

## The Search for Simplicity: A Fundamental Cognitive Principle?

Nick Chater

*University of Warwick, Coventry, U.K.*

It is proposed that the cognitive system imposes patterns on the world according to a simplicity principle: Choose the pattern that provides the briefest representation of the available information. The simplicity principle is normatively justified—patterns that support simple representations provide good explanations and predictions on the basis of which the agent can make decisions and actions. Moreover, the simplicity principle appears to be consistent with empirical data from many psychological domains, including perception, similarity, learning, memory, and reasoning. Thus, the simplicity principle promises to serve as the starting point for the *rational analysis* of a wide range of cognitive processes, in Anderson's (1990, 1991a) sense. The simplicity principle also provides a framework for integrating a wide range of existing psychological proposals.

The cognitive system must cope with a world that is immensely complex but that is, nonetheless, highly patterned. The patterns are crucial. In a completely random world, prediction, explanation, and understanding would be impossible—there would be no patterns on which prediction could be based, to which explanations could refer, or the comprehension of which could amount to understanding. Even more fundamentally, without any patterns relating actions to consequences, there would be no basis to choose one action rather than another.

The ability to find patterns in the world is therefore of central importance throughout cognition. Without the ability to find such patterns an agent might as well be in a random world: It would be able to predict, explain, and understand nothing; and it would have no basis on which to choose its actions. By contrast, the cognitive systems of people and

---

Requests for reprints and correspondence concerning this article should be sent to Nick Chater, Department of Psychology, University of Warwick, Coventry CV4 7AL, U.K. E-mail: [nick.chater@warwick.ac.uk](mailto:nick.chater@warwick.ac.uk)

I would like to thank Frans Boselie, Mark Ellison, Steven Finch, Ulrike Hahn, Peter van der Helm, Emmanuel Leeuwenberg, James McClelland, Mike Oaksford, Martin Pickering, Emmanuel Pothos, Martin Redington, Jerry Seligman, David Shanks, Julian Smith, and Paul Vitányi for valuable discussions of these ideas at various stages in their development or for comments on this manuscript. A brief and informal outline of some of the material here is given in an article for *The Psychologist* (Chater, 1997).

This paper is based on an EPS Prize Lecture (for distinguished work in experimental psychology or a cognate discipline by a person at an early stage of her or his career), delivered at the March 1997 EPS meeting at Oxford University.

---

animals appear to be conspicuously successful in coping with the world. Somehow, cognitive processes are able to find patterns successfully.

How is this success achieved? Any proposal must meet two adequacy criteria: (a) It must be *normatively justified*—without such normative justification, the *success* of the method of finding patterns is mysterious; (b) It must be *descriptively correct*—it must accord with empirical data—at least to some approximation. A theory that is both normatively justified and descriptively correct provides a *rational analysis* of a cognitive process (for discussion of this concept see, for example, Anderson, 1990, 1991a, 1991b; Anderson & Schooler, 1991; Oaksford & Chater, 1994, 1995a, in press). So explaining how the cognitive system successfully finds patterns requires providing a rational analysis of the cognitive system's pattern-finding capabilities.

I propose that patterns are found by following a fundamental principle: Choose the pattern that provides the *simplest* explanation of the available data. Moreover, I suggest that the simplicity principle has very broad scope and hence can be used as a starting point for the detailed rational analyses of a wide range of cognitive processes.

The idea that cognition involves a search for simplicity has a long lineage, in the discussion of both normative and descriptive issues. On the normative side, the injunction to favour simple scientific theories can be traced to William of Ockham<sup>1</sup> (1285–1349) and is endorsed by Newton (see Li & Vitányi, 1997, p. 317). Simplicity was also assigned fundamental importance in early positivist epistemology (e.g. Mach, 1883/1960), and it remains a standard principle in modern philosophy of science (e.g. Sober, 1975). Simplicity is also recognized as important in statistics. If a straight line and a cubic fit the same data equally well, then the straight line is the preferred model because it is simpler—it contains fewer adjustable parameters. Indeed, without some implicit adherence to a simplicity principle, classical statistical approaches to modelling data would be incoherent, because increasing the generality of a model (e.g. switching from a straight line to a cubic) can only improve the fit with the data. Aside from its importance in statistics, a preference for simple explanations is also a standard methodological principle in informal scientific discourse—for a prominent psychological example, see Pylyshyn's (1984) discussion of the importance of having fewer model parameters than data points in cognitive modelling. But although the preference for simple patterns has been widely recognized, simplicity has typically remained a largely intuitive notion. Over the last thirty years, however, a rich and important theory of simplicity—Kolmogorov complexity—has been developed and widely applied by mathematicians (Chaitin, 1966; Kolmogorov, 1965; Solomonoff, 1964; for an overview, see the excellent textbook by Li & Vitányi, 1997), statisticians (Rissanen, 1987, 1989; Wallace & Freeman, 1987), and computer scientists (Quinlan & Rivest, 1989; Wallace & Boulton, 1968). This theory allows rigorous normative justifications to be given for why choosing the simplest pattern leads to the best explanations and predictions, and it also allows the more concrete formulation of the

---

<sup>1</sup> The original formulation of "Occam's razor" is that explanations postulating the smallest number of entities should be preferred; thus it embodies a specific measure of simplicity in terms of number of objects involved in an explanation. Occam's razor has since been interpreted more broadly as expressing a preference for simple explanations.

psychological proposal that cognition seeks to find the simplest pattern. This account of simplicity and its potential application to cognition is outlined in this paper.

Simplicity has also been frequently viewed as important from the point of view of *describing*, rather than justifying, cognitive processes. Mach (1886/1959), one of the strongest advocates of simplicity of a normative principle in science, also proposed that the perceptual system seeks to find the simplest representations of sensory input. This viewpoint is echoed in the proposal in the Gestalt tradition that perceptual organization is chosen to maximize “*prägnanz*” (Koffka, 1935/1962), a notion closely related to simplicity, which aims to integrate the range of specific Gestalt principles of perceptual organization (good form, good continuation, and so on). Moreover, Hochberg and McAlister (1953) explicitly identified the goal of perceptual organization as maximizing simplicity, and this work, was followed by a variety of related proposals, where simplicity is measured in different ways (Buffart, Leeuwenberg, & Restle, 1981; Garner, 1962, 1974; Leeuwenberg, 1969, 1971). Moving from perception to the psychological processes involved in scientific inference, simplicity has also frequently been invoked as an important guiding principle. For example, scientists frequently report strong aesthetic preferences in theory construction and evaluation, using terms such as “simplicity”, “elegance”, “parsimony”, and so on, to describe desirable properties of theoretical proposals. Einstein has been attributed with the remark that “Everything should be made as simple as possible, but not simpler” (Eysenck & Keane, 1990). This preference for simplicity (or more generally, beauty) is sometimes expressed so strongly that it even overrides the concern to fit the data (e.g. this view is attributed to the great theoretical physicist Paul Dirac in Stewart & Golubitsky, 1992). Thus simplicity has been implicated as a guiding principle in finding patterns, from perceptual processing to scientific reasoning. I propose that simplicity may have an even more general role in cognition: ranging from reasoning and memory to learning and similarity.

This paper has three parts. The first introduces the problem of finding patterns in data, and why it is normatively and descriptively puzzling, essentially because there are an infinite number of patterns consistent with any finite set of data. The second part considers the normative question of how patterns *should* be found. I outline how simplicity can be quantified in terms of the mathematical theory of Kolmogorov complexity and how this theory explains why searching for simple patterns is normatively justified as a strategy for predicting and explaining the world, and as a partial basis for deciding how to act. The third part considers the descriptive problem of how various cognitive processes actually *do* find patterns. The approach is programmatic—I aim to provide an integrated framework for apparently diverse cognitive problems and to suggest directions for future research, rather than attempting a definitive account in any one area. Overall, I hope to show that simplicity is both a normatively justified and descriptively plausible account of how the cognitive system finds patterns in a range of domains.

## THE PROBLEMS OF FINDING PATTERNS

Consider the problem of finding patterns in a finite portion of an infinite sequence. In the portion of the sequence that we observe, just two states are found. Let us call the binary values “black” and “white” to allow the visual representation shown in Figure 1a. In this

finite sequence, an intuitively evident pattern is that there is an alternation of the two states. If this pattern is correct, then the sequence should continue as shown in Figure 1b. But another pattern, equally consistent with the observed data, is that there is an infinite sequence of “white”, followed by an alternating sequence of white and black, and then an infinite sequence of “black”. The observed data are assumed to correspond to the middle part of this sequence (Figure 1c). Moreover, a further pattern consistent with the data consists of a jumble of states (many not occurring in the observed part of the sequence at all—represented by patterned squares in the figure) to the left and right of the alternating white and black items that are observed. Again this kind of pattern is precisely consistent with the observed data. More generally, it is clear that an infinite number of patterns are consistent with any finite set of data.

A similar example, of traditional psychological interest (e.g. Dinnerstein & Wertheimer, 1957; Kanizsa & Gerbino, 1982), concerns the completion of occluded figures (Figure 2). The intuitively natural completion of the occluded region in Figure 2a interprets the figure as a square partially occluded by another square (Figure 2b). This completion is predicted by two Gestalt principles: good continuation, which states that lines should be assumed to continue as smoothly as possible, and good form, which states that completions should prefer regular underlying figures. But, again, an infinite number of alternative completions are possible (Figure 2c).

