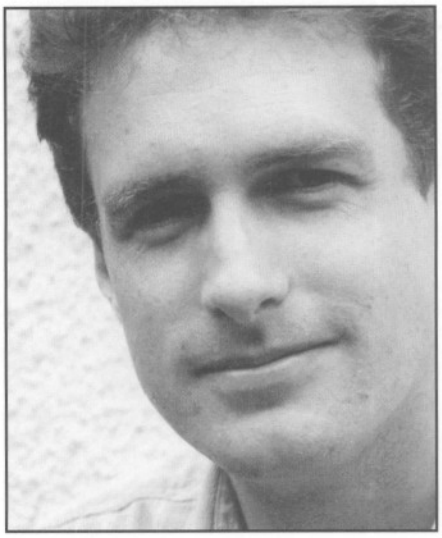
Simplicity and the mind



Nick Chater *gave his Spearman Medal Award lecture at the Society's London Conference in December 1996.*

Requests for reprints should be addressed to:
Professor Nick Chater
Department of Psychology
University of Warwick
Coventry CV4 7AL
E-mail: nick.chater@warwick.ac.uk

WE all seem to be concerned about simplicity. Psychologists, like other scientists, are impressed by simple explanations. We ask: why accept complicated theory X, when simple account Y fits the facts just as well? And our intuitions that simple is best are also expressed in methodological injunctions to 'apply Occam's razor', adopt a principle of parsimony, and to avoid explanations with 'too many degrees of freedom'. Scientists frequently admit to being guided by 'aesthetic' criteria, preferring 'beautiful', 'elegant' theories, and such aesthetic intuitions appear to be closely bound up with judgements about simplicity. For example, what makes Ptolemaic astronomy turn out to be so ugly is the need to add in the complexity of all those epicycles.

A preference for simple explanations extends beyond science. Juries are unimpressed by convoluted explanations of a defendant's whereabouts or actions. The more complex the story becomes, the more 'epicycles' are added to fit the evidence and the more implausible it starts to sound. And a preference for simplicity applies in our attempts to understand the everyday world around us — when a light goes out suddenly, I assume that the bulb has blown; but it could be that an intruder has very quietly entered the house, hit the light switch, and hidden themselves in the shadows. This *is* possible; but wildly implausible. I'll put my money on the simpler explanation.

Science Museum/Science and Society Picture Library

Ptolemy of Alexandria

Finally, consider aesthetics. I've already mentioned aesthetic judgements about scientific theories, but aesthetic judgements quite generally seem sensitive to simplicity. From our pleasure in 'economy of language' to our preferences for the symmetries of a snowflake rather than the muddle of a lump of mud, simplicity appears to be a guiding aesthetic principle. Simplicity is not, of course, the whole story about aesthetics — simple shapes like a plain white square seem to hold little aesthetic interest — but it seems to be at least *part* of the story.

Why do we care about simplicity? I propose that this is because the search for simplicity is a fundamental goal of cognition — and that this is an appropriate goal, not for aesthetic reasons, but in view of profoundly utilitarian concerns. Simple patterns, hypotheses, explanations or theories, are the *most reliable* patterns, hypotheses, explanations or theories. 'Most reliable' can be spelled out in various ways: the most likely to be true; providing the most accurate predictions; or providing the best basis for deciding how to act. So simplicity is not just aesthetically appealing — it is of crucial practical importance.

Quantifying simplicity

We all have intuitions about what is simple and what is complex. But a theory of simplicity needs to explain these intuitions, rather than taking them for granted. Over the last 30 years, a beautiful (yes, simple!) theory of simplicity has been developed in mathematics and computer science, under the somewhat forbidding name of Kolmogorov complexity theory. This theory is little known in psychology, but has, I will suggest, important psychological applications.

In essence, Kolmogorov complexity theory defines the complexity of an object as the length of the shortest description that uniquely specifies that object. The idea is that simple things have short descriptions; complex things

