Politics, Bit by Bit: A Formal Link Between Entropy and the Effective Number of Parties

Jack Bailey*

January 15, 2025

Word count: 2,461

Abstract

Information theory now informs fields as diverse as quantum physics, neuroscience, and economics. In this short note, I show that it informs political science too. In particular, I show that the effective number of parties—perhaps the most widely-used measure in all of comparative politics—is the antilogarithm of a common measure of information entropy: the Rényi entropy. This equivalence bridges the two fields, allowing advances in information theory to inform our understanding of electoral systems.

*Lecturer in Quantitative Political Science, Department of Politics, The University of Manchester, UK. If you have any comments or questions, please feel free to contact me at jack.bailey@manchester.ac.uk.

Introduction

In 1948, Claude Shannon introduced *information theory*: a mathematical framework that characterizes how signals pass over noisy communication channels. Eight decades later, Shannon's framework underpins our digital lives. It lets us communicate over vast distances, secure any data that we share through encryption, and compress, store, and retrieve photos, videos, and media at will. Though Shannon sought only to resolve problems in telecommunications research, others have since applied his insights to a diverse range of fields. They include quantum physics, linguistics, computer science, ecology, neuroscience, psychology, biology, and economics (Stone 2022; Cover and Thomas 2006).

A single concept, *entropy*, lies at the heart of information theory. It measures the average amount of uncertainty that is associated with a given random variable. When variables have *uncertain* outcomes—like a fair coin—we say that they have *high* entropy. Conversely, when variables have *certain* outcomes—like an unfair coin that always lands heads side up—we say that they have *low* entropy instead. To measure this concept, information theorists most often use the *Shannon entropy* (1948), defined as $\mathrm{H}(X) = -\sum_{i=1}^{N} p_i \log_2(p_i)$. However, several other measures of information entropy also exist, each with its own application or use case.

In this short research note, I show that information theory shares a striking connection with political science. Laakso and Taagepera's (1979) measure of the "effective" number of parties is perhaps the most widely-used metric in the field of comparative politics. It counts how many share-winning parties exist *conditional on their relative shares*. I show, however, that it also shares an *exact* equivalence with a common measure of information entropy: the Rényi entropy. Simply put, the former is the antilogarithm of the latter. This fundamental equivalence presents new opportunities for students of political science. Information entropy is both well-understood and well-known to have many useful decompositions. Thus, by bridging information theory and political science, the connection that I establish here allows us to use information-theoretic advances to better our understanding of electoral systems and the many ways that they constrain our politics.

Quantities of Interest

To establish that Laakso and Taagepera's (1979) measure of the effective number of parties is equivalent to Rényi's (1961) generalized measure of entropy, it is first necessary to motivate each measure in turn. I begin with the effective number of parties and then move on to information entropy, paying especial attention to the Rényi entropy of order $\alpha = 2$.

The Effective Number of Parties

Party systems can be more or less fragmented despite having equal numbers of parties. Consider a simple two-party system where both parties hold 50% of all seats (left-hand panel, figure 1). Now imagine two scenarios where a new third party emerges. In the first, it becomes just as popular as both existing parties. Thus, each now comes to hold ≈ 33% of all seats (middle panel, figure 1). In the second, however, the third party is much less successful. As a result, it comes to hold just 2% of all seats, reducing the seat shares of both existing parties by just a single point in each case (right-hand panel, figure 1). Though both scenarios include three seat-winning parties, their characters are fundamentally different: the first is more *fragmented* than the second.

To study such fragmentation, Laakso and Taagepera (1979) propose the following measure:

$$N_2 = \frac{1}{\sum_{i=1}^{N} p_i^2} \tag{1}$$

Where p is some distribution of party shares and N a count of the number of share-winning parties. Rather than count how many parties win some share of seats, votes, or anything else, Laakso and Taagepera's measure of the "effective" number of parties instead counts the number of share-winning parties *conditional on their relative shares*. This, they explain, is akin to measuring "the number of hypothetical *equal*-size parties that would have the same total *effect* on fractionalization of the system as have the actual parties of *unequal* size" (1979, 4, emphasis in original). As a result, where one party wins *all* shares and any $p_i = 1$, $N_2 = 1$. Likewise, where all parties win an *equal* proportion of shares and all $p_i = \frac{1}{N}$, $N_2 = N$. Thus, in general, the effective number of parties may take any real value such that $1 \leq N_2 \leq N$.