The hard-headed psychologist may feel tempted to dismiss the rather bizarre patterns shown in Figures 1 and 2 as “silly”. Of course, such a psychologist might say that the cognitive system is only concerned with “sensible” patterns and bases its explanations, predictions, and decisions on these. The psychologist might go on to point out that the really interesting issue is how the cognitive system copes with cases where two “sensible” patterns can be imposed on a stimulus, and some choice must

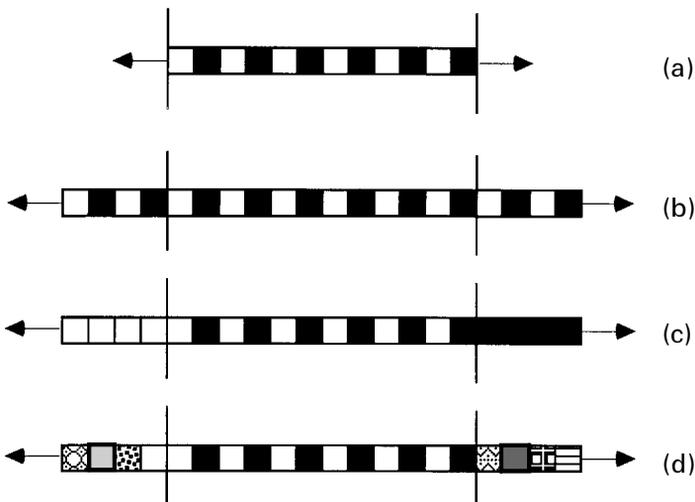


FIG. 1. An infinite number of incompatible patterns are compatible with any finite sequence of data.

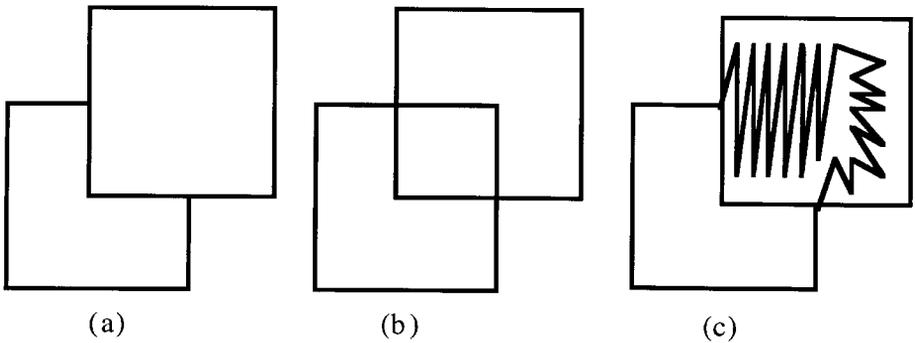


FIG. 2. Occluded figures are consistent with an infinite number of continuations.

be made between them. But this impatient response misses the point. The psychologist must explain our intuitions about which patterns are “silly” and which are “sensible”, and cannot take them for granted, because these intuitions are themselves the outcome of psychological processes. Indeed, these intuitions must be explained in two ways. First, some *normative justification* must be given for assuming that the cognitive system is justified in favouring “sensible” patterns and basing its predictions, explanations, and decisions on these. This issue is the focus of the next section. Second, the *descriptive* question of how the cognitive system differentiates between “silly” and “sensible” patterns must also be addressed—I leave descriptive issues to the final section of the paper.

## FINDING PATTERNS: THE NORMATIVE PROBLEM

Despite our strong intuitions that not all patterns consistent with a finite set of data are equal (i.e. that some are plausible and others are absurd), there has been a long sceptical tradition in philosophy arguing that no normative justification can be given for such preferences (e.g. Goodman, 1983; Hume, 1739–1740/ 1965; Popper,<sup>2</sup> 1934/ 1959). But this scepticism is unattractive, because it makes utterly mysterious the remarkable and consistent success that cognitive systems enjoy on the basis of favouring some patterns over others.

Fortunately, the sceptical challenge can be addressed by applying the mathematical theory of Kolmogorov complexity. This theory quantifies simplicity and shows that a preference for simpler patterns is justified, because describing the world in terms of simple patterns consistently leads to better predictions, explanations, and decisions.

Before considering how this theory measures simplicity, however, we must first ask: *What* should be measured for simplicity?

<sup>2</sup> Popper allows that different hypotheses might be differentially favoured for investigation on the basis regarding their falsifiability—but this does not bear on the examples given here, because they are all equally specific, and therefore equally falsifiable.

## The Simplicity of What?

In choosing patterns on the basis of simplicity, the most obvious suggestion is that the simplest available pattern should be preferred. This principle correctly favours an indefinitely long sequence of alternating black and white squares in Figure 1 and the “square” completion in Figure 2. But, taken at face value, it also has a paradoxical consequence: A very simple pattern, such as when the pattern in Figure 1 is an infinitely long sequence of black squares or when the pattern in Figure 2 is a simple uniform field, will always be preferred. Such possibilities are, of course, ruled out by the constraint that the pattern has to be consistent with the available data—thus, these “null” patterns are just too simple. But this point itself raises difficult questions: What does it mean for a pattern to be consistent with the available data? Can consistency with the input be traded against simplicity of interpretation? If so, how are simplicity and consistency with the data to be jointly optimized? We shall see that the theoretical account of simplicity presented below answers these questions.

There is, however, a further and more subtle difficulty: What rules out the simplest possible “null” pattern, if such a pattern could be interpreted as saying that “anything goes”? The null pattern will be consistent with the available data; indeed it would be consistent with any data because it rules nothing out. Mere consistency or compatibility with the data is plainly not enough; the pattern must also, in some sense, capture regularities in the data (Harman, 1965). But this appears to imply that choosing a pattern involves the joint optimization of two factors; and the relative influence of these two factors is unspecified. Moreover, this conclusion is unattractive because two notions—simplicity and explanatory power—must be explicated rather than just one.

Fortunately, there is an alternative way to proceed. This is to view a pattern as a way of encoding the data; and to propose that the pattern chosen is that which allows the simplest encoding of the data. This view disallows null or nearly null patterns, which bear little or no relation to the data, because these organizations do not help encode the data simply. It also provides an operational definition of the “explanatory power” of a pattern—as the degree to which that pattern helps provide a simple encoding of the data. If a pattern captures the regularities in the pattern (i.e. if it “explains” those regularities), then it will provide the basis for a brief description of the data; if an organization fails to capture regularities in the data, then it will be of no value in providing a brief description. Explanatory power is therefore not an additional constraint that must be traded off against simplicity; maximizing explanatory power is the same as maximizing the simplicity of the encoding of the data.

## Quantifying Simplicity

To apply the injunction to choose the pattern that provides the simplest encoding of the data, we need a measure of simplicity. There is a long tradition in philosophy of equating simplicity with brevity in some coding language (e.g. Kemeny, 1953). In psychology, this general approach has been applied in a variety of contexts, from the organization of simple sequences, such as the example we have just considered (Leeuwenberg, 1969; Restle,

1970; Simon, 1972; Simon & Kotovsky, 1963; Vitz & Todd, 1969), to judgements of “figural goodness” (Hochberg & McAlister, 1953), the analysis of Johansson’s (1950) experiments on the perception of motion configurations (Restle, 1979), and figural completion (Buffart et al., 1981). It has also been advanced as a general framework for understanding perceptual organization (e.g. Attneave & Frost, 1969; Leeuwenberg, 1971; Leeuwenberg & Boselie, 1988).

Approaches based on brevity of encoding in some description language appear to be dogged by two problems: (a) that a fresh description language must be constructed for each fresh kind of pattern; and (b) that the predictions of the theory depend on the description language chosen, and that there is no (direct) empirical means of deciding between putative languages.

Kolmogorov complexity theory addresses these problems. The first problem is avoided by choosing a general coding language. Specifically, the language chosen is a *universal programming language*. A universal programming language is a general purpose language for programming a computer. The familiar programming languages such as PROLOG, LISP, and PASCAL are all universal programming languages. How can an object, such as a perceptual stimulus, be encoded in a universal programming language such as LISP? The idea is that a program in LISP encodes an object if the object is generated as the output or final result of running the program. By the definition of a universal programming language, if an object has a description from which it can be reconstructed in *any* language, then it will have a description from which it can be reconstructed in the universal programming language. It is this that makes the programming language *universal*.

Moreover, in solving the first problem, the second problem—that different patterns of languages give different code lengths—is, at least partially, addressed. A central result of Kolmogorov complexity theory, the invariance theorem (Li & Vitányi, 1997), states that the length of the shortest description of an object,  $x$ , is invariant (up to a constant) between different universal languages. Due to language invariance, we can speak of the code length required to specify an object,  $x$ , independent of the particular universal language in which the shortest code for  $x$  is written. This quantity is defined as the Kolmogorov complexity,  $K(x)$ , of that object. Similarly, we can define the conditional Kolmogorov complexity,  $K(y|x)$ , between two objects  $x$  and  $y$ . This is the length of the shortest program that transforms  $x$  into  $y$ .

So, by assuming that the coding language that the cognitive system uses is very general (i.e. universal) we can avoid having to provide a detailed account of the codes that the cognitive system uses and still develop a simplicity-based account of cognition. But the project of specifying how information is coded by the cognitive system remains, of course, one of the central goals of psychology. Moreover, as I discuss below, the choice of language may be crucially important in developing simplicity-based theories of particular cognitive processes—essentially because although code lengths in any two universal languages are equivalent up to a constant, that constant may be very large. Nonetheless, in some cognitively relevant contexts, different languages seem to give surprisingly well-correlated shortest code lengths (Simon, 1972) and hence surprisingly similar measures of simplicity. In these cases, psychological theorizing may be possible without any prior certainty concerning which coding language the cognitive system actually uses.

We have noted that Kolmogorov complexity theory assumes that the cognitive system uses a general purpose coding language—specifically, a universal programming language. This talk of universal programming languages may appear rather unpsychological—after all, the cognitive system presumably does not represent information in PROLOG, LISP, or PASCAL! But the notion of a universal programming language is actually very broad—almost any reasonably rich system of representation, including most proposals concerning mental representation, are universal, and hence the Kolmogorov complexity measure can be applied.

So we now have a definite interpretation of the claim that patterns are chosen on the basis of simplicity: The pattern is chosen to encode the data as briefly as possible. We now consider why this preference for simplicity is justified.

