

Understanding the timing of cognitive processes with a variable rate neural code

S. Thomas Christie (tchristie@umn.edu)

Cognitive Science, 75 E River Rd
Minneapolis, MN, 55455, USA

Paul R. Schrater (schrater@umn.edu)

Cognitive Science, 75 E River Rd
Minneapolis, MN, 55455, USA

Abstract

Cognitive processes all require time, as they universally depend on information transmission between brain regions limited by physical and biological constraints. The time required for behavior also exhibits surprisingly lawful variation with task demands, success and failure, stimulus and response complexity, familiarity, practice and learning. Here we consider these regularities as consequences of constraints on information transmission, which we show provide rational predictions for timing effects across a surprising range of cognitive domains. We use a simple model for neural information transmission based on a variable-length rate coding model built with Poisson processes, Bayesian inference, and an entropy-based decision threshold that simultaneously replicates a broad array of well-known reaction-time effects. By providing a principled connection between a high-level normative decision framework with time-dependent neural rate codes, we integrate several disjoint ideas in cognitive science through translating plausible constraints into information theoretic terms.

Keywords: information theory; bayesian inference; rate coding; response time; learning

Introduction

Whatever the task at hand, neurons performing task-related computations must infer, in a continuous-time and streaming manner, which 'messages' are being transmitted from other brain regions (Rieke et al., 1999). This inference process is noisy, imperfect, and time-dependent, and enforces a bound on behavioral reaction time to stimuli. Despite the complex and chaotic nature of neural coding, simple changes in experimental conditions have 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 & Rosenbloom, 1981). In this paper, we consider information transmission from the environment, through the brain, to behavior as an information channel using a neural rate code. By constraining both the channel encoding and each transmitted signal to be optimally inferred under normative assumptions, we can construct a message-transmission system that replicates these regular phenomena, and produces human-like response time distributions. Our information-theoretic approach affords a principled way to connect levels of analysis (Marr, 1982) by integrating energetic resource availability, message

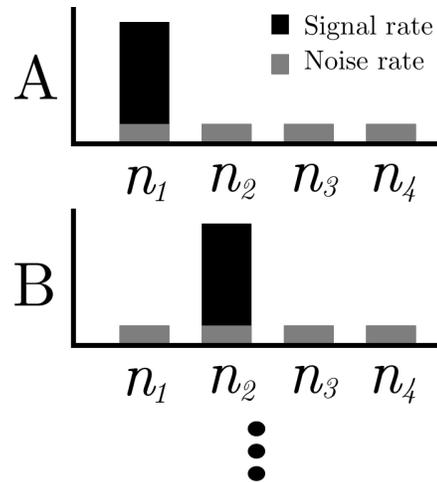


Figure 1: A codebook converts symbols A, B, etc. from a symbol alphabet into configurations of firing rates across Poisson processes n_1, n_2, \dots . In this simple model, the codebook assigns a signal rate λ_S to a single Poisson process for a given symbol. Each Poisson process also emits spikes at a noise rate λ_N . As Poisson process rates are additive, this results in a total emission rate of $\lambda_N + \lambda_S$ for the 'activated' process.

encoding and decoding schemes, and task performance characteristics into a single framework.

In what follows, we present a continuous-time variable length coding mechanism, built using entropy and inference, that adheres to the principles of information theory while providing normative predictions of signal transmission time and accuracy. We emphasize that the continuous-time nature of the code means that signals are not discretized. Because of this, we are able to transmit messages such that transmission time is linearly related to message surprisal, replicating the Hick-Hyman law. By presenting such a code, we show that appropriate information-theoretic concepts can be applied to the study of neural information transmission.

Implementation

We model information transmission by having a sender encode a message into a configuration of Poisson process firing rates, and a receiver watch the generated spikes until they are confident about the configuration of underlying rates, and thus about the content of the encoded message.

