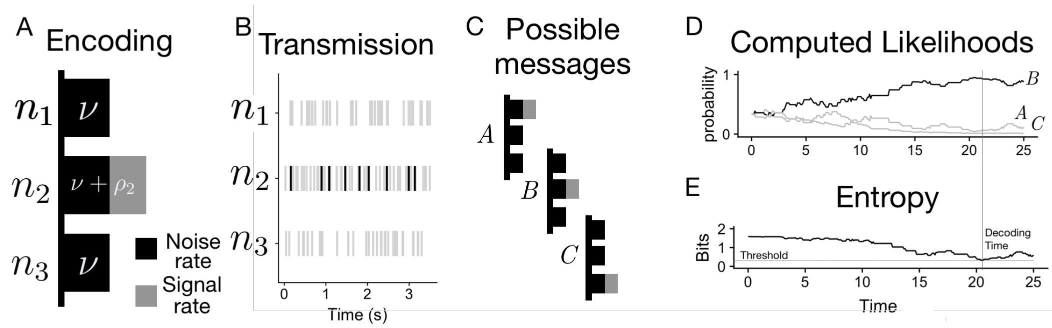


Figure 1: Schematic of a message transmission. See text for description.

## 2 The Model

We consider the scenario in which a discrete feature of the environment, such as the utterance of a word or a symbol on a screen, is represented by firing rates of a population of neurons. To model this, we let an environmental feature $i$ be represented by an $n$-length vector of firing rates $\rho_i$ over an array of $n$ neurons. We will call this vector the *signal* and refer to the population of neurons as the *encoder*. We also assume that all neurons in the encoder fire at some typical rate $\nu$ when feature $i$ is not being represented, perhaps related to other representations or to sustain a baseline level of recurrent activity. We call $\nu$ the baseline *noise*, where $\nu > 0$. The encoder neurons fire at a total rate $\tau_i = \nu + \rho_i$ when representing $i$. We model neural firing as Poisson processes with rate parameters $\tau_i$. For example, a feature $i$ might result in a population of 3 neurons producing spikes at the rates $\tau_i = [\nu + 0, \nu + 4, \nu + 0]$, in which case $\rho_i = [0, 4, 0]$, as in Figure 1A.

The decoder must decide which feature, selected from a set of possible features $M$, is currently encoded in the neural firing rates. Let us assume that the decoder has the tools typically available to a decoder in an information theoretic analysis: it has a codebook of possible messages $C_M$ specifying the signal rates $\rho_m$ used to encode each feature $m \in M$ (see Figure 1C); it knows the value of the noise rate $\nu$; it knows the likelihood function of spike counts given a Poisson process rate; it has perfect memory for counting spikes observed in the signal; and finally, it maintains a vector of prior probabilities $P_M$ of each message being encoded, where where $\sum_{i \in M} P(m = i) = 1$. Critically, $P_M$ contains the decoder's *belief* about message probabilities and may not equal actual transmission frequencies. Let us further assume that it begins the decoding process at a time $T = t_0$.

How long will this ideal decoder take to infer the feature being encoded? The decoding cannot be instantaneous: in zero time, zero spikes will have been observed. The random nature of the observations, generated as they are by Poisson processes, means that the decoder should never be 100% confident in any decoding judgment, no matter how much time has passed. Taken together, these imply that the decoder should observe spikes emitted from the encoder until it is sufficiently confident in the encoded message; that is, until an entropy over posterior message probabilities reaches a pre-specified threshold.

The decoder begins each transmission with an uncertainty over possible messages captured by $\mathcal{H}(P_M)$, where $\mathcal{H}$ is the Shannon entropy in bits. As an encoding $e$ is observed, the decoder computes the likelihood of the observation $P_t(e|M = i)$ given the spikes observed from each neuron by time $t$ (Figure 1B,D), resulting in a vector of posterior probabilities over messages $P_t(M|e) \propto P_t(e|M)P_M$, and the corresponding posterior entropy $\mathcal{H}(P_t(M|e))$. The result is an entropy time-series, as shown in Figure 1E.

The decoding time is the time from $T = t_0$ until the entropy threshold is first reached at $T = t_{thresh}$. The inferred message is the message with the greatest posterior probability at decoding time. The decoding accuracy is determined by whether the decoder's judgment about which feature being encoded is correct at $T = t_{thresh}$. In the context of task performance, $T = t_{thresh}$ is a lower bound on behavioral response time. This can be understood as a model of forced-choice reaction times; unlike popular models for forced-choice tasks, the proposed model naturally extends to more than 2 choices.