## The Justification of Simplicity

There are various criteria by which a particular choice of pattern in a set of data might be justified. The best pattern might be the pattern that is the most likely explanation of how the data were generated; the pattern that gives rise to the best predictions; or the pattern that provides the best basis for decision making. I now consider the justification of simplicity in each of these ways.

*Simplicity and the Most Likely Explanation.*<sup>3</sup> Suppose that we have data,  $D$ , and a set of hypotheses concerning patterns in the data. By definition, the most likely hypothesis is the hypothesis,  $H$ , that has the greatest probability, given the data. In symbols, this is the  $H$  that maximizes  $P(H|D)$ . Bayes' theorem, a standard theorem of probability theory, states that:

$$P(HD) \propto P(D|H) P(H) \quad 1$$

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, choosing the  $H$  that maximizes (1) is equivalent to choosing the  $H$  that minimizes (2):

$$-\log_2 P(D|H) - \log_2 P(H) \quad 2$$

Under very general conditions,  $-\log_2 P(x)$  is approximated by the Kolmogorov complexity of  $x$ ,  $K(x)$ , and  $-\log_2 P(y|x)$  is approximated by the conditional Kolmogorov complexity of  $y$  given  $x$ ,  $K(y|x)$  (see Li & Vitányi, 1997, for a rigorous analysis, and Chater, 1996, for a more informal discussion). This duality between probabilities and code lengths is of great importance and has been widely used in statistics (e.g. Rissanen,

---

<sup>3</sup> I follow the analysis in Li & Vitányi (1997) in this subsection.

1987, 1989), artificial intelligence (e.g. Cheeseman, 1995), and computer vision (e.g. Mumford, 1992), as well as having direct psychological implications (Chater, 1996).

Given these equivalences, we can rewrite (2) in terms of Kolmogorov complexities:

$$K(D|H) + K(H) \tag{3}$$

This means that choosing the  $H$  that maximizes (1) is equivalent to choosing the  $H$  that minimizes (3). But (3) has the following interpretation:  $K(H)$  is the length of the shortest code to specify the hypothesized pattern,  $H$ ; and  $K(D|H)$  is the length of the shortest code that specifies the data,  $D$ , given  $H$ . The sum of these quantities is therefore the code length of the data, using the hypothesized pattern—the code consists of two parts: first, the pattern is specified, and second the specific data are specified in terms of the pattern. Therefore, (3) can be informally glossed as follows:

$$\text{Shortest code for } D, \text{ using } H \tag{4}$$

According to Bayes' theorem,  $H$  should be chosen to be as probable as possible, that is to maximize (1). But we have seen that this is equivalent to choosing  $H$  to minimize (4): that is, the pattern should be chosen in order to provide the simplest specification of the data. Therefore, choosing the simplest hypothesized pattern is justified because it amounts to choosing the pattern that is the most likely explanation of the data.<sup>4</sup>

*Simplicity and Prediction.* Let us consider prediction in the simple setting where the environment consists of a string of 0s and 1s. A continuous portion  $x_1, \dots, x_n$ , of the sequence is observed—the task is to predict the next item,  $x_{n+1}$ , in the sequence. By elementary probability theory:

$$P(x_{n+1} | x_1, \dots, x_n) = \frac{P(x_1, \dots, x_n, x_{n+1})}{P(x_1, \dots, x_n)} \tag{5}$$

---

<sup>4</sup> The mathematically inclined reader may feel that the above identification between probability and Kolmogorov complexity is bizarre (readers without such qualms may prefer to skip this footnote). The hypothetical reader's question is: How can Kolmogorov complexity be used to approximate an *arbitrary* probability distribution,  $P$ ? If  $P$  is really arbitrary, why can it not simply be chosen so that  $-\log_2 P(x)$  is not approximated by  $K(x)$ , or so that  $-\log_2 P(Y|x)$  is not approximated by  $K(y|x)$ ? If  $P$  can be chosen so that these approximations are not valid, then the crucial relationship between probability and Kolmogorov complexity is undermined. This concern is legitimate—a hidden and crucial assumption is that the probability measures are computable. Assuming the Church–Turing thesis, these distributions are, in any case, the only ones that the cognitive system can entertain, so this is a mild restriction in this context. Remarkably, it then turns out that, for almost all  $y$ ,  $-\log_2 P(y|x) = K(x)$ , up to a constant (see Li & Vitányi, 1997, for details). Thus, under these restrictions, the translation between probability and Kolmogorov complexity is legitimate. Thus, under these conditions, the simplest explanation of a set of data is also the most likely to be true.

The best prediction of  $x_{n+1}$  is the one that has the highest probability of being true—that is, that maximizes (5). Because the denominator does not contain  $x_{n+1}$  the best prediction will also maximize

$$P(x_1, \dots, x_n, x_{n+1}) \quad 6$$

and will minimize:

$$-\log_2 P(x_1, \dots, x_n, x_{n+1}) \quad 7$$

Using the equivalence between Kolmogorov complexity and probability, as above, the best prediction  $x_{n+1}$  therefore minimizes (aside from rare “pathological” cases, which can be ignored in practice):

$$K(x_1, \dots, x_n, x_{n+1}) \quad 8$$

Thus, prediction is achieved by finding the pattern that is the basis for the shortest code for  $x_1, \dots, x_n$  and then choosing the next item  $x_{n+1}$  that follows according to that pattern. We can therefore conclude that the predictions of the pattern chosen by the simplicity principle are the most likely to be true.

This heuristic argument has been made rigorous in Li & Vitányi (1997). In addition, mathematical justifications for prediction based on patterns that are chosen to maximize simplicity have been provided in other mathematical contexts (e.g. Rissanen, 1987, 1989; Vapnik, 1995). Predictions based on simplicity have also been successful in a range of practical applications (e.g. Goa, Li, & Vitányi, 1989; Quinlan & Rivest, 1989), and indeed simplicity (often under the label “Occam’s razor”) is a fundamental principle of contemporary machine learning theory (e.g. Kearns & Vazirani, 1994). Thus, choosing patterns on the basis of simplicity appears justified as a basis for prediction.

*Simplicity and Decision Making.* Finding patterns by simplicity allows an agent to predict and explain the world. These are abstract goals, but nonetheless goals that are of fundamental importance to guiding decisions about practical action. The standard normative theory of how decisions should be made—decision theory—requires associating each possible outcome of the decision maker’s actions with a number representing its utility; and assessing the probability of each outcome if each particular action is taken (Berger, 1985). The rough goal is to choose actions that are likely to lead to outcomes with positive utility and unlikely to lead to outcomes with negative utility. At a technical level there are various ways in which probabilities and utilities can be combined to implement this goal—the most popular is to choose the action that has the highest expected utility (i.e. where expected utility is the utility of each possible outcome, weighted by its probability).

This account of how decisions should be made has a clear role for the simplicity principle—simplicity determines the probability of possible events, which are then combined with utilities to determine what action should be taken. Thus, the simplicity principle can be related, albeit indirectly, not merely to the abstract normative goals of

inferring the most probable pattern, or predicting what will happen, but to the concrete problems of deciding how to act.

## FINDING PATTERNS: DESCRIBING COGNITIVE FUNCTION

I have argued that finding patterns should proceed by choosing patterns that support the shortest encoding of the relevant data. This suggests a possible (although of course partial) account of the remarkable success of the cognitive system in prediction, understanding, and acting in an uncertain and complex environment: that cognitive processes search for simplicity. I now consider whether this proposal provides a basis for plausible descriptive psychological theories. I begin by giving a broad outline of how the proposal that cognition is guided by simplicity should be understood. I then consider two case studies, taken from the study of perception and similarity, which show how this approach can lead to specific theoretical proposals. Finally, I outline in more general terms the potential implications of the simplicity principle for understanding some core aspects of cognition: learning, memory, and reasoning.

### Simplicity and Cognition: The broad picture

*A Psychological Simplicity Principle.* The normative discussion suggests that the cognitive system should aim to find the simplest possible interpretation of the information available to it. But this will not be possible in general. From a computational point of view, the problem of finding the shortest encoding of a set of data is generally computationally intractable. Moreover, it is empirically obvious that people do not always find the simplest interpretation of the information they are given. A well-known example arises in Glass patterns (Glass, 1969; Glass & Pérez, 1973) where there is a spatial translation between two identical superimposed copies of a random dot pattern. The simplest interpretation of such patterns will, of course, exploit the identity between the two copies of the pattern. The relevant code might, for example, first describe one of the random patterns, and then a translation, which specifies the relative location of the other copy of the pattern. This means that the dots in the second copy of the pattern do not have to be specified individually, but can be captured at a stroke by their relation to the first copy. But in order to find such a code, the perceptual system must recognize that the stimulus consists of two copies of the same pattern. If the spatial separation between the patterns is sufficiently small in relation to the spacing between neighbouring dots in each individual pattern, then the perceptual system does recover the identity between the two copies of the pattern (subjectively, this is perceived as a “flow” in the direction of the translational shift between the two copies). But if the spatial separation is sufficiently large, then the Glass pattern is perceived as a completely random field of dots—each of these dots must therefore be encoded separately, without being able to exploit the relationship between the dots in each of the superimposed patterns. Thus, there is a short description of the stimulus, but the perceptual system is unable to find it. Chater (1996) considers a more extreme case: A binary expansion of  $\pi$  represented as a pattern of black or white squares

would appear completely random; its simple description (as an expansion of  $\pi$ ) cannot be discovered by the cognitive system. Indeed, this simple description would also elude any known statistical procedure for detecting structure, because the expansion appears to “pass” all known statistical tests for randomness (Li & Vitányi, 1997).

A psychological form of the simplicity principle, therefore, cannot specify that the cognitive system succeeds in finding the shortest description of the information available to it. Rather, simplicity should be viewed as a goal of cognitive processing: The cognitive system chooses the simplest interpretation of this information that it can find.