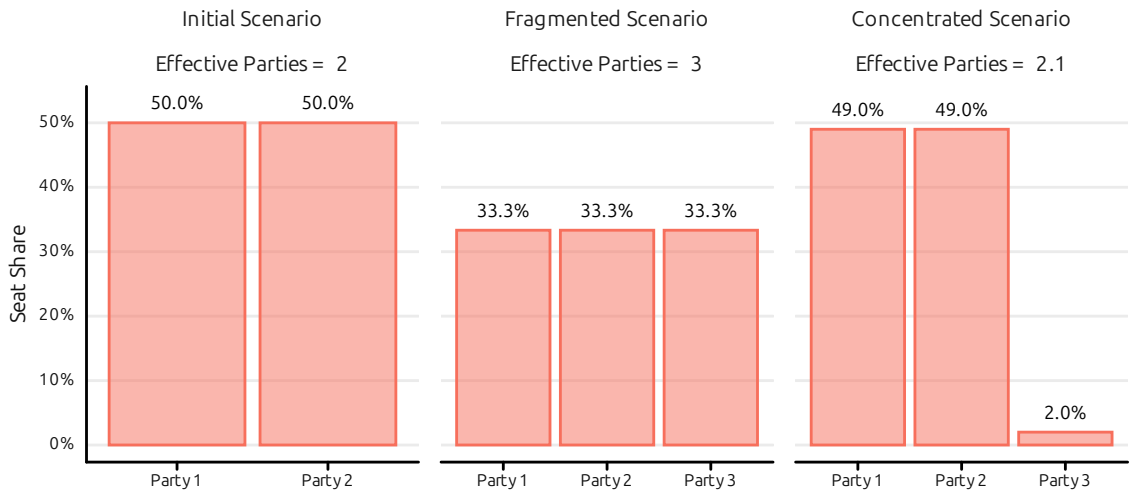

Figure 1: Seat shares from three hypothetical party systems. Note that the degree of fragmentation, measured by the effective number of parties, highlights differences in party system dynamics despite the second and third panels having identical party counts.

Comparing these scenarios shows why Laakso and Taagepera's measure is so useful. The fragmented scenario (middle panel, figure 1) and the concentrated scenario (right panel, figure 1) both feature the same *actual* number of seat-winning parties. Yet, since the parties' shares differ in each scenario, so too do the *effective* number of parties. In the fragmented scenario, all three parties have an equal share of all seats, giving $N_2 = 3$. The concentrated scenario, however, much more closer resembles the initial scenario (left panel, figure 1), giving $N_2 = 2.1$. Thus, the effective number of parties lets us compare party systems based on their degree of fragmentation.

Information Entropy

Information reduces uncertainty, which then reduces surprise. So any useful information that you learn about some system should make its outcomes both easier for you to predict and less surprising when they come to pass. What's more, this reduction in your uncertainty and in your surprise should also scale with how much you learn. When you learn only a little, any reduction will be only small. But, by contrast, if you learn a lot, this reduction will be large instead. This implies, by extension, that if you were to learn *everything* that there is to know about some particular system, you would eliminate any uncertainty that you once had. It would be as though you made the system deterministic, removing any ability that it had to surprise.

We can quantify the uncertainty inherent in some system by measuring its *entropy*. In general, the information entropy of some probability distribution, p, of length N, can range from a low of 0 where there is no uncertainty to a high of $\log_2(N)$ where uncertainty reaches its maximum.[1] As a result, the entropy of any given random variable is also contingent on its length. Most often, information theorists measure this uncertainty using the Shannon entropy, first introduced by Shannon (1948) and defined as follows:

$$\mathrm{H}(X) = -\sum_{i=1}^{N} p_i \log_2(p_i) \tag{2}$$

Consider, for example, a pair of coins: one fair and one loaded to always land heads side up. Both coins have only two outcomes: they can be either heads or tails. Yet the outcomes that they produce are different: whereas the outcome of the fair coin is uncertain (heads, 50%; tails; 50%), the outcome of the loaded coin is not (heads, 100%; tails; 0%). The coins' entropy also reflects this difference in uncertainty. The fair coin has an entropy of $-0.5 \log_2(0.5) - 0.5 \log_2(0.5) = 1$, the *maximum* possible, since $\log_2(2) = 1$. The loaded coin, instead, has an entropy of $-1 \log_2(1) - 0 \log_2(0) = 0$, the *minimum* possible, since entropy must be greater than or equal to zero.

We can extend this example further by considering *all* possible coins with *all* possible probabilities of landing heads or tails side up. Figure 2 shows how the entropy changes as the probability of receiving a heads moves from 0% to 100%. Since entropy measures uncertainty, the entropy of the coin is zero if either heads or tails is certain. The coin's entropy then increases as its outcome's uncertainty grows. It then reaches its peak—the *maximum entropy*—where both heads and tails have an equal chance of occurring. Were we to increase the number of outcomes by, for example, considering dice with an arbitrary number of sides, this pattern would continue to hold. Thus, in general, for some discrete probability distribution, p, entropy is *maximized* where all $p_i = \frac{1}{N}$ and each outcome is equally probable. Likewise, entropy is *minimized* where any $p_i = 1$ and any one outcome is totally certain.

