Environment and Planning B: Planning and Design 2012, volume 39, pages 174 – 182

doi:10.1068/b37063

The entropy of LEGO®

Andrew Crompton
School of Architecture, University of Liverpool, Leverhulme Building, Abercromby Square, Liverpool L69 3BX, England; e-mail: crompton@liv.ac.uk
Received 31 May 2010; in revised form 27 April 2011

Abstract. Can the information content of a building be given meaning? Provided it can be modeled as a set of repeating parts the entropy of those parts can be measured. One way to do this is to make a LEGO® model then base the calculation on how many pieces of each type have been used. It is claimed that Lego is a language-like method of representing the built environment and that Lego pieces are analogous to words because their rank – frequency distribution follows a Zipf – Mandelbrot distribution similar to words in English. The zero-order entropy of Lego was measured to be about 8.5 bits, compared with 9.4 bits per word of English. Lego models of famous buildings were found to have entropy of between 6 and 10 bits per piece, giving a quantitative measure of how unusual their pieces were.

Keywords: Lego, language, information, Zipf – Mandelbrot

1 Introduction

The information content of an object can be defined as the number of binary decisions needed to identify it out of a set of known alternatives. This can be applied to a building provided it is first modeled as a set of repeating parts so that the calculation amounts to counting how many questions are needed to name each part in turn. The entropy of the building is the average information content of these parts, measured in bits per part. When this is multiplied by the number of parts it gives the information content of the whole building. Unfortunately entropy can vary wildly depending on both how the model is made and how the questions are asked. Only if a standard method of measurement is used can different buildings be compared meaningfully.

The entropy of an object tells us how unlikely it is, supposing that we know how it could have been different. In thermodynamics an ordered object like a crystal has less entropy than the same material melted. Similarly, in information theory, ordered sequences of symbols carry less information than disordered sequences. Since in ordinary life we often connect information with order and think of randomness as communicating nothing, this can seem odd. An example of our paradoxical understanding of entropy occurs in the writings of artist Robert Smithson. He associated minimalist modern art, like the work of Sol LeWitt, with heat-death, which is a state of maximum entropy (Smithson, 1979 [1966]). Work like this would have low entropy as measured by information theory, but high entropy as described by Smithson. I take Smithson's description to be poetic. Normally order and structure reduce entropy, as seen in the following examples.

2 Zero and higher order entropies

Let us compare different ways of measuring the entropy of a building. Suppose it is modeled as a set of parts which come in n types numbered $i = 1, 2, ..., n$. The zero-order entropy is the average number of yes – no decisions needed to identify randomly chosen parts of this model. Since they come in n types this ought to be no harder than identifying a random number less than or equal to n; the number of questions this requires is no more than the number of digits in the binary number n, which is roughly $\log_2 n$.

Entropy is greatest when the parts are equally probable; if some of them occur more frequently than others they can be identified with fewer guesses than this. In all cases the zero-order entropy, H, measured in bits per part, can be calculated using equation (1), in which p_i is the probability that a randomly chosen part is of type i, (MacKay, 2003, page 66–73; Shannon, 1950).

$$H = \sum_{i=1,\ldots,n} p_i \log_2 \left(\frac{1}{p_i} \right) . \qquad (1)$$

A simple way to model a building as a set of repeating parts is to use a grid whose cells are, or are not, filled. For example, the Empire State Building in figure 1 can be read as an $8 \times 9 \times 50$ block of cells, of which 563 are occupied. The probability that a unit is empty is 0.84 and that it is occupied is 0.16. From equation (1), its zero-order entropy is found to be 0.63 bits per cell.

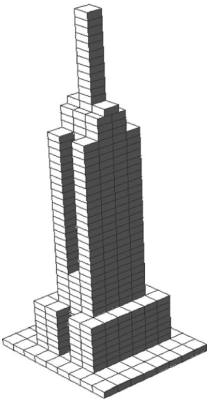

Figure 1. A grid model of the Empire State Building.

This result would be unchanged whichever 563 cells were filled. Taking into account the fact that what we know about the building helps us to predict whether a cell is filled or not leads to higher order measures of entropy. Suppose we go through the model one line at a time, one layer at a time, guessing whether the next cell will be occupied or not. Once we realise that the model has two axes of symmetry and that each layer is usually the same as the one below it is surprisingly easy to guess what is coming next. I estimate that you could get to the top with as few as thirty-five to forty-five mistakes, which is therefore the information content in bits. Measured this way the entropy of the model is about 0.01 bits per cell, far lower than the zero-order estimate because it takes structure into account.