A further important issue in the psychological interpretation of the simplicity principle concerns mental representation. I noted previously that Kolmogorov complexity theory abstracts over representation languages, so that the theory can be used as a general framework for theorizing about cognition without a detailed understanding of the nature of mental representation. Nonetheless, the specific representations used by the cognitive system will be of crucial importance in detailed psychological explanation. Indeed, note that, according to the relationship between simplicity and probability above, the coding language can be viewed as encoding a set of prior probabilities concerning possible patterns. Evidence concerning mental representation from any source may thereby be useful in providing constraints on the predictions of simplicity-based accounts of cognition. For example, evidence from linguistics or psycholinguistics concerning the nature of the mental representations involved in understanding natural language must be taken into account in any simplicity/ likelihood account of how the cognitive system finds structure in speech.<sup>5</sup>

Having considered how the simplicity principle can be interpreted as a psychological proposal, I now consider how it can be applied to understanding cognition. I first outline applications to two specific areas: perception and similarity. I then sketch, in broad terms, how simplicity can be related to other major topics in cognitive psychology.

### Case Study 1: Perception

Perception is, from an abstract point of view, a process of finding patterns in sensory input. Thus, a simplicity criterion for choosing between patterns may potentially be applied across a wide range of aspects of perceptual analysis. For example, in low-level perception, it has been conjectured that the *compression* of the information in the sensory signal is a central goal (Atick & Redlich, 1990; Barlow et al., 1989; Blakemore, 1990). For example, lateral inhibition in the retina may be explained as removing local correlations in retinal input, thus providing a less redundant and hence more compressed representation of that input. The goal of compression is frequently viewed as stemming from limitations in the information-carrying capacity of the sensory pathways. However, the viewpoint

---

<sup>5</sup> Although neither the simplicity nor the likelihood principle is currently explicitly advocated in the context of human language processing, many proposals in this area can be viewed in these terms, as Martin Pickering (personal communication) has pointed out. For example, Frazier (1979) can be viewed as arguing that the parser prefers structures that are syntactically simple; and other theorists (e.g. MacDonald, Pearlmutter, & Seidenberg, 1994) can be viewed as arguing for a likelihood approach to parsing, where probabilistic information is drawn from a wide range of sources.

outlined here suggests a complementary interpretation. This is that compressed (i.e. simple) perceptual representations are preferred because of their “cognitive” advantages as outlined above: they tend to involve the extraction of features likely to have generated the sensory input, to support the best predictions, and to provide the best basis for making decisions. Thus, the compression in the sensory signal may provide the best representations, as well as minimize problems of information transmission in sensory pathways.

Moreover, the same principles might equally well be at work in high-level perceptual processing—the simplicity principle seems equally valuable in attempting to understand the causal structure of a sequence of observed actions or events. The key goal is to find patterns that are a reliable basis for explanation and prediction; we have seen that following a simplicity principle is a way of achieving goals of this kind. The simplicity principle therefore finds potential applications in understanding perception at many scales. Which areas of applications prove to be theoretically fruitful remains for future research—I now discuss some areas where the notion of simplicity has already been usefully applied.

*Perceptual Organization.* How does the perceptual system derive a complex and structured description of the perceptual world from sensory input? Two apparently competing theories of perceptual organization have been influential. The first, initiated by Helmholtz (1910/ 1962), advocates the *likelihood principle*: that sensory input will be organized into the most probable distal object or event consistent with that input. The second advocates what Pomerantz and Kubovy (1986) call the *simplicity principle*: The perceptual system is viewed as finding the simplest, rather than the most likely, perceptual organization consistent with the sensory input. But we have already seen that there are close connections between simplicity and probability—which suggests that these theories can perhaps be viewed as identical rather than diametrically opposed.

Interestingly, there have long been suspicions that the two principles are not in fact separate but are two sides of the same coin. Pomerantz and Kubovy (1986) cite Mach (1886/ 1959): “The visual sense acts therefore in conformity with the principle of economy [i.e. simplicity], and at the same time, in conformity with the principle of probability [i.e. likelihood]” (p. 215), and they suggest that some resolution between the two approaches might be possible—particularly in view of the fact that both likelihood and simplicity explanations typically appear to be available for most phenomena in perceptual organization. Chater (1996) notes that the simplicity and likelihood principles are indeed equivalent under natural interpretations, because of the mathematical equivalence between simplicity and likelihood that we discussed above.

The unification of the simplicity and likelihood views appears to be challenged, however, by empirical results that appear to distinguish between likelihood and simplicity. If the two principles are identical, empirical evidence distinguishing between them should not be possible. Chater (1996) argues, however, that such evidence can be interpreted in both the simplicity and likelihood frameworks. I briefly consider such evidence, and how it can be seen as compatible with both the simplicity and likelihood viewpoints.

Likelihood is widely assumed to be favoured by evidence that shows that the preferred perceptual organization is influenced by factors concerning the structure of the everyday environment. For example, consider two-dimensional projections of a shaded pattern,

which can be seen either as a bump or an indentation (see e.g. Rock, 1975). The preferred interpretation is consistent with a light source from above, as in natural light. Thus, the perceptual system appears to choose the interpretation that is most *likely*; but there is no intuitive difference between the *simplicity* of the two interpretations. But such phenomena also have a simplicity-based explanation. Consider the simplest description not of a single stimulus, but of a typical sample of natural scenes. Any regularity that is consistent across those scenes need not be encoded afresh for each scene—rather, it can be treated as a “default”. That is, unless there is a specific additional part of the code for a stimulus that indicates that the scene violates the regularity (and in what way), it can be assumed that the regularity applies. Therefore, other things being equal, scenes that respect the regularity can be encoded more briefly than those that do not. Moreover, perceptual organizations of ambiguous scenes that respect the regularity will be encoded more briefly than those that violate it. In particular, then, the perceptual organization of an ambiguous stimulus obeying the natural regularity of illumination from above will be briefer than the alternative organization with illumination from below. In general, preferences for likely interpretations also give rise to preferences for simple interpretations: If the code for perceptual stimuli and organizations is to be optimal when considered over all (or a typical sample of) natural scenes, it will reflect regularities across those scenes.

Simplicity is assumed to be favoured by cases of perceptual organizations that violate, rather than conform to, environmental constraints. Leeuwenberg and Boselie (1988) show a schematic drawing of a symmetrical two-headed horse. The more likely interpretation, also consistent with the drawing, is that there are two horses, one occluding the other. But the perceptual system appears to reject likelihood. Instead, the drawing is interpreted as a single, two-headed animal. But we can also provide a likelihood explanation of this phenomenon, where likelihood applies locally rather than globally. That is, the perceptual system may determine the interpretation of particular parts of the stimulus according to likelihood (e.g. the fact that there are no local depth or boundary cues may locally suggest a continuous object). These local processes may not always be guaranteed to arrive at the globally most likely interpretation (see Hochberg, 1982).

Thus, the evidence that distinguishes between the simplicity and likelihood principles is actually compatible with both and therefore does not challenge the unification between them.

*Figural Goodness.* Some perceptual patterns are intuitively judged to be more “regular” or “better” than others. These intuitive judgements of “figural goodness” appear to correlate reliably with the resistance of such patterns to noise and the speed with which such patterns are detected. Perhaps the most well developed accounts of goodness are based on the assumption that the goodness of a figure relates to the number of symmetries that it possesses with respect to transformations (e.g. reflection, translation, and so on) (e.g. Palmer, 1983). But Hochberg and McAlister (1953) for a different viewpoint proposed identifying figural goodness with simplicity.

One line of argument in favour of Hochberg and McAlister’s viewpoint is that there is an interesting connection between the simplicity principle in perceptual organizations and noise resistance of patterns, one of the standard litmus tests for the goodness of a pattern.

Specifically, if the simplicity principle is right, then it follows that simple patterns will be the most noise resistant.

Why is this true? The intuitive idea is that the noise resistance of a pattern depends on a comparison between a “null” organization, in which the pattern is not imposed and the stimulus is viewed purely as noise, and a “pattern + noise” interpretation, in which the stimulus is viewed arising from a pattern that has been corrupted by noise. Choosing between these two organizations is, of course, a special case of the general problem of choosing between perceptual organizations, and according to the simplicity principle, the preferred organization will be the one that supports the briefest code for the stimulus. The “null” organization merely treats the entire stimulus as noise. Suppose that encoding the stimulus in this way uses  $S$  bits of information. The “pattern + noise” interpretation is associated with a code that first specifies the pattern and then specifies the way in which the pattern is “disturbed” by the noise. Suppose that encoding the pattern requires  $P$  bits and the noise applied to the pattern requires  $N$  bits. Hence encoding the stimulus using the “pattern + noise” organization requires  $N + P$  bits. According to the simplicity principle, the “pattern + noise” organization will be preferred when it supports the shortest code for the data—that is, when  $N + P < S$ . Thus, the pattern will be perceived if the amount of noise  $N$  is less than  $S - P$ . Thus, as the complexity of the pattern,  $P$ , increases, noise tolerance is reduced. Thus, simple patterns should be the most tolerant to noise.

We can now see why the simplicity principle in perceptual organization implies that simple patterns will be the most noise resistant—because pattern complexity is inversely related to noise tolerance. Given that noise resistance is a litmus test for figural goodness, this suggests the possibility that simple patterns will be particularly good. Thus, the simplicity principle in perceptual organization appears to suggest that simplicity also governs goodness, as Hochberg and McAlister propose.

*Randomness.* If *simplicity* determines judgements of “goodness” or “regularity”, then this suggests that *complexity* might determine judgements of “randomness” or “irregularity”. That is, perhaps judgements of randomness can be viewed as the inverse of goodness judgements (see e.g. Alberoni, 1962). If perceived goodness is determined by the degree to which the cognitive system succeeds in finding structure in the stimulus, then this suggests that perceived randomness may be determined by the degree to which the cognitive system *fails* to find such structure. Interestingly, Falk and Konold (1997) have recently provided support for this view. They give a persuasive theoretical analysis as well as empirical confirmation of the suggestion that subjective judgements of the randomness of a stimulus are inversely related to the success of people’s attempts to find a brief code for that stimulus. Indeed, Falk and Konold’s (1997) analysis proposes an algorithmic definition of randomness drawn from Kolmogorov complexity theory (Li & Vitányi, 1997), thus using the same tools as the current analysis of simplicity at a technical level. This is a straightforward inversion of the simplicity account of goodness: A stimulus is perceived as random to the extent that no simple organization can be found for it. Thus, the simplicity approach promises to unify the literature on goodness with that on judgements of randomness (e.g. Bar-Hillel & Wagenaar, 1991; Budescu, 1987; Lopes & Oden, 1987).