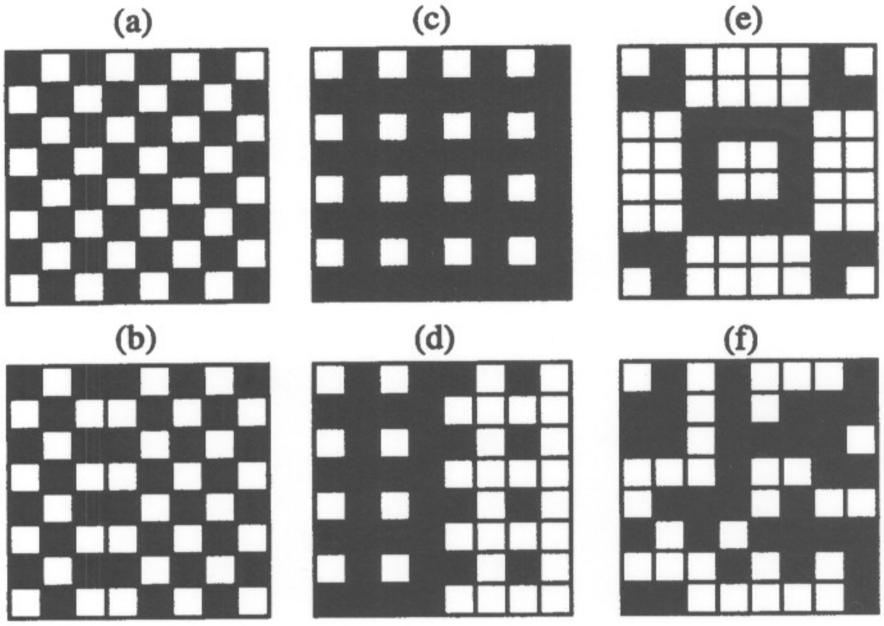


Figure 1: Patterns of varying simplicity. Note that (b) seems intuitively more complex than (a). This may reflect the fact that specifying (b) requires specifying 'discontinuity' between the third and fourth columns. Similarly, (c) seems simpler than (d). Specifying (d) requires specifying the inversion between black and white between the left and right halves of the figure. (e) is simple and its symmetries make it easy to specify briefly. (f) appears to be least simple, and does not have a short description. It was generated by flipping a coin. In general, simple patterns appear to be those which can be specified briefly.

have long descriptions. But, as stated, this formulation falls victim to the celebrated Richard-Berry paradox. Consider the following description:

the smallest natural number that cannot be uniquely specified in less the twenty words of English (1)

There must be infinitely many numbers which cannot be described in less than 20 words, because there are infinitely many natural numbers, but only finitely many descriptions in English of 20 words or less. So out of this infinite set of numbers, one must be the smallest — call it *N*. But now note that the description in (1), which contains just 16 words, *does* uniquely specify *N*. This contradicts the definition of *N*, which states that *N* cannot be specified in less than 20 words. Hence the paradox!

Kolmogorov complexity avoids the paradox by a simple restriction — that descriptions must be sets of instructions which, when followed, will lead to the *construction* of the object described. That is, descriptions are computer programs, and the Kolmogorov complexity of an object, $K(x)$, is the length of the shortest computer program that will generate x.

So a description like (1) is not allowed: partly because it is stated in English, rather than a programming language where there is a precise interpretation of the language as a set of instructions; and crucially because it does not provide a *procedure* which produces the number which it describes. So the paradox disappears.[a]

The paradox removed, we have a definition of simplicity (or rather, complexity) as the shortest computer program that generates an object. So, the strings 111111111111111111 and 101010101010101010 are simple — they have short descriptions as a string of '1's or alternating '1's and '0's. But the sequence 10011101011011010001000 is not simple, as it follows no pattern. It can only be specified by simply listing it verbatim (in fact, it was generated by flipping a coin), and hence no brief encoding is possible. Figure 1 shows some examples of simple objects as compared with complex objects.

Three immediate concerns regarding this definition of simplicity naturally arise. The first is that, if simplicity is shortest description length, then surely it matters in which language we write our descriptions. Surprisingly, Kolmogorov complexity is language-invariant. A central and remarkable result in Kolmogorov complexity theory states that $K(x)$ assessed in one language will be the same as $K(x)$ assessed in another language, up to a constant. This is what justifies talking about *the* Kolmogorov complexity of an object. But, nonetheless, the constant, by which description lengths may differ, may be large enough to be important in many psychological contexts. Thus, a detailed psychological theory of similarity requires an account of how information is represented by the mind. Much of perceptual and cognitive psychology is concerned with providing evidence about the nature of mental representation — so it should be possible to use this evidence to constrain a psychological theory of simplicity. But, as we shall see below, the approximate language-invariance of Kolmogorov complexity means that we can make a lot of progress in developing psychological theory without having to make specific assumptions about mental representation.