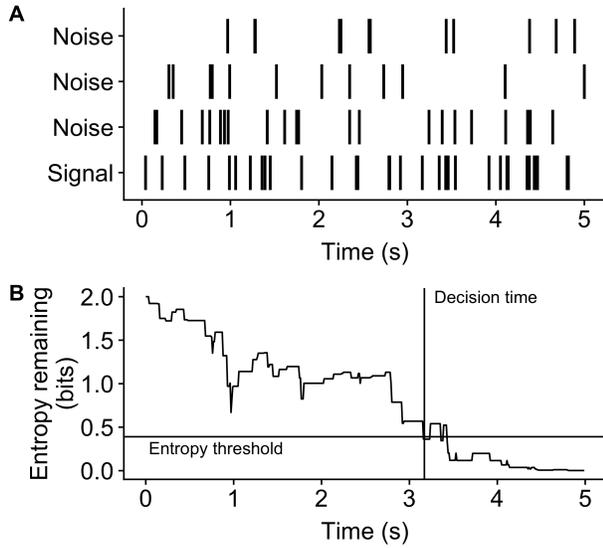


Figure 2: (A) Spikes are randomly emitted by each Poisson process as a function of time. The lower-most Poisson process is firing at a higher $\lambda_N + \lambda_S$ rate, while the others are firing at rate λ_N . (B) The receiver observes the spikes and infers which process is firing at rate $\lambda_N + \lambda_S$. The initial entropy is 2 bits, indicating a weak belief in equal probabilities for each of the 4 possible signals. The receiver's remaining entropy changes as the processes are observed and the posterior probability of each signal is calculated.

In more detail, the transmission mechanism consists of an encoder, a transmitter, a receiver, and a codebook. The transmitter is an array of Poisson processes, each continuously producing points or 'spikes' independently at a given noise rate λ_N . This can be viewed as a basic model of a neural rate code, as neural spikes trains are often modeled as Poisson processes (Rieke et al., 1999). The symbols to be communicated are taken from an alphabet of discrete symbols \mathcal{A} . The codebook describes a mapping between each symbol and a configuration of Poisson rates, and the mapping from a given symbol to rate configuration is carried out by the encoder. For the sake of expositional simplicity, we restrict the codebook to increasing the rate for a single Poisson process from the noise rate λ_N to a signal rate $\lambda_N + \lambda_S$, as shown in Figure 1. The neural analogue is that each Poisson process is 'tuned' to 'prefer' a particular symbol in a 1-hot manner, resulting in a sparse code.

The receiver observes the sequence of spikes emitting from each Poisson process and continuously attempts to infer which rate configuration is producing the spikes it observes, and thereby which symbol is being transmitted. We assume, again for simplicity and consistent with common information-theoretic analysis, that the receiver knows the values of both λ_N and λ_S . In standard binary or Gaussian channels, transmission is a discrete vector of amplitudes that takes a fixed

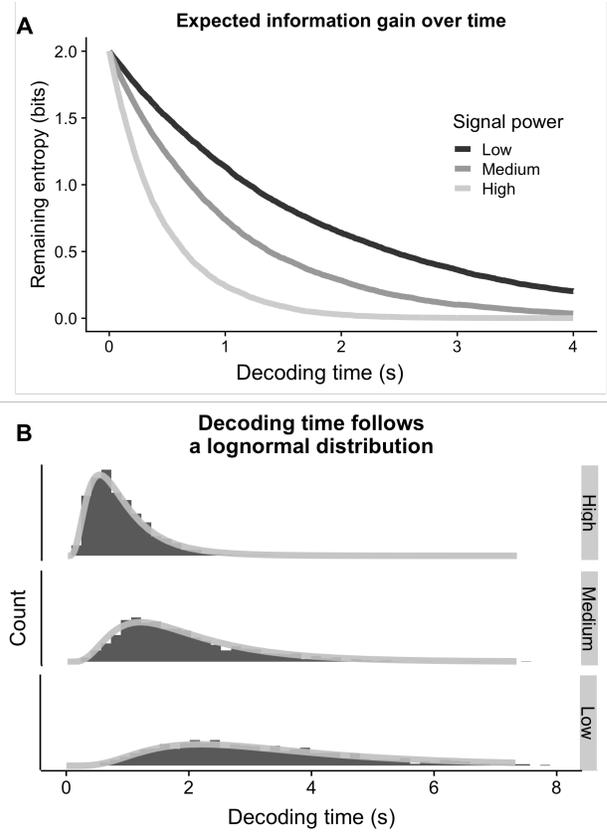


Figure 3: (A) The expected value of the receiver's entropy regarding four possible messages decreases as spikes are observed. Increasing the signal power λ_S changes the information transmission rate. (B) Response time distributions vary as a function of signal power λ_S , and in each case are well-fit by a log-normal distribution.

time to transmit. Because of this, practitioners typically speak in terms of transmitting bits-per-signal, or bits-per-second (which are a constant multiple of each other). In our case, the receiver accumulates information about each transmission gradually, over time. In effect, observing for a longer period of time adds redundancy to the signal.

