A Formal Link Between Information Entropy and the Effective Number of Parties

Jack Bailey*

January 09, 2025

Word count: 2,466

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, thereby allowing us to use advances in information theory to improve our understanding of electoral systems and their consequences.

*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 proposed *information theory*: a mathematical framework that characterizes how signals pass over noisy channels. Eight decades later, this framework underpins most aspects of 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 with ease. In developing his framework, Shannon sought only to resolve problems in telecommunications research. Yet others have since used his insights to solve problems in a diverse range of fields, including quantum physics, linguistics, computer science, ecology, neuroscience, psychology, and economics (Stone 2022; Cover and Thomas 2006).

Central to information theory is the concept of *entropy*: the average amount of uncertainty associated with a given random variable. Variables with *uncertain* outcomes—such as a fair die or coin—are said to have *high* entropy. Conversely, variables with more *certain* outcomes—such as an unfair coin that always lands heads side up or a loaded die that always produces a six—are said to have *low* entropy instead. To measure this concept, information theorists tend to 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 note, I show that, much like other fields, information theory and political science share a striking connection. In particular, I show that there is an *exact* equivalence between Laakso and Taagepera's (1979) measure of the "effective" number of parties—perhaps the most widely-used metric in the field of comparative politics, if not in all of political science—and a common measure of information entropy: the Rényi entropy. Simply put, the former is the antilogarithm of the latter. Though this equivalence is important in its own right, it also presents new opportunities for students of political science. Information entropy is both well-understood and well-known to have many useful decompositions and extensions. Thus, the connection that I establish should allow us to use information-theoretic advances to better our understanding of electoral systems and the consequences that they have for 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 entropy measure, 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 exactly the same number of seat-winning parties. Consider a simple two-party system where each party holds 50% of all seats (left-hand panel, figure 1). Now imagine that a third party emerges. In one scenario, it becomes as popular as both existing parties. Thus, each comes to hold ≈ 33% of all seats (middle panel, figure 1). In another scenario, however, it establishes much less support. 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 each (right-hand panel, figure 1). Though both scenarios feature three seat-winning parties, their characters are fundamentally different: one is more fragmented than the other.

In order to study these differences in fragmentation, Laakso and Taagepera (1979) propose a measure of the "effective" number of parties, which they denote N_2 and define as follows:

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

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 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). Thus, 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 measure can take any real value such that $1 \leq N_2 \leq N$.

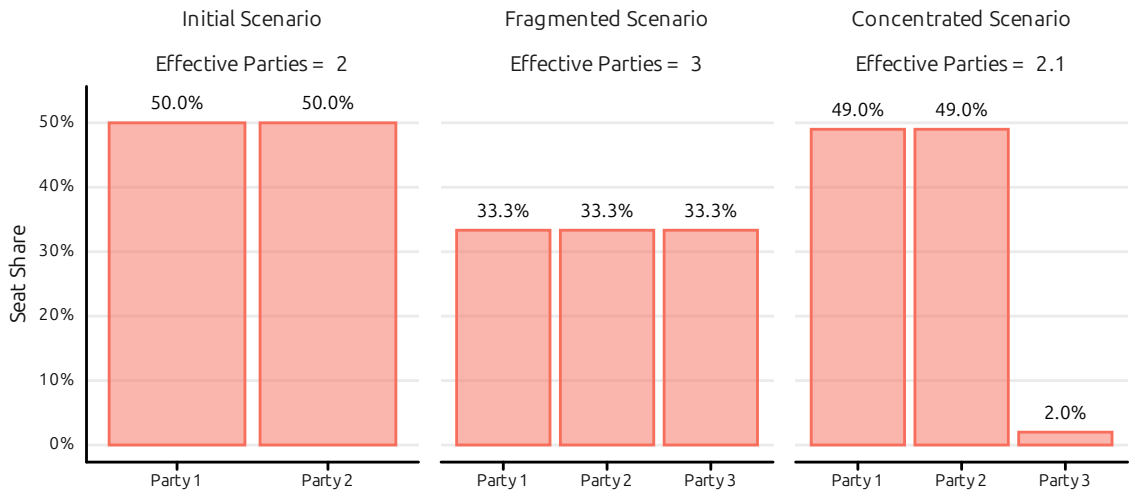

Figure 1: Seat shares from three hypothetical party systems. Note that the degree of fragmentation, given by the effective number of parties, differs for scenarios 1 and 2, despite both having the same actual number of parties.

If we compare all three scenarios shown in figure 1, we can see why Laakso and Taagepera's measure is useful. Note that the fragmented scenario (middle panel) and the concentrated scenario (right panel) both feature the same *actual* number of seat-winning parties. Yet each features a different *effective* number of parties, reflecting each scenario's differing level of fragmentation. In the case of the fragmented scenario, $N_2 = 3$, since all three parties have an equal share of all seats. But, in the case of the concentrated scenario, $N_2 = 2.1$, not much different to the initial scenario (left panel), since two parties still hold almost all of the seats.

Information Entropy

Information reduces uncertainty and uncertainty reduces surprise. As such, anything that you learn about a system should also make the outcomes it produces less surprising. Likewise, your reduction in surprise should also scale according to how much you learn. If you were to learn just a little, your reduction in surprise should be small. But, if you were to learn a lot, your reduction in surprise might be large. Thus, by extension, if you were to learn *everything* that there is to know about some system, any outcome that it produced should *never* surprise you at all. By learning so much and by reducing your uncertainty about the system to zero, it is as though you have rendered it deterministic, thereby removing its ability to surprise.