[1] Note that information theorists tend to use a log base of 2, which also defines a unit of information known as a *bit*. Due to the use of base 2, one bit corresponds to the amount of information needed to reduce some probability space by half (for example, the amount of information necessary to predict the outcome of a fair coin toss). For more information on bits as a unit of measurement, see Stone (2022).

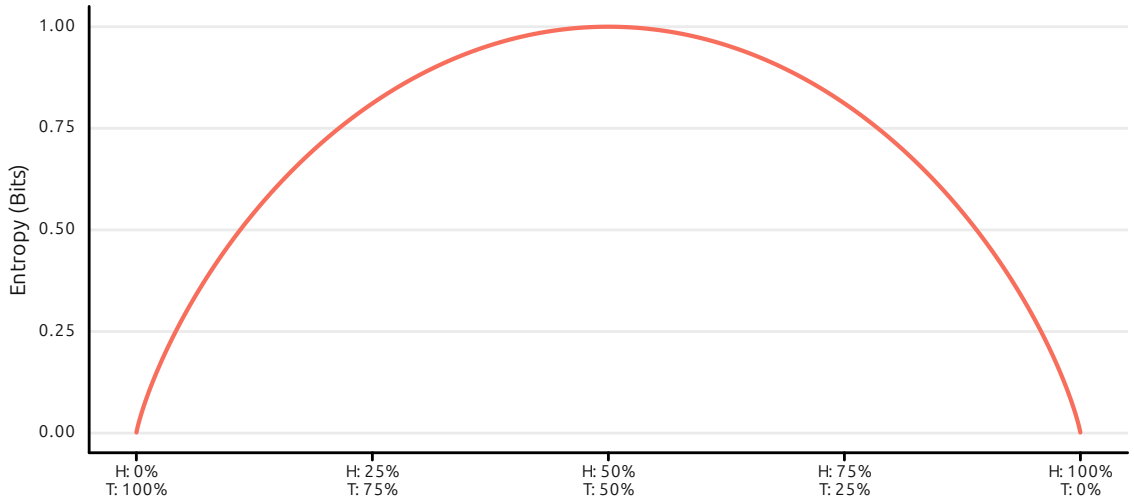

Figure 2: The Shannon entropy of a loaded coin, measured in bits, where the probability of recieving either a heads or a tails varies from 0% to 100%.

The Shannon entropy weights all probabilities equally. However, there are some cases where we might want to place greater weight on more rare or more frequent events (for instance, to stress uncommon but potentially catastrophic outcomes or to design efficient compression algorithms that focus on the outcomes that occur most often). As such, Rényi (1961) proposed a generalized entropy measure that allows one to adjust the weight one places on more or less frequent events. This measure—now known as the Rényi entropy—is defined as follows:

$$\mathrm{H}_{\alpha}(X) = \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^{N} p_i^{\alpha} \right) \tag{3}$$

Like the Shannon entropy, the Rényi entropy takes as its input some probability distribution, p, of length N. Likewise, it also produces equivalent measures of entropy that range from a low of 0 where there is no uncertainty and a high of $\log_2(N)$ where uncertainty reaches its peak. Unlike the Shannon entropy, however, the Rényi entropy includes an additional parameter, $\alpha \geq 0$, that determines its "order": how sensitive it is to more or less common events. This parameter may take any real non-negative value, though they span three specific domains. First, where $0 \leq \alpha < 1$, the measure emphasizes less frequent events. Second, in the limit where α tends to one, the measure converges on the Shannon entropy and weights all events equally. Third, where $\alpha > 1$, the measure emphasizes more frequent events instead.

Proof of Equivalence

It is now clear that both the effective number of parties and the Rényi entropy quantify how uniform (and, thus, how uncertain) some probability distribution might be. This suggests the prospect of a much closer connection. To show that this is the case, first recall Laakso and Taagepera's (1979) definition of the effective number of parties:

$$N_2 = \frac{1}{\sum_{i=1}^{N} p_i^2} \tag{4}$$

Note that, if we take the logarithm of both sides of this equation and then apply the reciprocal rule, $\log_b\left(\frac{1}{X}\right) = -\log_b(X)$, to its right-hand side, we get:

$$\log_2(N_2) = -\log_2\left(\sum_{i=1}^{N} p_i^2\right) \tag{5}$$

Now consider the Rényi entropy of order $\alpha = 2$, which simplifies such that:

$$\mathrm{H}_2(X) = \frac{1}{1-2}\log_2\left(\sum_{i=1}^{N} p_i^2\right) = -\log_2\left(\sum_{i=1}^{N} p_i^2\right) \tag{6}$$