The second concern is that the mathematical notion of Kolmogorov complexity refers to the shortest possible description of an object, but the cognitive system may not be able to find that shortest description. For example, the digits of the expansion of π are simple (a short program will compute them), but they appear to be completely random to the cognitive system (indeed they even appear random to intensive statistical analysis — the digits of π pass all known tests for randomness). The cognitive system cannot find the shortest possible description for an object; but it can choose the shortest description *that it can find*. So a cognitive notion of simplicity must inevitably only approximate the ideal notion of simplicity given by the mathematical theory of Kolmogorov complexity.

The third concern is that all this discussion of *computer* programs sounds rather unpsychological, and hence may appear to suggest that Kolmogorov complexity cannot apply to human cognition. But, in this context, the notion of a computer programming language is very broad — all that is meant is that the language of instructions can be followed by some mechanistic procedures (and the brain presumably embodies mechanistic procedures, even though these are astonishingly complex and little understood). In addition, we require that the language with which information is represented by the mind is 'universal'. But, despite its rather imposing name, the constraint that the language must be universal turns out to be surprisingly weak, and almost all psychological theories about how the mind represents information meet this constraint. In sum, despite its unpsychological sound, the notions of Kolmogorov complexity may nonetheless apply to human cognition.

Having outlined how simplicity can be quantified, I now consider *why* simplicity matters: specifically, simple explanations are, other things being equal, more likely to be true than complex explanations.

Simple = most probable

I started by noting that people find simple explanations the most believable. The more convoluted and elaborate an explanation, the less likely it seems. But can this intuition be justified? It turns out that it can: simple explanations have, other things being equal, the highest probability of being true. Suppose that we have data, D, and a set of hypotheses about that data. The most likely hypothesis is the hypothesis, H, that has the greatest probability, given the data. In symbols, this is the H that maximises $P(H|D)$. Bayes' theorem, a standard theorem of probability theory, states that:

$P(H|D) \propto P(D|H)P(H)$ (2)

That is, the probability of the hypothesis given the data is proportional to the product of the probability of the data given the hypothesis and the prior probability of the hypothesis. By elementary mathematics, the H that *maximises* (2) is the same as the H that *minimises* (3):

$-\log_2 P(D|H) - \log_2 P(H)$ (3)

Most psychologists have, at one time or another, encountered information theory, which was widely applied to understanding perception and cognition in the 1950s and 1960s. Information theory deals with how information can be represented in codes, and considers properties of codes such as their length, noise-resistance, redundancy, and so on. It specifies that (for an optimal code, which minimises average code length), probable events are given short codes, and improbable events are given long codes. Specifically, for an event x with probability $P(x)$, the code length is $-\log_2 P(x)$. But the terms in equation (3) have exactly this form, so we can interpret (3) as follows:

code length for D in terms of H +
code length for H (4)

But note that the sum of (4) just gives the total code length for specifying the data, D, by means of the hypothesis H. Specifically, this code has two parts — a part which specifies the hypothesis, and a part which uses the hypothesis to specify the data. The equivalence between (2) and (4) says that the *most probable* hypothesis is also the hypothesis that provides the *simplest* (i.e. shortest) overall specification of the data. More informally, the most probable hypothesis is the one which provides the simplest explanation. So our intuitive preference for simple explanations is vindicated — simple explanations are the most likely to the true.[b]

Cognition as a search for simplicity

The story so far: people have a preference for simple explanations; simplicity can be measured by Kolmogorov complexity; and preferring simple explanations seems to be a good idea, because simple explanations are the most probable. But is there reason to believe that the mind really does search for simplicity?

Perhaps the richest source of evidence comes from the study of perceptual organization: how a description of the structure of a perceptual stimulus is derived from sensory data. There are indefinitely many possible descriptions compatible with the sensory input — what determines which is chosen? A description of the structure of the stimulus can be viewed as an 'explanation' of the sensory data — thus, according to the present account, the best pattern should be the one which allows the simplest (briefest) encoding of the sensory data. This 'simplicity principle' (Pomerantz & Kubovy, 1986) can be traced from Mach (1959/1886) through Gestalt psychology (Koffka, 1962/1935) to information processing research on perception (Buffart *et al.*, 1981; Garner, 1962, 1974; Hochberg & McAllister, 1953; Leeuwenberg, 1969, 1971; Leeuwenberg & Boselie, 1988).