This guessing game could be automated by digitising the model cell by cell, writing 0 for empty and 1 for occupied. Since portions of the resulting sequence repeat themselves, just as the building does, it would then be compressible by an algorithm such as Lempel–Ziv. This shortens binary strings by compiling a dictionary of substrings as it goes along, so that it can replace those that have occurred before with a pointer to an entry in the dictionary. Asympototically Lempel–Ziv approaches the entropy of the source (MacKay, 2003, page 120). If information is seen as the amount of the data after data compression, then the length of this compressed number is a measure of the information content of the model.

These higher order entropies are calculated after using algorithms, space grammars, symmetries, repetitions, and what have you, to describe a model in as few bits as possible after decomposing it into parts that may, or may not, correspond to how we read it visually. Results depend upon which method is used. At first sight, the file size of the most compact description of a building appears to be a way of defining information content in a way everyone could agree on. Unfortunately, it is impossible to prove that a particular method is minimal, so an absolute measurement of algorithmic complexity is an inaccessible ideal (Chaitin, 1975).

3 An architectural approach to measuring information

A method of parsing a building that proceeds like a draughtsperson, starting with a grid then filling in the detail, can lead to a measure of information that corresponds to how we perceive a building architecturally. This approach was first used by Kiemle in 1967. He sketched buildings at increasing levels of complexity, at each stage listing and counting the different parts that he drew, such as blocks, windows, and pilasters. The information content of each level was then found using equation (1). An example of his stage-by-stage process is shown in figure 2; in its most elaborate form (top left) the building had 260 elements with a zero-order entropy of about 3.6 bits (Wolpert, 1973). This is a low figure because the probability that a part is of such and such a type is calculated only with respect to a single building elevation. Kiemle's method is similar to a computer-aided design model in which repeating elements are defined as objects so that when they occur the file refers to the original in a library. An efficient draughtsperson will produce a naturally compressed file in which a model is resolved into a set of objects occurring with certain frequencies whose entropy can be measured just as Kiemle did.

Kiemle's method has been used to justify the preservation of older buildings, even those with no special historical or artistic merit, on the grounds that buildings with more information are more interesting (von Buttlar et al, 1972). Whether high information is correlated with being valuable is unproven, even questionable: after all, a chaotic building could have high information yet still be dull. However, experiments have shown that zero-order entropy is correlated with perceived visual diversity, establishing a useful link between entropy and something we can grasp (Stamps, 2002; 2003).

Both Kiemle and Stamps calculated the probabilities of parts being of such and such a type based on how often they occurred within a single elevation or a few houses. But this may not reflect how unusual we feel those parts to be. The probability of a part depends upon what we consider its alternatives to be: are those to be found only

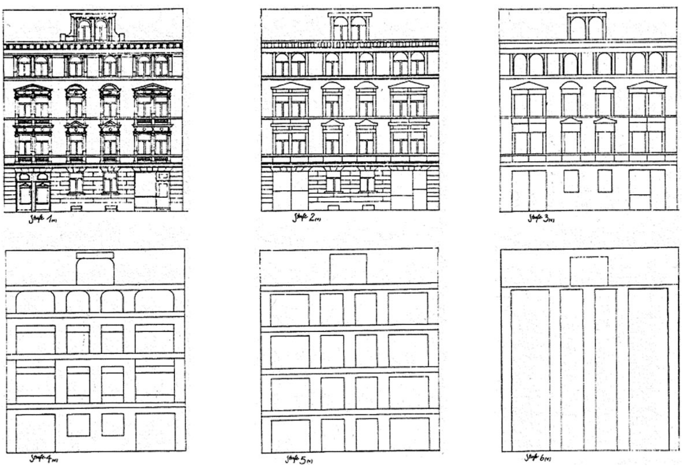

Figure 2. Six levels of detail in a Munich building (von Buttlar et al, 1972).

within the building itself, or in its neighbourhood, or an entire city? When a building is seen in the company of others, its entropy will increase because its features become less probable.

Can we find a standard method of representing buildings that avoids this vagueness? What is needed is a method for which the zero-order measure is based on parts that have some architectural character. These parts should be well defined and yet be able to represent almost anything. Furthermore, we ought to be able to calculate their probabilities with respect to a universal set of parts not just the ones that appear in the particular model in which we are interested. All this can be achieved by using Lego.