## Case Study 2: Similarity

Consider the problem of finding patterns in a stimulus consisting of two distinct objects. Each object may contain internal patterns; but in addition, there may be patterns that interrelate the two objects. For example, a short description of the stimulus shown in Figure 3a would exploit the common patterns between the left and right object; specifically by noting that one is the mirror image of the other in a vertical axis of symmetry. The pattern interrelating the two parts of the stimulus is very strong; once one half of the stimulus is described, the other can be generated very simply, by specifying the axis of symmetry. Figure 3b shows a pair of objects that share somewhat less structure—specifying one in terms of the other requires a reflection and the interchange of black and white. Figure 3c shows a case where there is less structure still; to specify one object in terms of the other requires an additional translation of the inner figure.

Suppose that we ask: How similar are the pairs of objects in Figure 3? Intuitively, similarity appears to decrease from (a) to (c). Thus, the more shared patterns between two stimuli, and therefore the more simply one can be specified in terms of the other, the more similar they are. Generalizing this observation leads to the proposal that the judged similarity between two objects depends on the complexity of the *transformation* from the representation of one object to the representation of the other. If the shortest transformation is simple, then the representations are similar; if the shortest transformation is complex, the representations are dissimilar. In terms of Kolmogorov complexity, the complexity of the transformation between two representations is measured by the conditional Kolmogorov complexity,  $K(y|x)$ , introduced previously—the length of the shortest program that transforms  $x$  into  $y$ . Bennett, Gács, Li, Vitányi, & Zurek (in press) have developed a deep mathematical theory showing why measures based on conditional Kolmogorov complexity provide a natural definition of the “distance” between two representations (see also Li & Vitányi, 1997). Note, however, that just as the cognitive system can judge simplicity only approximately (because there are simple codes that it cannot find), it can judge only approximately the complexity of a transformation between two representations (because there will be simple transformations that it cannot find).

Thus, the degree of similarity between two cognitive representations will depend on the code length of the shortest transformation that “distorts” one representation into the other that the cognitive system can find. Ulrike Hahn and I (Chater & Hahn, 1996; Hahn & Chater, 1997) have called this the *representational distortion* theory of similarity—the simpler the transformation the cognitive system finds between the representations of a pair of objects, the more similar those objects are assumed to be.

Representational distortion provides an interesting generalization of current psychological theories of similarity. The two leading accounts—the geometric and featural views—also treat similarity as a relation between mental representations. But whereas representational distortion applies to any kind of representation and allows arbitrary computable transformations between them, these theories are committed to specific types of representations and particular relations between them.

The geometric view (Shepard, 1987) assumes that objects are represented as points in an internal space. The similarity between two objects is inversely related to the distance between their representations in this space. By contrast, the set-theoretic view (Tversky,

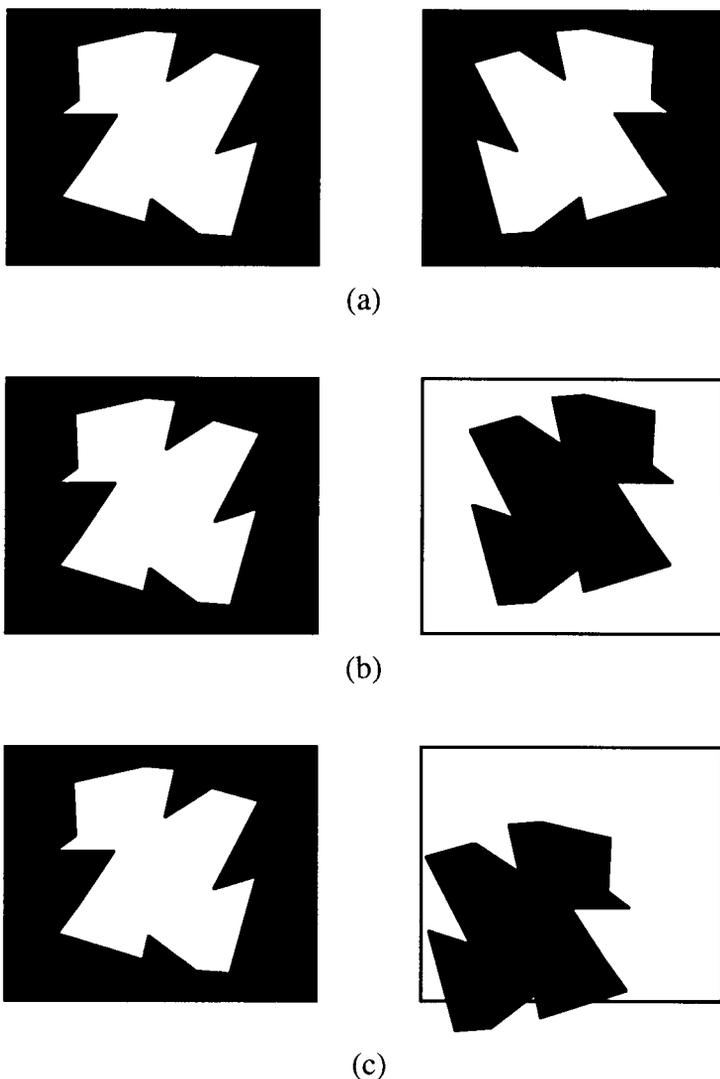


FIG. 3. Pairs of objects that contain shared patterns.

1977) assumes that objects are represented as sets of features. The similarity between two objects depends on the degree of overlap between their sets of features. The representational limitations of both accounts are severe. It does not seem possible to represent perceptual organizations, parsed sentences, schemas for world knowledge, or sequences of motor commands either as points in an internal space or as sets of features. Rather, they appear to require *structured representations*, which are able to capture relations between parts and wholes and to capture systems of relations between parts (Chomsky, 1965; Fodor, 1975; Fodor & Pylyshyn, 1988; Marr, 1982; Minsky, 1977). In short, structured

representations appear to be required to represent almost all cognitively significant stimuli; and judgements of similarity between such stimuli thereby fall outside the scope of both geometric and set-theoretic accounts of similarity.

I stress that representational distortion, like the geometric and set-theoretic views, is defined over mental representations of objects—not over the objects themselves. To see why this is crucial, consider the psychological similarity of two unrelated bursts of white noise. At an acoustic level of description, where the bursts are considered as amplitudes varying over time, a very long set of instructions would be required in order to transform one of these bursts into the other. But the two noises may, nonetheless, be judged to be similar, even to the extent that the auditory system cannot distinguish the two. According to this account, this is because the mental representations of the two bursts do not include minute details of each aspect of the noise. Instead, they give a more general description, perhaps concerning the duration, loudness, or location of the burst. These properties may be largely or completely matched between stimuli, so that the mental representations of the two sounds are identical, or they may differ only slightly, and hence the representational distortion between them is small.

I stress again that the representational distortion found by the cognitive system will correspond only approximately to the conditional Kolmogorov complexity—that is, the code length of the shortest transformation found by the cognitive system will only approximate to the code length of the absolutely shortest transformation. Discovering a transformation with a short code between one representation and another may require arbitrary amounts of computation. For example, the sequences 1 5 3 7 2 3 9 0 6 and 3 0 7 4 4 7 8 1 2 are very simply related—if they are interpreted as base-10 numbers, the second is double the first. Hence the representational distortion between the two sequences is small; however, the cognitive system may not find this short transformation; and the similarity between the two representations may be judged to be low. We assume therefore only that the cognitive system can approximate representational distortion to some degree.

## Geometric and Set-theoretic Theories are Special Cases of Representational Distortion

I have noted that representational distortion is a very general account of similarity, in that it applies to representations of any kind. But it is also possible to view representational distortion as a generalization of both the standard psychological accounts of similarity: the geometric and set-theoretic models. I shall give an intuitive sketch of why this is so, omitting mathematical details for brevity.

*The Geometric Account.* Representations are limited to vectors of numbers. Transformations are limited to sequence of “nudges” of unit length (this length can be thought of as a limit of resolution in the space), and a “program” consists of a sequence of such nudges. If nudges can be in any spatial direction, then the simplest transformation between two points is given by the distance of the straight line path between the points (this is the length of the program of concatenated nudges, ignoring the cost of specifying the direction of a nudge). This gives the Euclidean version of the spatial model, which is

frequently used in psychological modelling. If the nudges can occur only along the axes, then the representational distortion between any two items will depend on the sum of the distance between them for each axis. With this restriction, the representational distortion account is identical to a spatial model with a “city-block” metric, which is another commonly used model in psychological models of similarity. Finally, if the axes in space can be non-orthogonal axes, the resulting representational distortions models correspond to the general Euclidean scaling model (Ashby & Townsend, 1986). So a range of geometric models of similarity can be viewed as generalizations of the representational distortion account.

*The Set-theoretic Account.* Representations are limited to sets of features. Transformations are limited to the deletion and addition of features one by one. Thus a program consists of a sequence of deletions and additions. Assuming differential length for deletion and addition (specifically, deletion has the shorter code, because additions require specifying what is to be added), program length is then determined from a weighted sum of the number of features that object A has and object B does not (which must be deleted) and that B has but A does not (which must be added). The length of this program is a close variant of Tversky’s (1977) theory of similarity.

## Properties of the Representational Distortion Theory of Similarity

I now consider briefly some basic properties of representational distortion that suggest that it is a promising starting point for a psychological theory of similarity.