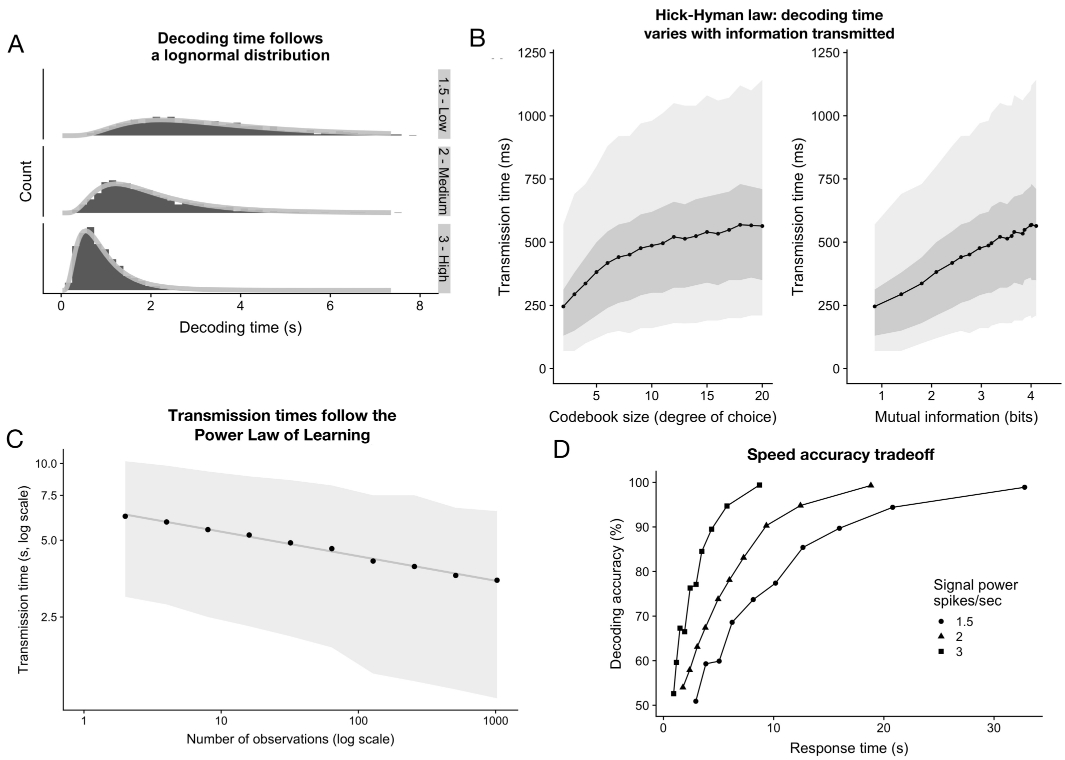


Figure 2: Decoding response time and accuracy from simulated signal transmissions. For shaded regions in B and C, points represent mean transmission times and shaded regions represent the 50% (dark gray) and 90% (light gray) percentiles of the transmission time distribution. (*A*) Distribution of decoding times fit by lognormal distributions. (*B*) Mean transmission time increases logarithmically with codebook size and linearly with information transmitted, mirroring the Hick-Hyman law. For each transmission, an entropy threshold of 0.3 bits was used, with $\rho_i$ = 16 and $\nu = 10$, though different rates produce different time scales. (*C*) Simulations reproduce the Power Law of Learning. (*D*) Varying the within-decoder entropy threshold and across-decoder signal power creates speed-accuracy trade-off curves. Simulation code is available at *https://github.com/tom-christie/transmit*.

## 3 An Instantiated Example With Simulations

We characterize the decoding accuracy and duration by performing repeated decoding simulations using the model described above. Decoding time is a function of the decoder's prior beliefs, the message being transmitted, the desired decoding confidence, and the firing rates $\nu$ and $\rho$. In each of the simulations below, the encoder uses sparse coding, where the signal rate of a single neuron $\rho_i > 0$ and $\rho_{j \neq i} = 0$ for all other neurons $j$.

**Decoding time distribution** Human response times are characteristically noisy, with substantial variation in response time even given repeated variation of the same stimuli. This variability is well-modeled by a lognormal distribution for a wide range of tasks. The variability of simulated decoding times produced by our model is shown in the histogram in Figure 2A, and overlaid with a lognormal distribution. Decoding time is a function of signal power.

**The Hick-Hyman Law** One of the earliest and most controversial application of information-theoretic concepts to human behavior was the discovery by William Hick and Ray Hyman that response times vary with the amount of information in the stimulus: with the number of lights [Hick, 1952], for example, or their relative probabilities [Hyman, 1953]. The legitimacy of an information theoretic analysis of this finding has been regularly disputed, with detractors claiming that the linear relationship between information and response time must rely on "sophisticated coding" [Laming, 2010], which cannot be employed when transmitting "single stimuli, one at a time".