As observations continue, the receiver calculates and continuously updates a posterior probability distribution over possible messages, and stops decoding when the entropy of the posterior reaches a pre-specified stopping threshold. Let transmitted symbols be treated as realizations of a random variable X . The receiver begins each transmission at time $t = 0$ with an initial uncertainty $H_Q(X)$ regarding the symbol being transmitted, reflecting its prior distribution $Q(X)$ of the possible codewords. As time passes and observations $Y_t = \{y_1, \dots, y_t\}$ are made, the receiver uses Bayesian inference to update the prior to obtain a posterior distribution $Q_t(X|Y_t)$ over messages according to Bayes rule, which yields an updated posterior entropy $H_{Q_t}(X|Y_t)$. The posterior en-

trophy decreases non-linearly with time and reflects the degree of confidence that a message has been correctly received. Transmission stops when $H_{Q_t}(X|Y_t)$ reaches a threshold. Figure 2 shows the change in posterior entropy over time for an example transmission.

Variable length transmissions

In the coding scheme introduced here, messages are *variable-length*: transmissions of messages with higher surprisal takes more time than messages with low surprisal, where surprisal is calculated using the prior probability distribution $Q(X)$ of the receiver. Recall that the surprisal $h(x)$ of a message x drawn from a distribution $P(X)$ is the logarithm of the inverse probability of the message, $h(x) = \log_2 \frac{1}{P(X=x)}$.

In ‘entropy codes,’ codeword length (and thus transmission time of each codeword) is roughly proportional to the surprisal of the encoded symbol in the absence of noise. When symbols are independently drawn according to a categorical probability distribution, this can manifest in two ways. In the first, increasing the number of possible symbols increases the surprisal of each individual symbol, and consequently the length of the code needed to encode its value. In the second, symbols drawn from a categorical distribution with unequal probabilities will have different surprisal values: more frequently transmitted messages will have lower surprisal and shorter codes than less frequent messages. We performed simulations to explore these scenarios in turn using our transmission model.

First, we varied codebook sizes and recorded transmission times using a fixed entropy threshold and a uniform source distribution. The nonzero entropy threshold occasionally results in transmission errors, as we see in human subjects. Information transmitted is thus less than the surprisal of each individual message, on average. We computed actual information transmitted by calculating the mutual information between transmitted symbols and received symbols, for each codebook size. The results are shown in Figure 4 and are a close qualitative match for the Hick-Hyman observations of human response times reported by Hick (1952) and Hyman (1953).

We next transmitted messages drawn from a non-uniform distribution $P(X)$ and measured transmission time for each message. For each transmission, we measured the information transmitted by comparing the receiver’s prior probability distribution $Q(X)$ (which equals the source distribution $P(X)$), an assumption we relax below) with their posterior distribution $Q(X|Y)$ at decision time. We measured the difference in these distributions using the Kullback–Leibler divergence between the two distributions, $D_{KL}(Q(X|Y)||Q(X))$. The change between the receiver’s prior and posterior distributions is equivalent to the decrease in the receiver’s subjective uncertainty about which message is being transmitted. From the point of view of the receiver, this is equivalent to the amount of information transmitted, in bits. Figure 5 shows a linear relationship between message surprisal and transmission time, again

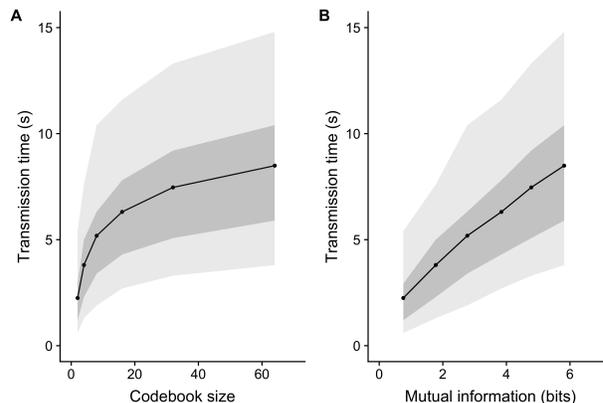


Figure 4: Mean transmission time increases logarithmically with codebook size and linearly with information transmitted, mirroring the Hick-Hyman law. Points represent mean transmission times and shaded regions represent the 50% and 90% high-density interval of the transmission time distribution. In each case, messages were transmitted according to a discrete uniform distribution $P(X)$ over messages, and the receiver maintained a uniform prior distribution $Q(X) = P(X)$ of the same dimensionality. For each transmission, an entropy threshold of 0.3 bits was used, with $\lambda_S = 4$ and $\lambda_N = 10$.