*Flexibility.* The fact that similarity is defined over general representations takes account of the great flexibility of human similarity judgements (e.g. Medin, Goldstone, & Gentner, 1993), because similarity is defined over representations of objects, and the goals and knowledge of the subject may affect the representations that are formed. As with the set-theoretic models (Tversky, 1977), this flexibility has both advantages, in terms of accounting for the flexibility of people’s similarity judgements, and disadvantages, from the point of view of deriving testable empirical predictions.

*Similarity and Identity.* According to representational distortion, each object is more similar to itself than to any other object. This is because the shortest possible programme is the “empty” programme, which, clearly, leaves any representation unchanged: in symbols,  $K(x|x) = 0$ . Thus, the representational distortion viewpoint automatically captures the fundamental intuition that identity is the most extreme form of similarity—an intuition that appears central to the very notion of similarity (Hahn & Chater, 1998).

This property of representational distortion seems attractive, but it appears to run counter to data obtained by Tversky (1977), which appear to show that the similarity between distinct objects can sometimes exceed the similarity between identical objects. However, it may be too early to take the drastic conceptual step of rejecting the intuition that identity is the most extreme form of similarity.

*Asymmetry.* Representational distortion allows for asymmetry in similarity judgements:  $K(x|y)$  is not in general equal to  $K(y|x)$ . This asymmetry is particularly apparent when the representations being transformed differ substantially in complexity. Suppose that a subject knows a reasonable amount about China but rather little about Korea, except that it is “rather like” China in certain ways. Then transforming the representation of China into the representation of Korea will require a reasonably short program (which simply deletes large amounts of information concerning China that are not relevant to Korea), whereas the program transforming in the reverse direction will be complex, as the minimal information known about Korea will be almost no help in constructing the complex representation of China. Thus, we would predict that  $K(\text{China}|\text{Korea})$  should be greater than  $K(\text{Korea}|\text{China})$ , and hence that Korea should be judged to be more similar to China than China is to Korea. This is observed experimentally (Tversky, 1977).

*Background Knowledge.* Similarity judgements are influenced by background knowledge. For example, if the Arabic number system is part of your background knowledge, then you may perceive similarities between otherwise dissimilar patterns. To choose a simple example, a typewritten “7” may, when considered as a mere pattern of dots, look completely different from a scrawled hand-written “7”. But for a person who knows the Arabic number system, the two patterns are perceived as very similar, because they are both encoded as “sevens”. It is difficult for any theory of similarity to explain the role of background knowledge. In the spatial view, the natural role of knowledge is in specifying the dimensions of the space in which the comparison takes place, as well as assigning weights, which determine the relative importance of each dimension (effectively by stretching or squashing the space along the relevant dimension). In the set-theoretic view, background knowledge can play a role in determining the features that are taken into account in the comparison. In both cases, the role of knowledge is to affect the representations that are the *input* to the comparison process. Similarly, background knowledge may affect the representations that are compared, according to the view that similarity is representational distortion. But, moreover, the representational distortion account allows an additional way in which background can affect similarity comparisons: by assuming that background knowledge forms an additional input to the program that transforms one object into another. Thus, background knowledge affects what operations are available in transforming one representation into another—for example, a knowledge of the number system might suggest all manner of numerical transformations that might relate two numbers (e.g. having the concept of a prime might increase the degree to which people judge 43 and 47 to be similar—partly because one can be generated from the other by the instruction “next prime” or “previous prime”). People with different mathematical knowledge might thereby have different judgements about which numbers are similar. More drastically, people who use different mathematical notations will thereby have dramatically different judgements concerning the similarities between patterns corresponding to formulae expressed in various notations.

Thus, representational distortion provides a rich framework for understanding how background knowledge influences similarity judgements—knowledge can readily influence the nature of the similarity comparison itself, as well as changing the representations that are inputs to the comparison (see Hahn & Chater, 1998, for discussion of related

issues). It remains for future work to determine to what extent this account can capture in detail the way in which people's similarity judgements are influenced by their background knowledge.

*Summary.* To sum up, the representational distortion approach to similarity arises as follows. If the cognitive system searches for the simplest interpretation of the information available, then it will aim to exploit regularities between different representations. The strength of the shared regularities between two objects can be measured by the conditional Kolmogorov complexity between them: let us call the cognitive system's approximation to this quantity "representational distortion". Representational distortion can be viewed as a generalization of the two standard psychological models of similarity: the spatial and set-theoretical models. Moreover, it has a number of intuitively attractive properties. An interesting project for future research is to attempt to develop this theoretical account in more detail, and to provide experimental tests for this approach (see Chater & Hahn, 1996).

## Simplicity as a General Principle in Cognition

I now turn from specific case studies, to consider more generally how the simplicity principle may apply to cognition, briefly considering three areas: learning, memory, and reasoning.

*Learning from Experience.* Learning from experience<sup>6</sup> is a problem of finding patterns in what are typically large amounts of complex and often noisy data. It therefore falls naturally within the domain of application of the normative theory of finding patterns by searching for simplicity. Moreover, theorists have directly proposed that certain aspects of language acquisition may proceed by finding the shortest possible encoding of the input linguistic data. For example, Brent and Cartwright (1997) show how morphological structure can be found within isolated words, Wolff (1977; see also Redlich, 1993) considers how higher level structure can be found automatically in text. Less directly, connectionist networks, currently perhaps the most popular computational models of human learning (Elman et al., 1996) can be interpreted as implementing Bayesian probabilistic inference (MacKay, 1992; Neal, 1993), and thus, by the connection between probability and simplicity, as maximizing simplicity. Indeed, much recent interest in the technical literature on connectionists networks has focused on directly viewing networks as minimizing description length and, therefore, as maximizing simplicity. Thus, many current psychological models of learning are compatible with the thesis that the cognitive system maximizes simplicity.

*Memory.* The claim that the cognitive system searches for patterns that provide the briefest encoding of available information has an important side-effect in terms of memory: that the cognitive system thereby seeks to minimize memory load. Brief

---

<sup>6</sup> A different analysis may be appropriate for understanding learning from instruction, and learning from both instruction and experience. The question of how or whether the search for simplicity, as outlined here, can be applied in such contexts is an important topic for future research.

explanations are, I have argued, useful for explanation, prediction, and action—but fortunately they also require the least space in memory.

As pointed out by pioneers of the application of information in psychology (Attneave, 1959; Garner, 1962, 1974), this point of view leads to the prediction that the richer the patterns that the cognitive system can find in a stimulus, the shorter its code will be, and hence the better it will be remembered. This is a ubiquitous finding in all areas of memory research, from the memory advantage for words over nonsense letter strings, to the memory advantage for meaningful pictures over meaningless patterns, and for comprehensible stories over incomprehensible passages.

It is important to note that this account, although focusing on brevity of encoding as a theoretical construct, does not depend on the assumption that the memories are stored as briefly as possible—that is, with no redundancy. Indeed, it has frequently been observed that this kind of storage would be inappropriate for the cognitive system, because it would not be robust to noise. But even if memories are stored redundantly, the cognitive system should still search for shortest descriptions to achieve the best memory performance. Information theory specifies that constructing an optimal redundant code is achieved by first finding the simplest encoding, and then introducing 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 to use for storage, which will thereby be noise resistant and hence better remembered.

*Reasoning.* The simplicity principle provides a normative account of inductive inference, as we have seen. Thus, it may provide a starting point for understanding aspects of reasoning that are traditionally viewed as inductive: for example, how scientists choose between theories or how people learn the causal structure of the world from experience. But here we focus on a more general connection between the simplicity principle and a general conception of human reasoning as fundamentally *probabilistic* rather than *logical* in character.

In a series of papers (Chater & Oaksford, 1990, 1993; Oaksford & Chater, 1991, 1992, 1993, 1995b), Mike Oaksford and I have argued that almost all everyday reasoning is uncertain: people draw conclusions that are plausible, but not certain, given the available premises. We have argued that probability theory, the calculus of uncertainty, is therefore a more appropriate starting point for understanding human reasoning than logic, the calculus of certainty. Moreover, we have argued that people interpret classic psychological reasoning tasks, which are typically assumed to be deductive, in probabilistic terms, and solve them using strategies that can be understood in probabilistic terms. Thus, we argue that people are not logical but that they are rational; Logic is simply the wrong standard against which to assess most human reasoning. This viewpoint has proved useful in providing detailed models of a range of standard reasoning tasks, including Wason's selection task (Wason, 1966, 1968; Oaksford & Chater, 1994, 1995a; Oaksford, Chater, Grainger, & Larkin, 1997; see Almor & Sloman, 1996; Evans & Over, 1996b; Laming, 1996 for critical discussion and Oaksford & Chater, 1996, for a response), syllogistic reasoning (Chater & Oaksford, in press) and conditional inference (Oaksford, Chater, & Larkin, 1997). Because of the duality between simplicity and probability, discussed

previously, the probabilistic interpretation of human inference is immediately compatible with the simplicity principle: Any problem of maximizing probability is equivalent to a problem of maximizing simplicity. So the probabilistic view of human reasoning can be reinterpreted as embodying the simplicity principle.

Note that although this viewpoint proposes that simplicity/probability is a goal of reasoning, it allows that this goal will be only approximately achieved. Theorists differ on how good such an approximation might be. Kahneman and Tversky have argued that their experimental results showed strong departures from the norms of probability theory under certain conditions (e.g. Kahneman & Tversky, 1973; Kahneman, Slovic, & Tversky, 1982), although the reasoning heuristics (availability and representativeness) that they proposed people use, are usually reasonably reliable in normal circumstances. But Gigerenzer and his colleagues (e.g. Gigerenzer, Hell, & Blank, 1988; Gigerenzer & Murray, 1987) have argued that Kahneman and Tversky may have substantially underestimated the normative correctness of human probabilistic reasoning, and they have shown that experimental manipulations that clarify the task for the experimental participant can dramatically improve the fit between reasoning performance and probabilistic norms.