4 LEGO® is like a language

LEGO® began in 1949 as a system that could grow and any of its pieces, no matter how old, can still be joined to any other using the original snap connection. Although early kits were designed for making models of the built environment, the system has since developed to be able to construct mechanisms, fantasy environments, and at LEGOLAND®, parts of cities. A Lego model is always recognisable as being Lego just as a sentence is always heard as belonging to a particular language. Indeed, if we treat pieces as words and models as utterances then Lego is like an artificial language: a finite system applied with different degrees of creativity to an infinite variety of situations. In linguistics this potential to represent almost anything is called *productivity*, and is one of six core features of language in general (Yule, 1996). The other core features also appear in Lego to some degree. They are, *displacement*: Lego models are communicative and not just informative because they can refer to things not immediately present; simpler forms of communication lack this property if like animal signs they have fixed reference. *Arbitrariness*: words, onomatopoeia excepted, are said to be arbitrary because there is no connection between how they sound and what they represent. The same is true of many, though not all, Lego pieces. The 'minifig' human figures, for example, resemble and represent one thing only; but for the most part what pieces mean depends upon how they are situated in a model. *Cultural transmission*: Lego, like other languages, is passed on from one generation to the next by cultural transmission. To help this process Lego has an equivalent to baby-talk: DUPLO®, double-sized bricks for infants, which can nonetheless join to ordinary pieces. *Discreetness*: like words Lego pieces are all different and distinct; they do not blend one into the other. *Double articulation*: sentences are made of words which are in turn made of phonemes; similarly Lego models are made of pieces which are in turn made of lesser repeating parts, such as studs, holes, and a size module. Clearly this is not as well developed in Lego as in language, but parts of broken Lego bricks can still be recognised as such showing that the quality of being Lego goes deeper than the brick.

More interestingly Lego resembles a language statistically: the rank-frequency distribution of Lego bricks follows a proper law similar to that for words in English known as Zipf's law. This was established with data on Lego obtained from the Peeron website, an independent database of Lego sets (http://peeron.com). In 2008 I extracted the inventories of all the 6200 sets listed on this site, which at that time listed 12 588 different types of piece. This list does not distinguish bricks which differed only in colour. Mixed together these 6200 Lego sets contained 892 952 pieces. I made a list of the number of times that each type of brick occurred in this ensemble. Lego pieces were ranked $r = 1, 2, ..., n$, 12 588 according to how frequently they occurred, with the most common piece ranked number one. The number-one piece is the type '3004 Brick 2×1' which accounts for 4.6% of all pieces ever issued. Roughly a third (4647 out of 12 588) of pieces appear in only one Lego set: for example, the '6272c01 Galidor Torso' which is unique to *Set 8317-1*. The number of types is similar to the number of words in a working vocabulary of English. For comparison, Pidgin English makes do with about 750 words, and MECCANO® (ERECTOR® in the US) had, in 1964, only about 250 pieces. Lego, like other languages, grows from the high-rank end, which is to say that new pieces are introduced as unique; only later do some of them become more common.

Let $n(r)$ be the number of instances of each type in the ensemble of Lego sets. The result of plotting $\log r$ against $\log n$ is shown in figure 3. A straight line would indicate that the distribution is scaling: in fact, the points seem to lie on a slightly curved line. A better model is the Zipf–Mandelbrot distribution in equation (2). Let $\mathrm{P}(r)$ be the probability of a randomly chosen brick being of rank r, and let k, v, and a be constants, with a being close to unity:

$$\mathrm{P}(r) = k(r+v)^{-a} . \quad (2)$$

I find by minimising the sum of squares of differences from the actual distribution that a good fit is achieved with $k = 0.64$, $v = 6$, and $a = 1.32$. Figure 3 compares the actual distribution with equation (2) using these values. It is similar to the fit of the equation to English text: for example, using the words in Jane Austen's *Emma*: $k = 0.56$, $v = 8$, $a = 1.26$ (MacKay, 2003, page 262). Power laws like this are common