Nevertheless, a sophisticated decoding scheme is not required for our model to simulate Hick and Hyman's results (see Christie [2019] for a more extensive set of simulations mirroring Hyman's experiments). Figure 2B shows signal decoding times in our simple model as a function of codebook size. Information transmission is computed as the mutual information between encoded and inferred messages over many transmissions. For each set of messages, each message was sent with equal probability and the decoder had an accurate prior belief that message probability was uniform. It turns out that for the linear relationship between information transmission and reaction time to hold, it is necessary for the decoder to have a stable estimate of the relative probabilities of messages as the task complexity increases. In human tasks, this can only be developed after extensive practice. Indeed, Hyman collected 15,000 reaction times per subject over the course of three months.

**The Power Law of Learning** Decoding time is a function of the decoder's posterior probability over possible messages. The posterior probability is a function of both the decoder's prior probability and the conditional likelihood of each possible message given the observed signal. Accordingly, the decoder's prior belief can either increase or decrease decoding time for individual messages, though *expected* decoding time always increases when the prior does not match the true message probabilities.

Suppose a decoder begins decoding with a prior belief $Q(m)$ over message probabilities, while messages are transmitted with probability $P(m)$. If the decoder keeps track of which messages are sent and updates $Q(m)$ accordingly (as e.g. parameters of a Dirichlet distribution), $Q(m)$ will gradually become a better approximation of $P(m)$ as captured by the Kullback–Leibler divergence between $Q$ and $P$. We simulated this scenario with results shown in Figure 2C. The linear relationship in the log-log plot between decoding events and decoding time is characteristic of the Power Law of Learning [Newell and Rosenbloom, 1981], a widely observed in behavioral tasks. Successive updates to $Q(m)$ can be interpreted as Bayesian updates, updating a Dirichlet prior by 1. An update of less than 1 alters the slope of the line, and may be optimal when updates are costly [Lieder et al., 2018].

The Power Law of Learning (also called the Power Law of Practice) has been criticized as describing aggregate behavior over subjects rather than the expected learning rate of any individual. We agree, and note that the results in Figure 2C are aggregated across 1,000 decoders, with each decoder updating its own prior $Q(m)$. We would expect each decoder to start with a distinct value for $Q(m)$ and maintain a distinct learning rate.

**Speed-accuracy trade-off** Varying the entropy stopping threshold produces a speed-accuracy trade-off curve qualitatively similar to those observed in behavioral experiments in humans [Heitz, 2014] and monkeys [Hanks et al., 2014]. Separate curves are often observed across subjects in experiments, and this is replicated by altering the firing rate vector of the signal.

## 4 Discussion

To date, information theory has been applied predominantly to the analysis of discrete-length signals with time-indexed samples [Shannon, 1948, Cover and Thomas, 2012]. The model proposed here instead uses a continuous-time coding scheme in which information is encoded in Poisson processes and the decoder's belief is continuously updated. This conceptual change leads directly to a coherent and normative account of pervasive behavioral phenomena. The simple model is derived from a first-principles analysis of rate coding and Bayesian updating, and relies on a task-indexed codebook, a prior distribution over messages, and an entropy stopping threshold. The codebook and prior can be understood as *contextual expectations*, a computation the human brain excels at making.

The application of information theoretic analyses to human behavior has been historically contentious, with detractors claiming that Shannon's ideas can only be applied when messages are transmitted in "very long strings of them so as to be rid of redundancies" [Luce, 2003, Laming, 2010], or that structure in stimulus sequences or features pollutes the required computations over 'pure' probabilities. We propose that such critiques confuse levels of analysis: they rightly worry that implementation decisions of a computation in one substrate (electronics) cannot be applied to another (the brain). However, when we abandon the baggage associated with discrete-time signals, we find that the core information-theoretic concepts of probability, entropy, and inference can still be profitably applied to neural and behavioral analysis.

## References


Scott Thomas Christie. *Information-theoretic bounded rationality: timing laws and cognitive costs emerge from rational bounds on information coding and transmission*. PhD thesis, University of Minnesota, 2019.

Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

Timothy Hanks, Roozbeh Kiani, and Michael N Shadlen. A neural mechanism of speed-accuracy tradeoff in macaque area lip. *Elife*, 3:e02260, 2014.

Richard P Heitz. The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in neuroscience*, 8:150, 2014.

William E Hick. On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4(1):11–26, 1952.

Ray Hyman. Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 45(3):188, 1953.

Donald Laming. Statistical information and uncertainty: A critique of applications in experimental psychology. *Entropy*, 12(4):720–771, 2010.

Falk Lieder, Thomas L Griffiths, Quentin JM Huys, and Noah D Goodman. The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review*, 25(1):322–349, 2018.

R Duncan Luce. Whatever happened to information theory in psychology? *Review of General Psychology*, 7(2):183, 2003.

David Marr. *Vision: A computational investigation into the human representation and processing of visual information*. Henry Holt and Co., Inc., New York, NY, USA, 1982. ISBN 0716715678.

Allen Newell and Paul S Rosenbloom. Mechanisms of skill acquisition and the law of practice. *Cognitive Skills and their Acquisition*, 1(1981):1–55, 1981.

Claude Elwood Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.