More recently, however, Gigerenzer and Goldstein (1996) and Evans and Over (1996a) have proposed that human performance should not be compared with any normative theories, such as logic or probability theory (or, in the present context, the simplicity principle). They argue that such normative accounts are entirely unnecessary for understanding human reasoning. Specifically, Gigerenzer and Goldstein (1996) argue that reasoning should be understood as consisting of “fast and frugal” heuristics, which are adaptively successful but not normatively justified; and Evans and Over (1996a) argue that much human reasoning is “rational<sub>1</sub>”, that is, successful with respect to achieving a person’s goals, but not “rational<sub>2</sub>”, that is, conforming to a normative analysis. Both these viewpoints suggest that reasoning may consistently succeed without conforming, even approximately, to any normative standard. This seems unsatisfactory, because it leaves the success of human reasoning unexplained (see Chater, Oaksford, Nakisa, & Redington, 1997). By contrast, the simplicity principle has a normative justification and also is intended to describe cognitive performance.

## SCOPE AND LIMITS

I have proposed that the search for simplicity is a fundamental cognitive principle, with potentially broad application in constructing psychological theories. In this section, I list a number of important limitations for this approach.

### Representation

If simplicity is defined in terms of brevity in a representation language, then simplicity will depend crucially on that representation language—thus, obtaining detailed psychological predictions from the simplicity principle requires making specific assumptions about mental representation. I noted above that Kolmogorov complexity theory is able to abstract away from the specific coding language being used, because code lengths in

any two languages are equal up to a constant—but this constant may be large in relation to the number of data available in specific psychological applications. This means that specific theories of representation concerning different cognitive domains will be required to build detailed psychological accounts based on simplicity.

The viability of simplicity-based accounts of cognition can be assessed by constraining the coding language to determine simplicity by independent theoretical and empirical evidence concerning the relevant aspect of mental representation. Leeuwenberg and colleagues (Buffart et al., 1981; Leeuwenberg, 1969, 1971) have pursued this approach in assessing the viability of their coding language for certain classes perceptual stimuli, by developing what they call structural information theory (e.g. van der Helm, van Lier, & Leeuwenberg, 1992; van Lier, van der Helm, & Leeuwenberg, 1994a, 1994b). Similar programs of research may be possible with respect to other applications of the simplicity principle.

## Search

I have argued that the simplicity is the criterion with which the cognitive system chooses between alternative patterns that may be imposed on the environment. But I have also noted that the cognitive system cannot, in general, maximize simplicity—in general, finding the shortest code for a set of data is uncomputable; and even restricted versions of the problem are generally combinatorially explosive. This means that the suboptimal solutions found by the cognitive system will depend on the nature of the search process.

## Speed

The discussion so far has also ignored cognitive limitations concerning speed of processing. In language processing, for example, immensely complex patterns in a speech wave can be decoded within fractions of a second (e.g. Marslen-Wilson, 1973). In view of the slowness of neural hardware, this suggests that the search process must take only a small number of steps—this places very strong constraints on the nature of the search process (Chater & Oaksford, 1990; Feldman & Ballard, 1982; Rumelhart & McClelland, 1986).

Speed also plays an important role in another way: The representations used by the cognitive system must not only be brief but be easy to use quickly. In some contexts, there may be a trade-off between brevity and speed. Consider an example from computer science: Arithmetical operations may be rapidly computed by consulting a large look-up table in which the answers to particular arithmetic operations are prestored (particularly if this table can be searched in parallel); by contrast, a more compact representation of arithmetic, for example, in terms of axioms in some logical language, may be much briefer, but they require much more computation to use. This tension between cognitive goals of speed and brevity may also be important in psychological contexts.

## Innate Constraints

The simplicity viewpoint outlined here may appear to be tied to a strong empiricist view of cognitive development. The emphasis has been on the criteria that the cognitive system can use to *find* patterns; this assumes that the patterns have to be found from experience,

rather than being innately specified. Nonetheless, the simplicity viewpoint is equally compatible with empiricist and nativist viewpoints. Even strong nativists require that the cognitive system searches for patterns—but they claim that this search is subject to strong innate constraints. For example, strong nativist viewpoints regarding language acquisition typically involve the claim that the child can entertain only a restricted set of grammars (e.g. Chomsky, 1980; Pinker, 1984). But the problem of finding the correct grammar most compatible with linguistic experience is still immensely difficult (e.g. see Redington & Chater, 1998, for discussion), and this pattern-finding problem may still be guided by simplicity. The restriction to a small set of grammars amounts to having constrained internal representation in terms of which linguistic hypotheses can be stated. But the simplicity principle may nonetheless apply: The grammar chosen may be that which provides the briefest encoding of linguistic input (see, e.g. Brent & Cartwright, 1997; Grünwald, 1996; Wolff, 1988). More generally, the simplicity principle applies to problems of finding patterns in the world from experience, whether or not finding such patterns is guided by innate constraints.

### The Importance of Interests

I have so far considered the cognitive system as engaged in a *disinterested* search for patterns. This leaves out the fact that some patterns are relevant and others irrelevant to the interests of the agent. Clearly, people are more concerned with each other's faces than with patterns of shadow; and they are more concerned with each other's voices than with the sounds of footsteps or distant traffic. Faces and voices are interesting not merely because they contain rich patterns, but because they are of fundamental importance in relating to other people, and thereby are relevant to achieving almost any goal a person may have. Equally, it seems plausible that the perception of the physical world is to some degree geared towards the detection of *affordances* (Gibson, 1979)—properties of objects that are potentially relevant to the actions of the agent (e.g. whether an object can be eaten, lifted, thrown, and so on).

The role of interests is beyond the scope of the simplicity principle but compatible with it. Interests affect how much cognitive effort is directed towards finding different kinds of patterns; but the pattern-finding process may, nonetheless, proceed without reference to interests and may be guided only by simplicity. Scientific research provides an appropriate analogy—various practical interests may determine the level of resources devoted to different areas of research, but the research itself should use disinterested scientific criteria, that is, that such research should proceed without reference to those interests. Indeed, it is generally assumed that interests must not be allowed to influence scientific research directly (e.g. the conclusions reached should be based purely on evidence, rather than choosing conclusions on the basis of political or social convenience), if scientific research is to be valuable to society. Similarly, I assume that a separation between interests and the criteria for finding patterns is cognitively desirable. Of course, such a separation may not always be achieved. It is anecdotally clear that people do tend to believe what they want to believe. But in general we do not see, hear, and infer only what we want to see, hear, and infer; and were this the case, the consequences would presumably be disastrous (see Fodor, 1983, for related discussion). Indeed, the remarkable

success of the cognitive system suggests that, to a large degree at least, human reason is successfully insulated from the interests that it serves.

## CONCLUSIONS

Many cognitive processes find patterns in experience—from perception to scientific thinking. I suggest that the cognitive system searches for the patterns according to simplicity—where simple patterns are those that allow a brief specification of the available data. This is normatively justified as providing a sound basis for prediction and explanation and provides an attractive framework for descriptive psychological theories in a range of cognitive domains. I propose that it is worth exploring further the hypothesis that the search for simplicity is a fundamental cognitive principle.

## REFERENCES

- Alberoni, F. (1962). Contribution to the study of subjective probability: I. *Journal of General Psychology*, *66*, 241–264.
- Almor, A., & Sloman, S.A. (1996). Is deontic reasoning special? *Psychological Review*, *103*, 374–380.
- Anderson, J.R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Anderson, J.R. (1991a). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*, 471–517.
- Anderson, J.R. (1991b). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.
- Anderson, J.R., & Schooler, L.J. (1991). Reflections of the environment in memory. *Psychological Science*, *1*, 396–408.
- Ashby, F.G., & Townsend, J.T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.
- Atick, J.J., & Redlich, A.N. (1990). Towards a theory of early visual processing. *Neural Computation*, *2*, 308–320.
- Attneave, F. (1959). *Applications of information theory to psychology*. New York: Holt, Rinehart & Winston.
- Attneave, F., & Frost, R. (1969). The determination of perceived tridimensional orientation by minimum criteria. *Perception & Psychophysics*, *6*, 391–396.
- Bar-Hillel, M., & Wagenaar, W.A. (1991). The perception of randomness. *Advances in Applied Mathematics*, *12*, 428–454.
- Barlow, H.B., Kaushal, T.P., & Mitchison, G.J. (1989). Finding minimum entropy codes. *Neural Computation*, *1*, 412–423.
- Bennett, C., Gács, P., Li, M., Vitányi, P., & Zurek, W.H. (in press). Information distance. *IEEE Transactions on Information Theory*.
- Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag.
- Blakemore, C. (Ed.) (1990). *Vision: Coding and efficiency*. Cambridge, England: Cambridge University Press.
- Brent, M.R., & Cartwright, T.A. (1997). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *63*, 121–170.
- Budescu, D.V. (1987). A Markov model for generation of random binary sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 25–39.
- 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. (1966). On the length of programs for computing finite binary sequences. *Journal of the Association for Computing Machinery*, *13*, 547–569.

- Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organisation. *Psychological Review*, *103*, 566–581.
- Chater, N. (1997, November). Simplicity and the mind. *The Psychologist*, 495–498.
- Chater, N., & Hahn, U. (1996). Representational distortion as a theory of similarity. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (p. 741). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Chater, N., & Oaksford, M. (1990). Autonomy, implementation and cognitive architecture: A reply to Fodor and Pylyshyn. *Cognition*, *34*, 93–107.
- Chater, N., & Oaksford, M. (1993). Logicism, mental models and everyday reasoning: Reply to Garnham. *Mind & Language*, *8*, 72–89.
- Chater, N., & Oaksford, M. (in press). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*.
- Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (1997). Fast, frugal and rational: Rational analysis and cognitive algorithms in human reasoning. ms., Department of Psychology, University of Warwick.
- Cheeseman, P. (1995). On Bayesian model selection. In D. Wolpert (Ed.), *The mathematics of generalisation* (pp. 315–330). Redwood City, CA: Addison-Wesley.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. (1980). *Rules and representations*. Oxford: Blackwell.
- Cover, T.M., & Thomas, J.A. (1992). *Elements of information theory*. New York: John Wiley.
- Dinnerstein, D., & Wertheimer, M. (1957). Some determinants of phenomenal overlapping. *American Journal of Psychology*, *70*, 21–37.
- Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness*. Cambridge, MA: MIT Press.
- Evans, J. St. B.T., & Over, D.E. (1996a). *Rationality and reasoning*. Hove, UK: Psychology Press.
- Evans, J. St. B.T., & Over, D.E. (1996b). Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review*, *103*, 356–363.
- Eysenck, M.W., & Keane, M.T. (1990). *Cognitive psychology: A student's handbook* (2nd Ed.). Hove, UK: Lawrence Erlbaum Associates Ltd.
- Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, *104*, 301–318.
- Feldman, J.A., & Ballard, D.H. (1982). Connectionist models and their properties. *Cognitive Science*, *6*, 205–254.
- Fodor, J.A. (1975). *The language of thought*. New York: Crowell.
- Fodor, J.A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Fodor, J.A., & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.
- Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies. PhD. thesis, University of Connecticut, West Ben, IN: Indiana University Linguistics Club.
- 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 Associates, Inc.
- Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.
- Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.
- Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base-rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 513–525.
- Gigerenzer, G., & Murray, D.J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Glass, L. (1969). Moiré effect from random dots. *Nature*, *223*, 578–580.
- Glass, L., & Pérez, R. (1973). Perception of random dot interference patterns. *Nature*, *246*, 360–362.
- Goa, Q., Li, M., & Vitányi, P. (1989). Learning on-line hand-written characters. In *11th International Joint Conference on Artificial Intelligence* (pp. 843–848). San Mateo, CA: Morgan Kaufman.

- Goodman, N. (1983). *Fact, fiction, and forecast* (4th ed.). Cambridge, MA: Harvard University Press.
- Grünwald, P. (1996). A minimum description length approach to grammar induction. In S. Wermter, E. Riloff, & G. Scheler (Eds.), *Symbolic, connectionist, and statistical approaches to learning for natural language processing* (pp. 203–216). Berlin: Springer Verlag.
- Hahn, U., & Chater, N. (1997). Concepts and similarity. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts and categories* (pp. 43–92). Hove, U.K.: Psychology Press.
- Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? exhaustive? empirically distinguishable? *Cognition*, 65, 197–230.
- Harman, G. (1965). The inference to the best explanation. *Philosophical Review*, 74, 88–95.
- Hochberg, J. (1982). How big is a stimulus? In J. Beck (Ed.), *Organisation and representation in perception* (pp. 191–218). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Hochberg, J., & McAlister, E. (1953). A quantitative approach to figure “goodness.” *Journal of Experimental Psychology*, 46, 361–364.
- Hume, D. (1965). *A treatise on human nature*. L.A. Selby-Bigge (Ed.). Oxford: Clarendon Press. (Original work published 1739–1740)
- Johansson, G. (1950). *Configurations in event perception*. Stockholm: Almqvist & Wiksell.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.
- Kanizsa, G., & Gerbino, W. (1982). Amodal completion: Seeing or thinking? In J. Beck (Ed.), *Organisation in representation and perception* (pp. 167–190). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Kearns, M.J., & Vazirani, U.V. (1994). *An introduction to computational learning theory*. Cambridge, MA: MIT Press.
- Kemeny, J.G. (1953). The use of simplicity in induction. *Philosophical Review*, 62, 391–408
- Koffka, K. (1962). *Principles of Gestalt psychology* (5th ed.). London: Routledge and Kegan Paul. (Original work published 1935)
- Kolmogorov, A.N. (1965). Three approaches to the quantitative definition of information. *Problems in Information Transmission*, 1, 1–7.
- Laming, D. (1996). On the analysis of irrational data selection: A critique of Oaksford and Chater (1994). *Psychological Review*, 103, 364–373.
- 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 ed.). New York: Springer-Verlag.
- Lopes, L.L., & Oden, G.C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 392–400.
- MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Mach, E. (1959). *The analysis of sensations and the relation of the physical to the psychical*. New York: Dover Publications. (Original work published 1886)
- Mach, E. (1960). *The science of mechanics*. La Salle, IL: Open Court. (Original work published 1883)
- MacKay, D.J.C. (1992). A practical Bayesian framework for backprop networks. *Neural Computation*, 4, 415–447.
- Marr, D. (1982). *Vision*. New York: Freeman.
- Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522–523.
- Medin, D.L., Goldstone, R., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254–278.

- Minsky, M. (1977). Frame system theory. In P.N. Johnson-Laird & P.C. Wason (Eds.), *Thinking: Readings in cognitive science* (pp. 355–376). Cambridge: Cambridge University Press.
- Mumford, D. (1992). Pattern theory: A unifying perspective. In A. Joseph, F. Mignot, F. Murat, B. Prum, & R. Rentschler (Eds.), *Proceedings of the First European Congress of Mathematics* (pp. 187–224). Basel: Birkhäuser Verlag.
- Neal, R.M. (1993). Bayesian learning via stochastic dynamics. In S.J. Hanson, J.D. Cowan, & C.L. Giles (Eds.), *Neural information processing systems 5* (pp. 475–482). San Mateo, CA: Morgan Kaufman.
- Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind & Language*, 6, 1–38.
- Oaksford, M., & Chater, N. (1992). Bounded rationality in taking risks and drawing inferences. *Theory & Psychology*, 2, 225–230.
- Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K.I. Manktelow & D.E. Over (Eds.), *Rationality* (pp. 31–60). London: Routledge.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608–631.
- Oaksford, M., & Chater, N. (1995a). Information gain explains relevance which explains the selection task. *Cognition*, 57, 97–108.
- Oaksford, M., & Chater, N. (1995b). Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning*, 1, 121–152.
- Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, 103, 581–591.
- Oaksford, M., & Chater, N. (Eds.). (in press). *Rational models of cognition*. Oxford: Oxford University Press.
- Oaksford, M., Chater, N., Grainger, R., & Larkin, J. (1997). Optimal data selection in the reduced array selection task (RAST). *Journal of Experimental Psychology: Learning Memory and Cognition*, 23, 441–458.
- Oaksford, M., Chater, N., & Larkin, J. (1997). *A probabilistic model of conditional inference*. ms., School of Psychology, University of Wales, Cardiff.
- Palmer, S.E. (1983). The psychology of perceptual organisation: A transformational approach. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 269–339). New York: Academic Press.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pomerantz, J.R., & Kubovy, M. (1986). Theoretical approaches to perceptual organisation: 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.
- Popper, K. (1959). *The logic of scientific discovery*. New York, Basic Books. (Original work *Logik der Forschung*, 1st ed., 1934)
- Polyshyn, Z.W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
- Quinlan, J., & Rivest, R. (1989). Inferring decision trees using the minimum description length principle. *Information and computation*, 80, 227–248.
- Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes*, 13, 129–191.
- Redlich, A.N. (1993). Redundancy reduction as a strategy for unsupervised learning. *Neural Computation*, 5, 289–304.
- Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review*, 77, 481–495.
- Restle, F. (1979). Coding theory of the perception of motion configurations. *Psychological Review*, 86, 1–24.
- Rissanen, J. (1987). Stochastic complexity. *Journal of the Royal Statistical Society, Series B*, 49, 223–239.
- Rissanen, J. (1989). *Stochastic complexity and statistical inquiry*. Singapore: World Scientific.
- Rock, I. (1975). *An introduction to perception*. New York: Macmillan.
- Rumelhart, D.E., & McClelland, J.L. (Eds.). (1986). *Parallel distributed processing* (2 volumes). Cambridge, MA: MIT Press.

- Shepard, R.N. (1987). Toward a universal law of generalisation for psychological science. *Science*, 237, 1317–1323.
- Simon, H.A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 79, 369–382.
- Simon, H.A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review*, 70, 534–546.
- Sober, E. (1975). *Simplicity*. Oxford: Clarendon Press.
- Solomonoff, R.J. (1964). A formal theory of inductive inference, Parts 1 and 2. *Information and Control*, 7, 1–22, 224–254.
- Stewart, I., & Golubitsky, M. (1992). *Fearful symmetry: Is God a geometer?* London: Penguin.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
- van der Helm, P.A., van Lier, R., & Leeuwenberg, E.L.J. (1992). Serial pattern complexity: Irregularity and hierarchy. *Perception*, 21, 517–544.
- van Lier, R., van der Helm, P.A., & Leeuwenberg, E.L.J. (1994a). Competing global and local aspects of visual occlusion. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 571–583.
- van Lier, R., van der Helm, P.A., & Leeuwenberg, E.L.J. (1994b). Integrating global and local aspects of visual occlusion. *Perception*, 23, 883–903.
- Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer.
- Vitz, P.C., & Todd, T.C. (1969). A coded element of the perceptual processing of sequential stimuli. *Psychological Review*, 76, 433–449.
- von Helmholtz, H. (1962). *Treatise on physiological optics*. (Vol. 3) (J.P. Southall, Trans.). New York: Dover. (Original work published 1910)
- Wallace, C.S., & Boulton, D.M. (1968). An information measure for classification. *Computing Journal*, 11, 185–195.
- Wallace, C.S., & Freeman, P.R. (1987). Estimation and inference by compact coding. *Journal of the Royal Statistical Society, Series B*, 49, 240–251.
- Wason, P.C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth, U.K.: Penguin.
- Wason, P.C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20, 273–281.
- Wolff, J.G. (1977). The discovery of segmentation in natural language. *British Journal of Psychology*, 67, 377–390.
- Wolff, J.F. (1988). Learning syntax and meanings through optimisation and distributional analysis. In Y. Levy, I.M. Schlesinger, & M.D.S. Braine (Eds.), *Categories and processes in language acquisition* (pp. 179–215). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

*Original manuscript received 1 August 1997*

*Accepted revision received 5 June 1998*