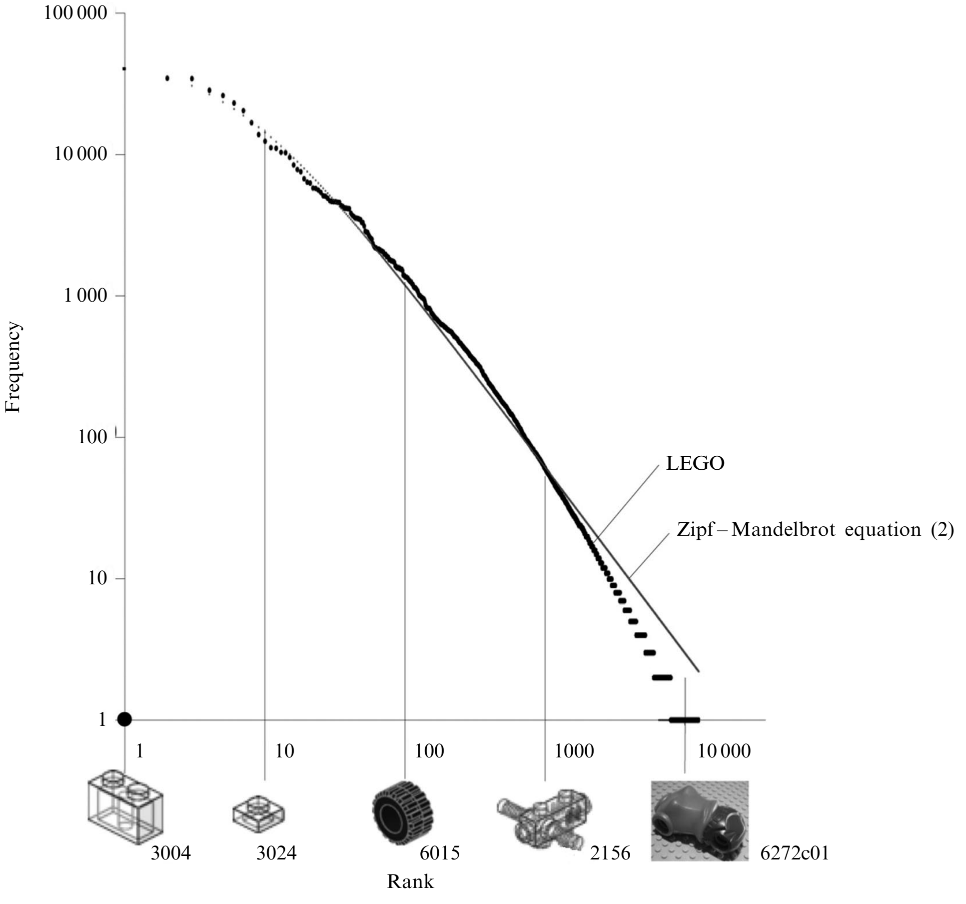

Figure 3. Rank–frequency Zipf plot for LEGO® in 2008 showing the 1st, 10th, 100th, 1000th, and 10 000th most frequently occurring pieces.

in nature and the social sciences and are unlikely to arise by chance, Newman (2005) has many examples. However, Zipf's law is controversial since there is no agreed explanation for it (Mandelbrot, 1997).

Figure 3 shows that transition from common to rare pieces occurs smoothly, indicating that there is no preferred Lego brick. It is not the case, as one might have expected, that Lego is based on cuboid bricks plus a few specials. Because of this it is not easy to sort Lego in a way that can assist a search for a particular brick because the pieces have no natural order. Pieces resemble each other in ways that form a family resemblance class so any method of sorting depending on one or two variables will always break down eventually. Children usually sort pieces into sets of the same colour and then rummage for particular bricks, which is why Lego is so noisy. In that case a trade-off occurs: do you spend time searching for the perfect piece or settle for an inferior brick which is at hand? Someone with a large heap would find it convenient to have more of the useful ones so that they are easier to find in the heap of bricks. This observation is similar to Zipf's original explanation of scaling in English. His qualitative argument based on least effort has been given quantitative support by a simple model of language that simultaneously minimises the effort of both the hearer and the speaker (Cancho and Sole, 2003). It would be an interesting project to try and adapt their model for Lego with the aim of seeing if the distribution of pieces provides an ideal compromise between pieces being easy to find and looking like what they are supposed to represent.

In any case, Zipfian distributions may be stable. A generalisation of the central limit theorem states that the sum of a number of random variables with power-law tail distributions can tend to a distribution of the same kind. In the right circumstances, if two information sources follow Zipf's law, then so will their mixture. The speech of children and adults can both follow Zipf's law, even though one may be more expressive than the other and as children imitate adult speech they learn to use words with a frequency that reinforces what they already do (Harremoes and Topsoe, 2001). Notice that Lego accumulates not piece by piece but by the addition of entire sets each of which approximately obeys Zipf's law, and which may obey it even more closely when combined.