So how does the simplicity principle work in practice? Consider how occluded figures are completed. In Figure 2(a) the upper left and right lines are viewed as part of a single line, occluded by a solid vertical bar. The vertical bar is assumed to continue in a straight line (see Figure 2(b)), although it could have any form (e.g. Figure 2(c)). The straight line is favoured by simplicity. Note that the lower lines, which are not aligned, are perceived as distinct, and as embedding into the side of the vertical bar. This illustrates the Gestalt law of good continuation — that organizations which allow continuous lines are preferred; but notice that this can be interpreted in terms of simplicity, because locating a single straight line (which is occluded) is simpler than independently specifying two separate straight lines to the left and right of the stimulus. But if the two lines are not aligned, then the simplicity advantage of postulating the 'hidden' line disappears, and no hidden line is perceived. Figure 2(d) is perceived as a cross, occluded by a circle (this interpretation is illustrated in Figure 2(e)); whereas good continuation would lead to the interpretation shown in Figure 2(f). The preference for the interpretation shown in Figure 2(e) follows the Gestalt principle of good form, but it also follows by assuming preference for simplicity, because the cross is simpler than the irregular structure in Figure 2(f).

A preference for simplicity may underly not just perception, but also learning from experience. Learning from experience, like perception, is a problem of finding patterns in what are typically large amounts of complex and often noisy data. A preference for simple explanations may guide how we learn the structure of the environment from experience. As for perception, there are existing theories of learning which assume that learning is guided by simplicity. For example, Brent and Cartwright (1996) show how structure within words can be found within isolated words, and Wolff (1977; see also Redlich, 1993) considers how other linguistic units, such as words and phrases, can be found automatically. But, less directly, theories of learning which are not superficially concerned with simplicity can, nonetheless, be viewed as maximising simplicity. For example, learning methods used by connectionist networks, currently the most popular models of learning in cognitive psychology, can be viewed as maximising simplicity. Thus, a range of current psychological proposals concerning learning are compatible with the thesis that the cognitive system maximises simplicity.

Finally, note that the claim that the mind searches for simplicity may have a natural application in understanding memory: that the cognitive system seeks to minimise memory load. This leads to

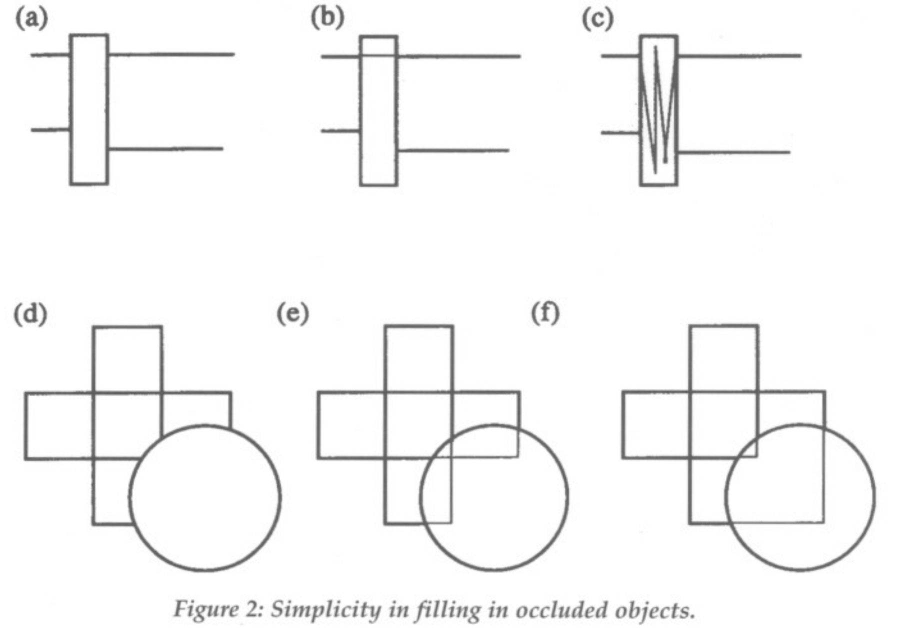


Figure 2: Simplicity in filling in occluded objects.

the prediction that the richer the patterns that the cognitive system can find in a stimulus, the more simply it can be encoded, and hence the better it will be remembered. This is a ubiquitous finding in all areas of memory research, from the advantage of memory for words over non-sense strings, meaningful over non-meaningful pictures, and comprehensible over non-comprehensible stories.

In the context of memory, however, the proposal that the cognitive system searches for the simplest possible explanation, and that memory load is therefore determined by the length of that explanation, may appear to involve an implausible presupposition: that memories are stored as briefly as possible — i.e. with no redundancy. It is often pointed out that this kind of storage would be inappropriate, because it would mean that memories would be lost if subjected to the slightest damage. But the link between simplicity and memory is actually quite compatible with the claim that memories are stored in a redundant, robust fashion. It turns out that the best way to produce a robust code is first to find the simplest encoding, and then introduce redundancy so that each part of this code is equally protected from corruption (Cover & Thomas, 1992). Thus, for a given stimulus, finding a brief encoding will allow the construction of a better redundant representation, which will thereby be more noise resistant and hence better remembered.