As we can see, the right-hand sides of equations (5) and equation (6) both have *exactly* the same functional form. This implies, by substitution, that the following must be true:

$$\log_2(N_2) = \mathrm{H}_2(X) \tag{7}$$

Which, then, in turn, implies that:

$$N_2 = \mathrm{antilog}_2\left(\mathrm{H}_2(X)\right) \tag{8}$$

In other words, Laakso and Taagepera's (1979) measure of the effective number of parties is information-theoretic in nature. More specifically, it is the antilogarithm of the Rényi entropy of order $\alpha = 2$ and measures how many equally-weighted components some probability distribution would need to generate the same entropy that we have observed.

Generalized Proof of Equivalence

This equivalence shows that the effective number of parties is not a *quantity* but a *function*: like the Rényi entropy, it takes some probability distribution and maps it onto a single summary measure. This raises an interesting possibility: if we adjust the order of the Rényi entropy, then take its logarithm, we might define a *family* of effective numbers of parties, such that:[2]

$$\mathrm{N}_{\alpha}(X) = \mathrm{antilog}_2\big(\mathrm{H}_{\alpha}(X)\big) \tag{9}$$

Though they do not recognize the information-theoretic nature of their measure, Laakso and Taagepera (1979) have much the same intuition. In their original study, they propose a "generalized expression for the effective number of components" (6), denoted N_{α}, as follows:

$$N_{\alpha} = \left(\sum_{i=1}^{N} p_i^{\alpha}\right)^{1/(1-\alpha)} \tag{10}$$

However, we can show that this generalized measure is *also* equivalent to the antilogarithm of the generalized measure of entropy that Rényi (1961) defined two decades earlier, suggesting a more fundamental connection between entropy and the effective number of parties. Indeed, if we take the logarithm of both sides of equation (10) and then apply the reciprocal rule, we find that $\log_2(N_{\alpha})$ is equal to the right-hand side of equation (3):

$$\log_2(N_{\alpha}) = \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{N} p_i^{\alpha}\right) \tag{11}$$

So, again, by substitution, we find that:

$$\mathrm{N}_{\alpha}(X) = N_{\alpha} = \mathrm{antilog}_2\big(\mathrm{H}_{\alpha}(X)\big) \tag{12}$$

[2]It is also worth noting that treating the effective number of parties as a function and denoting it using functional notation produces less unwieldy expressions than the subscript notation that Laakso and Taagepera (1979) use. For instance, they denote the effective number of seat-winning and vote-winning parties as N_{S2} and N_{V2}, respectively. The functional equivalent, however, would be $\mathrm{N}_2(s)$ and $\mathrm{N}_2(v)$, which leave room for subscripts on the distribution of seats or votes and avoid constructions like $N_{S\infty}$ in favor of $\mathrm{N}_{\infty}(s)$.

Conclusion

Information theory has a broad influence on a wide range of fields. In this short research note, I show that its influence has also reached political science. In particular, I demonstrate that a fundamental equivalence exists between Laakso and Taagepera's (1979) measure of the effective number of parties and the Rényi entropy of order $\alpha = 2$, a common measure of information entropy. Further, I also show that this equivalence applies to all effective numbers of parties, such that $\mathrm{N}_{\alpha}(X) = \mathrm{antilog}_2\big(\mathrm{H}_{\alpha}(X)\big)$.

This equivalence allows political scientists to adopt powerful tools from information theory, such as entropy decompositions, to study party-system fragmentation and electoral systems. We might also leverage known relationships between information-theoretic metrics to develop more precise logical models of electoral systems (Taagepera 2008). With any luck, these models might illuminate the mechanisms underlying electoral system design, offering insights that strengthen our democratic institutions.

References

Cover, Thomas M, and Joy A Thomas. 2006. *Elements of Information Theory*. 2nd ed. Hoboken, NJ: Wiley.

Laakso, Markku, and Rein Taagepera. 1979. "'Effective' Number of Parties: A Measure with Application to West Europe." *Comparative Political Studies* 12 (1): 3–27. https://doi.org/10.1177/001041407901200101.

Rényi, Alfred. 1961. "On Measures of Entropy and Information." In *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960*, 1:547–61. Berkeley, CA.

Shannon, C E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27 (3): 379–423.

Stone, James V. 2022. *Information Theory: A Tutorial Introduction*. 2nd ed. Sheffield, UK: Sebtel Press.

Taagepera, Rein. 2008. *Making Social Sciences More Scientific: The Need for Predictive Models*. Oxford, UK: Oxford University Press.