5 The entropy of LEGO®

By calculating the probability that a piece will be of a particular type with respect to the ensemble of sets issued up to 2008 I find that any unique piece carries 19.8 bits. The piece with the least information is the commonest, the 2×1 brick, which carries 4.4 bits. The average, found using equation (1), is the zero-order entropy of Lego, which is 8.51 bits. Since about a thousand new pieces appear per year, the entropy of Lego is increasing. From equation (2) the increase was calculated to be 0.02 bits per year. The entropy of Lego stood at 8.56 bits for the 14 028 pieces issued up to November 2010. Table 1 compares the entropy of Lego with MECCANO and English.

Table 1. Zero-order entropy of LEGO®, MECCANO®, and English.

Lego	8.51 bits/piece
Based on 12 588 types of pieces listed in the Peeron database (http://peeron.com), April 2008	
MECCANO	5.1 bits/piece
Based on 243 types of pieces in a Number 10 set (the largest) in 1954.	
English	9.41 bits/word
Based on 8 727 words with a frequency given by Zipf's law (Shannon, 1950). For 12 588 words obeying Zipf's law the entropy would be 9.71 bits. The figure given in Shannon (1950) of 11.82 bits per word appears to be a numerical error (see Mackay, 2003, page 40).	

6 Buildings compared using LEGO®

The entropy of a model tells us something about its character. Since Lego, en masse, has an entropy of about 8.5 bits, individual models will have an entropy higher than this if they contain a lot of unusual bricks and a lower figure if they use common bricks. The models of 20th-century buildings in figure 4 are sold as LEGO® Architecture kits. Table 2 shows the number of pieces in each set and compares their entropy with some other typical sets. The models are treated here as if they were minimal representations on the assumption that they are economical to produce as well as easy to build and to recognise. Large models can become pixilated and make do with low-entropy bricks; small model are more informative because they must ingeniously use bricks that imitate a building's features.

Lego Architecture kits do not exploit the exotic possibilities of Lego. The entropy of five of them is close to 7 bits per piece: a low value similar to the entropy of the thousand most common pieces which is 7.74 bits and far less than a typical modern outfit such as the *Steven Spielberg MovieMaker Set* with an entropy of 9.6 bits. Since they are basically rectilinear blocks, this is perhaps not surprising. The Seattle Needle

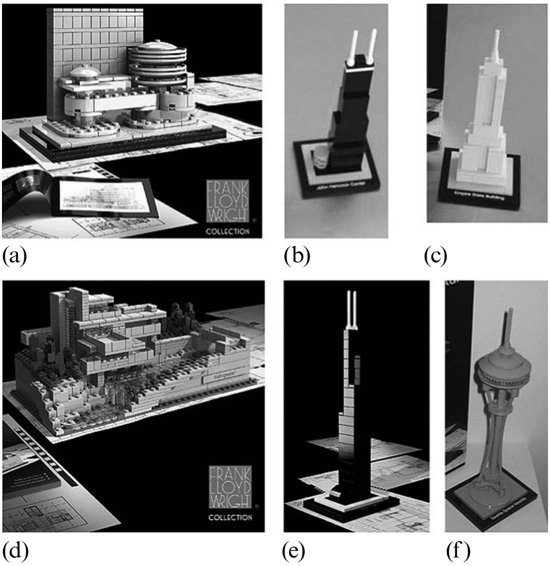

Figure 4. LEGO® models: (a) Guggenheim Museum, (b) Hancock Tower, (c) Empire State, (d) Falling Water, (e) Sears Tower, (f) Seattle Needle.

Table 2. The entropy, *H* bits per piece, of LEGO® sets.

	Pieces	*H*
Falling Water	811	6.8
Empire State	78	6.8
Sears Tower	70	7.2
Hancock Tower	69	7.5
Guggenheim	210	7.6
Seattle Needle	57	10.3
LEGO® System Hobby and Model Box (Modernist house, Set 752-1, 1963)	359	5.9
Universal Building Set (1973)	483	6.6
Steven Spielberg MovieMaker Set (Set 1349-1, 2000)	433	9.6

is the exception. Its space-age form, like a flying saucer on a tripod, makes use of unusual pieces that give it high entropy. Note that the two Frank Lloyd Wright models require far more pieces than the others: evidence that they are more complex than the towers, even though much smaller in reality.