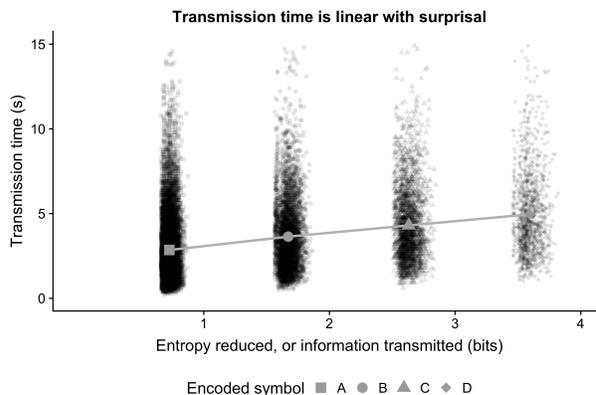


Figure 5: Mean transmission time increases linearly with actual information transmitted, echoing similar findings in humans by Hyman (1953). The quantity of information transmitted is calculated as the the KL-divergence between the prior distribution $Q(X)$ and the posterior distribution $P(X|Y)$ at decision time. Messages were drawn from a non-uniform source distribution $P(X)$. The receiver is assumed to know this source distribution and maintains a prior distribution $Q(X) = P(X)$. For each transmission, an entropy threshold of 0.3 bits was used, with $\lambda_S = 4$ and $\lambda_N = 10$.

qualitatively matching Hyman’s reported results from human subjects.

Learning to efficiently transmit

As with source-coding systems, expected message transmission times are faster when more frequently transmitted messages are transmitted in less time than less frequently transmitted messages. In our system, this is implemented by tailoring the receiver's prior distribution Q to match, as closely as possible, the source distribution P . This reveals an epistemic problem from the perspective of the receiver, which has no *a priori* knowledge of the source distribution: the prior must be learned and updated by observing message transmissions. The work of Hick and Hyman has been legitimately criticized for omitting this discussion (Laming, 2010).

Suppose we allow a receiver with an incorrect uniform prior message distribution Q_{init} to update its distribution to Q_{obs} in a Bayesian manner each time a message is received, so that the subsequent message transmission starts with the updated prior. As the receiver observes which messages are transmitted and at what relative frequency, Q_{obs} will become an ever-closer approximation to P , shrinking both $D_{KL}(P||Q_{obs})$ and the expected transmission times.

As observations accumulate, the rate at which response times decrease as Q approaches P mirrors the Power Law of Learning (Newell & Rosenbloom, 1981). The Power Law of Learning is a ubiquitous finding that task response times have a power-law relationship with the number of practice episodes, when averaged across many subjects. We constructed a categorical source distribution P with $k = 16$ categories, but with most of the probability mass in two categories. We initialized Q_{init} to have a Dirichlet prior with concentration parameters 2, representing a weak prior belief that the source distribution is uniform. We simulated N message transmissions, for $N = 2$ to $N = 1024$, taken evenly in log space. For each value of N , we averaged the results across 1,000 simulated observers, resulting in an expected posterior distribution Q_{obs} after N observations. For each Q_{obs} we then simulated more 2,000 message transmissions, with messages drawn with frequency defined by P , and calculated the transmission time for each. As illustrated in Figure 6, the relationship between observations N and transmission time is linear in log-log space, matching the Power Law of Learning.

Conclusion

We have applied the principles of information theory to a simple rate-coding model of neural information transmission. We showed that the Hick-Hyman Law, the Power Law of Practice, and the lognormal response time distributions are all produced through placing normative bounds on the inference of source distributions and the content of individual signals.

References

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4(1), 11-26. doi: 10.1080/17470215208416600

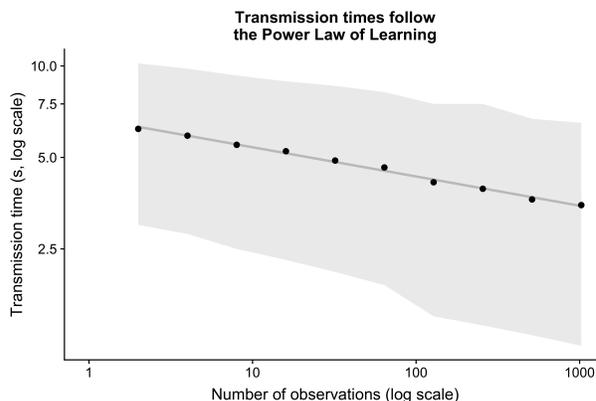


Figure 6: Simulated message transmission time decreases as a function of observations, as the prior Q approaches the source distribution P . Signals are transmitted with signal strength $\lambda_S = 4$, noise power $\lambda_N = 10$, and an entropy threshold of 0.3. Points represent mean transmission times, and the shaded region represents the 80% high-density interval of the response time distributions.

- Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of experimental psychology*, 45(3), 188.
- Laming, D. (2010). Statistical information and uncertainty: A critique of applications in experimental psychology. *Entropy*, 12(4), 72.
- Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. w. h. WH San Francisco: Freeman and Company.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition*, 1(1981), 1-55.
- Rieke, F., Warland, D., Steveninck, R. d. R. v., & Bialek, W. (1999). *Spikes: Exploring the neural code*. A Bradford Book.