Information entropy quantifies how uncertain, and, thus, how surprising, some random variable might be, on average. In general, the entropy of some probability distribution, p, of length N, can range from a low of 0 (no uncertainty) to a high of $\log_2(N)$ (maximum uncertainty). As a result, the entropy of any given variable is also contingent on its length. Most often, information theorists measure information entropy using the Shannon entropy, first introduced by Shannon (1948) in his path breaking paper and defined as follows:[1]

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

For example, consider 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 their inherent uncertainty is 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 or equal to zero.

We can extend this example by considering *all* possible coins with *all* possible probabilities of landing heads or tails side up. Figure 2 shows how the coin's entropy changes as the probability of receiving a heads (or not receiving a tails) 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, in general, entropy is *minimised* where any $p_i = 1$ and some outcome is totally certain.

[1]Information theorists tend to use log bases of 2, since it yields many desirable properties when it comes to measuring units of information and defining the quantity known as a *bit*. One bit corresponds to the amount of information needed to reduce some probability space by half (e.g. to predict the outcome of a fair coin toss). For more information on bits as a unit of measurement, see Stone (2022).

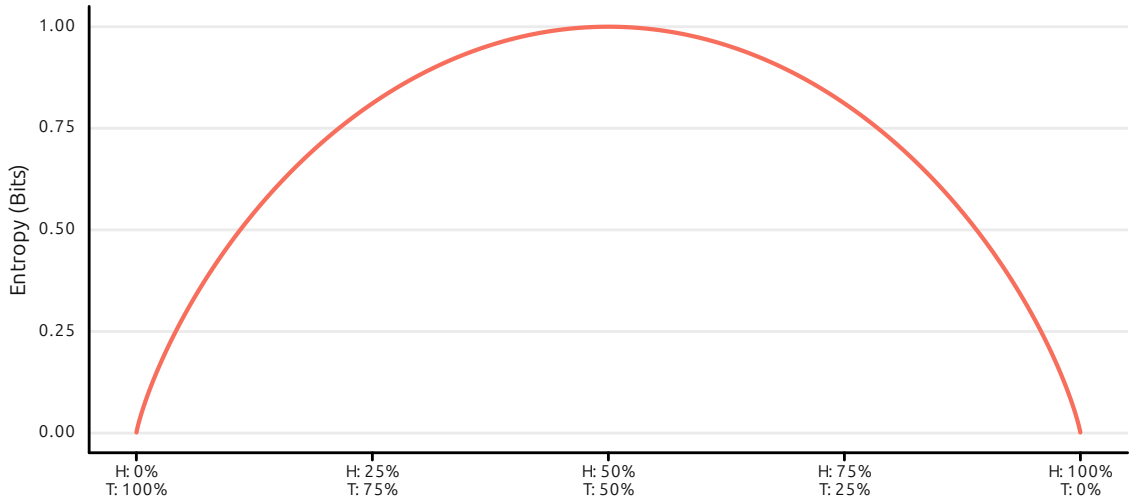

Figure 2: The Shannon entropy of a loaded coin, measured in bits, where the probability of recieving either a heads or a tails varies between 0% to 100%.

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 by focusing on common events). As such, Rényi (1961) introduced a generalized measure of entropy that allows one to adjust how much 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 is defined according to some probability distribution, p, of length N. Likewise, it also produces equivalent measures of entropy that also fall somewhere between a low of 0 (no uncertainty) and a high of $\log_2(N)$ (maximum uncertainty). However, unlike the Shannon entropy, the Rényi entropy includes an additional parameter, $\alpha \geq 0$, that determines its "order": how sensitive the measure is to more or less common events. This parameter may take any real non-negative value across 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

Having motivated both the effective number of parties and the Rényi entropy, we can now establish that these two measures are, in fact, equivalent. To do so, first recall Laakso and Taagepera's (1979) definition of the effective number of parties:

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

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 the right-hand side of the equation, we find that:

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

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 (6) are *exactly* the same. Thus, by substitution, we find that:

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

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 $\alpha = 2$. Similarly, for the same reason, the Rényi entropy of order $\alpha = 2$ is the logarithm of the effective number of parties.

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 distribution of party shares 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* 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 recognise the information-theoretic nature of their measure, Laakso and Taagepera (1979) have much the same intuition. In their original study, they 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 the generalized expression for the effective number of parties shown in 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 ugly constructions like $N_{S\infty}$ in favour of $N_{\infty}(s)$.

Conclusion

Information theory informs many fields. In this short note, I show that it also informs political science. More specifically, I show that Laakso and Taagepera's (1979) measure of the effective number of parties is the antilogarithm of the Rényi entropy, a common measure of information entropy. Further, I show that this is not just the case for the common measure of the effective number of parties: it applies to the general case too.

This equivalence, in effect, bridges the two fields, allowing us to use advances in information theory to inform how we understand politics. For example, we might adapt information-theoretic tools that decompose or extend entropy measures to study political phenomena like party-system fragmentation. Similarly, we might use known relationships between information theoretic measures to build better logical models (Taagepera 2008) of electoral systems. This might then reveal more of the mechanistic relationships that exist between elements of electoral system design, letting us build more robust democracies.

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.