The development of Lego to include machines and fantasy has moved in the opposite direction to architecture. The Seattle Needle excepted, the buildings in figure 4 could all be made with sets similar to the *Universal Building Set* of 1973 which reached only 6.75 bits per piece. Today it might be considered a little tame. It does not have to be like this. In times past, buildings often had unique features such as statues or decoration which would be difficult to represent in a minimal Lego model without introducing special high-entropy pieces. The entropy of architecture is, I conjecture, falling, whilst the entropy of Lego is growing.

It may be significant that where the Lego Company has been unsuccessful has been in launching low-entropy sets. Modulex, introduced by the Lego Company in 1963, was a variant of Lego marketed to architects. Its pieces were multiples of 5 mm cubes, compatible with ordinary Lego bricks though differently proportioned and with far fewer types of pieces than ordinary Lego. It did not sell well, although architect Eero Saarinen, was reputed to have made prototype structures with it. In the same year Lego marketed a series of architectural sets aimed at adults which included graph paper and a ruler along with instructions for designing scale models. These sets did not have any specialised roof, window, or door elements but were all blocks and plates suitable for the international-style buildings in the ideas book. The entropy of these three sets (numbers 750, 751, 752), was 5.9 bits per piece, the lowest of any I have measured. They were discontinued in 1965. Being restricted in what they could express they correspond, linguistically, to a dialect with a small vocabulary.

7 Conclusion

We can now appreciate why Lego is so versatile. An ideal modeling system ought to maximise entropy yet minimise the number of parts. These contradictory requirements are met by Lego offering parts with a range of entropies—from 4.4 to 19.8 bits—along with the possibility of new pieces being invented as required. Lego is not the arbitrary creation of a single mind but is a lawful language of form, which, since it is adapted to represent the built environment, ought to be of interest to readers of *Environment and Planning B*.

The entropy of Lego is increasing with the drive for fresh pieces coming from modeling new types of environment such as Harry Potter scenes and computer games like the *Prince of Persia*. Lego is a place where the built environment and virtual environments coexist and interact and its entropy has increased to enable this to happen. This is quantitative evidence of the limited vocabulary of modern architecture which, although it may be elegant and pure, makes little use of the growth in complexity that has come with digital worlds. Seen through the eyes of a child Lego makes modernist architecture look old fashioned.

It has been observed that the Shannon information content of a building depends upon its context. Cities intensify and speed up many aspects of our lives; here we have evidence that this may also apply to being architecturally interesting because parts of buildings will seem more improbable where there are a lot of other things to see. Cities amplify the entropy of buildings.

References

Cancho R F, Sole R V, 2003, "Least effort and the origins of scaling in human language" Proceedings of the National Academy of Sciences **100** 788 – 791

Chaitin G, 1975, "Randomness and mathematical proof" *Scientific American* May, pp 47 – 52; reprinted in 1987 *Papers on Algorithmic Information Theory* G Chaitin (World Scientific, Singapore) pp 3 – 13

Harremoes P, Toopsoe F, 2001, "Maximum entropy fundamentals" *Entropy* **3** 191 – 266

Kiemle M, 1967 *Ästhetische Probleme der Architektur unter dem Aspekt der Informationsåsthetik* (Verlag Schnelle, Quickborn)

MacKay D, 2003 *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, Cambridge)

Mandelbrot B B, 1997 *Fractals and Scaling in Finance* (Springer, New York) pp 198 – 218

Newman M E J, 2005, "Power laws, Pareto distributions and Zipf's law" *Contemporary Physics* **46** 323 – 351

Shannon C E, 1950, "Prediction and entropy of printed English", reprinted in *Claude Elwood Shannon, Collected Papers* 1993, Eds N J A Sloane, A D Wyner (IEEE Press, New York) pp 198 – 202

Smithson R, 1979 [1966], "Entropy and the new monuments", in *The Writings of Robert Smithson* Ed. N Holt (New York University Press, New York) pp 9 – 18

Stamps A E III, 2002, "Entropy, visual diversity, and preference" *The Journal of General Psychology* **129** 300 – 320

Stamps A E III, 2003, "Advances in visual diversity and entropy" *Environment and Planning B: Planning and Design* **30**

von Buttlar A, Selig H, Wetzig A, 1972, "Erhaltenswerte Stadtbildelemente des Münchener Cityrandgebietes Lehel" *Deutsche Kunst und Denkmalpflege* **30** 65 – 71

Wolpert L, 1973, "How many bits?" *Architectural Review* April, pp 251 – 252

Yule G, 1996 *The Study of Language* (Cambridge University Press, Cambridge) pp 19 – 25