Conclusion

People are attracted to simplicity — but, we have seen, this is more than a mere aesthetic preference. A preference for simplicity is a key to choosing the most likely explanations of the information that we receive. Hence the search for simplicity may be an important cognitive goal across many areas of cognition, from perceptual organization, to learning, to everyday reasoning and scientific thinking.

Acknowledgement

I would like to thank Mark Ellison, Steven Finch, Ulrike Hahn, Peter van der Helm, Emmanuel Leeuwenberg, James McClelland, Mike Oaksford, Martin Pickering, Emmanuel Pothos, Martin Redington, Jerry Seligman, Julian Smith, Paul Vitányi and Johan Wagemans for valuable discussions of these ideas. Li and Vitányi (1997) provide an excellent and comprehensive introduction to Kolmogorov complexity theory. Some key ideas are outlined more informally in Chater (1996, submitted).

Footnotes

a It is interesting to note, though, that the Richard-Berry paradox does re-emerge in a mutated form, and that Chaitin (1974, 1975) has used it to provide a new and remarkably simple proof of Gödel's theorem, a central theorem in modern mathematics.

b But, you may ask, what happened to Kolmogorov complexity? The argument above just mentioned the more familiar information theory, instead. The answer is that the information theory argument given above is suggestive but not really rigorous (specifically, no justification is given for assuming that an optimal code should be used, and thus that code lengths should be $-\log_2 P(x)$). The rigorous argument uses the fact that, under very general conditions, $-\log_2 P(x)$ is approximated by the Kolmogorov complexity of x, $K(x)$. In general, Kolmogorov complexity can be viewed as a generalization of information theory — so the spirit of the argument given above is correct.

References

Brent, M.R. & Cartwright, T.A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, **61**, 93-126.

Buffart, H., Leeuwenberg, E. & Restle, F. (1981). Coding theory of visual pattern completion. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 241-274.

Chaitin, G.J. (1974). Information-theoretic limitations of formal systems. *Journal of the Association of Computing Machinery*, **21**, 403-424.

Chaitin, G.J. (1975). Randomness and mathematical proof. *Scientific American*, **232**, 47-52.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organisation. *Psychological Review*, **103**, 566-581.

Chater, N. (submitted). The search for simplicity: A fundamental cognitive principle?

Cover, T.M. & Thomas, J.A. (1992). *Elements of Information Theory*. New York: Wiley.

Garner, W.R. (1962). *Uncertainty and Structure as Psychological Concepts*. New York: John Wiley.

Garner, W.R. (1974). *The Processing of Information and Structure*. Potomac, Md: Lawrence Erlbaum.

Hochberg, J. & McAlister, E. (1953). A quantitative approach to figure 'goodness'. *Journal of Experimental Psychology*, **46**, 361-364.

Koffka, K. (1962). *Principles of Gestalt Psychology*, 5th edn. London: Routledge and Kegan Paul. (Original work published in 1935.)

Leeuwenberg, E. (1969). Quantitative specification of information in sequential patterns. *Psychological Review*, **76**, 216-220.

Leeuwenberg, E. (1971). A perceptual coding language for perceptual and auditory patterns. *American Journal of Psychology*, **84**, 307-349.

Leeuwenberg, E. & Boselie, F. (1988). Against the likelihood principle in visual form perception. *Psychological Review*, **95**, 485-491.

Li, M. & Vitányi, P. (1997). *An Introduction to Kolmogorov Complexity and its Applications*, 2nd edn. New York: Springer-Verlag.

Mach, E. (1959). *The Analysis of Sensations and the Relation of the Physical to the Psychical*. New York: Dover. (Original work published 1886.)

Pomerantz, J.R. & Kubovy, M. (1986). Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In K.R. Boff, L. Kaufman & J.P. Thomas (Eds), *Handbook of Perception and Human Performance, Vol. II: Cognitive Processes and Performance*, pp.36:1-45. New York: Wiley.

Redlich, A.N. (1993). Redundancy reduction as a strategy for unsupervised learning. *Neural Computation*, **5**, 289-304.

Wolff, J.G. (1977). The discovery of segmentation in natural language. *British Journal of Psychology*, **67**, 377-390.